

Politechnika Rzeszowska im. Ignacego Łukasiewicza
Wydział Matematyki i Fizyki Stosowanej

Rozprawa doktorska

Wybrane własności i zastosowania funkcji kwantylowych

Promotor:
dr hab. Anna Kuczmaszewska, prof. PL

Autor:
mgr inż. Dagmara Dudek

Rzeszów 2026

Serdeczne podziękowania składam Pani Promotor
Dr hab. Annie Kuczmaszewskiej, prof. Politechniki Lubelskiej,
za nieocenioną pomoc, cenne wskazówki naukowe,
cierpliwość oraz czas poświęcony na każdym etapie powstawania tej rozprawy.
Dziękuję za życzliwość, wsparcie i motywację.

Niniejszą pracę dedykuję
mojej Córce Jagodzie, Mężowi Sebastianowi,
moim Rodzicom Marcie i Tadeuszowi oraz całej mojej Rodzinie.
Dziękuję za nieustające wsparcie, wiarę we mnie i za to, że zawsze mogę na Was liczyć.

Spis treści

Wstęp 5

Rozdział 1. Wybrane pojęcia związane z kwantylami i strukturami zależności 9

1.1. Wybrane własności funkcji kwantylowych 9
1.2. Estymatory kwantyla 12
1.3. Zależne zmienne losowe 14

Rozdział 2. Reprezentacja Bahadura dla kwantyli z próby 17

2.1. Silna zgodność estymatora kwantyla i reprezentacja Bahadura dla zależnych zmiennych losowych 18
2.2. Jednostajna asymptotyczna normalność kwantyli z próby 25
2.2.1. Symulacje 32

Rozdział 3. Reprezentacja Bahadura w przestrzeni podliniowej wartości oczekiwanej . . 35

3.1. Wprowadzenie i podstawowe założenia 36
3.2. Reprezentacja Bahadura 39

Rozdział 4. Reprezentacja kwantylowa dystrybuanty sumy, różnicy, iloczynu i ilorazu zmiennych losowych 51

4.1. Suma niezależnych zmiennych losowych 51
4.2. Reprezentacja kwantylowa różnicy, iloczynu i ilorazu zmiennych losowych 53
4.3. Aproksymacja i estymacja rozkładu różnicy niezależnych zmiennych losowych 62
4.3.1. Aproksymacja rozkładu 62
4.3.2. Błąd aproksymacji 63
4.3.3. Estymator nieparametryczny 67
4.3.4. Symulacje 73
4.4. Aproksymacja i estymacja rozkładu iloczynu zmiennych losowych 85
4.4.1. Aproksymacja rozkładu 85
4.4.2. Błąd aproksymacji 86
4.4.3. Estymator nieparametryczny 87
4.4.4. Symulacje 92

4.5. Aproksymacja i estymacja rozkładu ilorazu zmiennych losowych 101
4.5.1. Aproksymacja rozkładu 101
4.5.2. Błąd aproksymacji 102
4.5.3. Estymator nieparametryczny 105
4.5.4. Symulacje 112

Podsumowanie 123

Bibliografia 125

Wstęp

Funkcja kwantylowa stanowi jedno z podstawowych narzędzi współczesnej statystyki matematycznej oraz analizy rozkładów prawdopodobieństwa. Formalnie, dla danej dystrybuanty F, funkcję kwantylową (kwantyl) rzędu p definiuje się jako funkcję $Q\colon (0,1) \to \mathbb{R}$ spełniającą warunek

$$Q(p) = \inf\{x \in \mathbb{R} : F(x) \geqslant p\}.$$

Z praktycznego punktu widzenia kwantyl rzędu p to najmniejsza wartość x, dla której dystrybuanta osiąga lub przekracza wartość p, czyli taka, że co najmniej p-ta część rozkładu zmiennej losowej nie przekracza wartości x. Powszechnie znana i wykorzystywana w statystyce mediana, to kwantyl rzędu $0,5$. Pierwszy i trzeci kwartyl, to odpowiednio kwantyl rzędu 0,25 i kwantyl rzędu 0,75.

Funkcja kwantylowa dostarcza alternatywnego, równoważnego opisu rozkładu prawdopodobieństwa. Z praktycznego punktu widzenia funkcje kwantylowe odgrywają istotną rolę w wielu dziedzinach statystyki i w modelowaniu. Umożliwiają one między innymi:

- generowanie prób losowych z dowolnego rozkładu;
- analizę pozycyjnych miar rozkładu;
- modelowanie rozkładów o niestandardowych kształtach;
- opis zjawisk ekstremalnych;
- modelowanie ryzyka i asymetrii w ekonomii, finansach, inżynierii czy meteorologii;
- konstrukcję modeli regresji kwantylowej, które pozwalają badać wpływ zmiennych objaśniających na różne części rozkładu zmiennej objaśnianej, a nie tylko na jego średnią.

W badaniach empirycznych estymatory kwantyli odgrywają kluczową rolę. Bardzo często wykorzystywane są w statystyce teoretycznej i stosowanej, ponieważ pozwalają w bezpośredni sposób opisywać strukturę rozkładu zmiennej losowej, niezależnie od jego postaci czy parametrów. W przeciwieństwie do klasycznych miar tendencji centralnej, takich jak np. średnia, kwantyle dostarczają informacji odpornych na wartości odstające i silnie skośne rozkłady, dzięki czemu stanowią istotne narzędzie w analizie danych o niestandardowym charakterze. Szczególnie ważne jest to, że kwantyle pozwalają wyodrębnić zachowanie ogo-

nów rozkładu, co ma znaczenie w wielu obszarach badań, takich jak ekonometria, analiza ryzyka, uczenie maszynowe czy teoria niezawodności.

Z punktu widzenia teorii estymacji, konstrukcja dobrego estymatora kwantyla jest niezbędna, aby móc wnioskować o nieznanym rozkładzie zmiennej losowej na podstawie próby. Estymatory pełnią funkcję pomostu między obserwowanymi danymi a ich teoretyczną reprezentacją probabilistyczną. Umożliwiają zatem nie tylko szacowanie samych kwantyli, ale również budowę narzędzi wtórnych, takich jak przedziały ufności, testy statystyczne czy nieliniowe metody regresji kwantylowej. Estymatory kwantyli są fundamentem wielu algorytmów analizy danych, a ich własności probabilistyczne – takie jak zgodność czy asymptotyczna normalność – stanowią podstawę rzetelnych wniosków statystycznych. Zrozumienie roli kwantyli i ich estymatorów jest kluczowe zarówno w teorii, jak i zastosowaniach statystyki.

Funkcje kwantylowe stanowią zatem ważne narzędzie łączące teorię rozkładów prawdopodobieństwa i praktyczne metody analizy danych. Pozwalają na opis zmienności i niepewności w sposób bardziej kompletny niż klasyczne podejście oparte wyłącznie na wartościach oczekiwanych i wariancjach. Ich elastyczność i łatwość interpretacji sprawiają, że stanowią one podstawę wielu współczesnych technik modelowania probabilistycznego i statystycznego.

Inspiracją do rozważań zawartych w niniejszej rozprawie była przełomowa praca Bahadura [2], w której autor wyraził kwantyl empiryczny za pomocą dystrybuanty empirycznej. Taka reprezentacja estymatora kwantyla okazuje się bardzo użyteczna w badaniach nad granicznymi własnościami kwantyla z próby, a także może znaleźć zastosowania jako narzędzie pomocne w konstrukcji innych estymatorów. We wspomnianej pracy autor badał wyłącznie przypadek niezależnych zmiennych losowych oraz ograniczył swoje rozważania do kwantyla z próby zdefiniowanego jako element o numerze $\lfloor np \rfloor + 1$ (gdzie $0 < p < 1$) w uporządkowanej próbie. Jednak w sytuacjach praktycznych często pojawia się potrzeba analizowania prób, które pochodzą z populacji o pewnej strukturze zależności. Dlatego na przestrzeni lat pojawiało się wiele prac dotyczących reprezentacji Bahadura m.in. dla takich zmiennych losowych jak zmienne losowe ujemnie stowarzyszone, α-mieszające, φ-mieszające, ψ-mieszające czy zmienne losowe typu NOD (*negatively orthant dependent*).

Warto również zauważyć, że w ostatnich latach wzrasta potrzeba modelowania procesów o charakterze technicznym i ekonomicznym, które zachodzą w warunkach niepewności. W takiej sytuacji klasyczne miary prawdopodobieństwa okazują się niewystarczające i stąd pojawia się potrzeba wykorzystania innych narzędzi lepiej opisujących dany problem. W najnowszej literaturze można zaobserwować coraz większą liczbę prac analizujących liczne problemy w tzw. przestrzeni z niepewnością, czyli przestrzeni podliniowej wartości oczekiwanej. Wiele pojęć, parametrów czy twierdzeń z klasycznej liniowej przestrzeni probabi-

listycznej znalazło już swoje odwzorowanie w przestrzeni podliniowej wartości oczekiwanej. Do takich pojęć zaliczamy również pojęcie kwantyla.

Jednym z dwóch celów tej rozprawy jest sformułowanie reprezentacji Bahadura dla

– ośmiu różnych sposobów zdefiniowania kwantyla z próby;
– czterech różnych rodzajów struktury zależności w próbie (ujemnego stowarzyszenia, φ-mieszania, ρ-mieszania oraz ρ^* mieszania);
– niezależnych zmiennych losowych w przestrzeni podliniowej wartości oczekiwanej.

Drugą inspiracją do powstania niniejszej rozprawy była praca Midhu i in. [32], w której autorzy zaproponowali wykorzystanie funkcji kwantylowych do aproksymacji i estymacji rozkładu sumy niezależnych, nieujemnych zmiennych losowych. Wyniki zaprezentowane w tej pracy znajdują zastosowania np. w teorii niezawodności. Jednak modelowanie wielu istotnych problemów praktycznych o charakterze losowym wymaga zastosowania innych działań niż suma. Np. gdy chcemy modelować dopasowanie wałka i cylindrycznej tulei w montażu przy masowej produkcji pewnego urządzenia interesuje nas rozkład różnicy $C - S$, gdzie C to losowa średnica wewnętrzna tulei, a S to średnica wałka. Podobną sytuację mamy gdy analizujemy rynek energii elektrycznej rozliczanej na giełdzie energii. Końcowa wartość transakcji jest iloczynem wolumenu sprzedanej energii W i ceny za jedną megawatogodzinę P. Zatem znajomość rozkładu zmiennej zdefiniowanej jako $W \cdot P$ jest w tym przypadku kluczowa. Z kolei znajomość rozkładu ilorazu dwóch zmiennych losowych jest przydatna do badania losowego współczynnika wydajności wykorzystywanego do porównywania skuteczności dwóch różnych metod dla tego samego zadania. Praca Mihdu i in. [32], choć bardzo ważna, nie może być wykorzystana do rozwiązywania takich problemów. Stąd pojawia się naturalna potrzeba skonstruowania analogicznych narzędzi dla innych funkcji zmiennych losowych.

Drugim celem tej rozprawy jest wyznaczenie wzorów aproksymujących dystrybuanty dla

– różnicy,
– iloczynu,
– ilorazu

niezależnych zmiennych losowych (nieujemnych lub odpowiednio dodatnich) i na jej podstawie zaproponowanie nowych estymatorów rozkładu (dystrybuanty) dla tych funkcji.

Rozdział 1

Wybrane pojęcia związane z kwantylami i strukturami zależności

W poniższym rozdziale zostaną omówione najważniejsze własności i twierdzenia związane z funkcjami kwantylowymi, które najczęściej wykorzystywane są w rozważaniach teoretycznych i praktycznych. Jak również przedstawione zostaną definicje wybranych struktur zależności wykorzystywane w dalszej części rozprawy.

1.1. Wybrane własności funkcji kwantylowych

Niech F będzie dystrybuantą zmiennej losowej X. Funkcja kwantylowa $Q\colon (0,1) \to \mathbb{R}$, związana z dystrybuantą F, zdefiniowana jest wzorem

$$Q(p) = \inf\{x \in \mathbb{R}\colon F(x) \geqslant p\}, \quad \text{dla} \quad p \in (0,1).$$

Funkcję kwantylową Q można również zdefiniować dla $p = 0$ oraz dla $p = 1$, przyjmując $Q(0) = -\infty$ oraz $Q(1) = \infty$. Wtedy funkcja Q jest określona na przedziale $[0,1]$ i przyjmuje wartości w $\mathbb{R} \cup \{-\infty, \infty\}$, tzn.

$$Q\colon [0,1] \to \mathbb{R} \cup \{-\infty, \infty\}$$

oraz

$$Q(0) = -\infty \;\Leftrightarrow\; F(x) > 0, \quad \forall x \in \mathbb{R},$$

$$Q(1) = +\infty \;\Leftrightarrow\; F(x) < 1, \quad \forall x \in \mathbb{R}.$$

Ponadto:

– jeżeli $Q(0) = m$, gdzie $m \in \mathbb{R}$, to X ma ograniczenie dolne, tj.

$$P(X < m) = 0 \quad \wedge \quad P(X \geqslant x) > 0, \qquad \forall x > m;$$

– jeżeli $Q(1) = M$, gdzie $M \in \mathbb{R}$, to X ma ograniczenie górne, tj.

$$P(X > M) = 0 \quad \wedge \quad P(X \geqslant x) > 0 \qquad \forall x < M.$$

Podsumowując, niezależnie od tego, czy wartości $Q(0)$ i $Q(1)$ są skończone, można zapisać:

$$P(X < Q(0)) = 0 \quad \text{oraz} \quad P(X \geqslant x) > 0 \text{ dla każdego } x > Q(0),$$

$$P(X > Q(1)) = 0 \quad \text{oraz} \quad P(X \geqslant x) > 0 \text{ dla każdego } x < Q(1).$$

Spośród różnych własności funkcji kwantylowych warto podkreślić dwie następujące:

$$P(X \leqslant Q(p)) \geqslant p \quad \text{oraz} \quad P(X \geqslant Q(p)) \geqslant 1-p,$$

i

$$P(X < Q(p)) \leqslant p \quad \text{oraz} \quad P(X > Q(p)) \leqslant 1-p.$$

W kolejnej części podane zostały wybrane twierdzenia dotyczące własności funkcji kwantylowych. We wszystkich niżej podanych twierdzeniach funkcja kwantylowa odpowiadająca dystrybuancie F jest oznaczana jako Q.

Twierdzenie 1.1. *(Dufour [15]) Niech F będzie dystrybuantą zmiennej losowej X. Wtedy funkcja kwantylowa Q związana z dystrybuantą F ma następujące własności:*

(a) *dla każdego $p \in (0,1)$ istnieje dokładnie jedna liczba rzeczywista a taka, że $a = Q(p)$;*

(b) *$a = Q(p)$, $p \in (0,1)$, wtedy i tylko wtedy, gdy spełnione są dwa warunki:*

(i) *$F(a) \geqslant p$,*

(ii) *$x < a \implies F(x) < p$;*

(c) *$Q(p) = \inf\{x : P(X < x) \leqslant p \leqslant P(X \leqslant x)\}, \quad 0 < p < 1$;*

(d) *$Q(p) = \sup\{x : F(x) < p\}, \quad 0 < p < 1$;*

(e) *funkcja Q jest niemalejąca i lewostronnie ciągła;*

(f) *$F(x) \geqslant p \;\Leftrightarrow\; x \geqslant Q(p)$, dla każdego $x \in \mathbb{R}$ i $p \in (0,1)$;*

(g) *$F(x) < p \;\Leftrightarrow\; x < Q(p)$, dla każdego $x \in \mathbb{R}$ i $p \in (0,1)$;*

(h) *$F((Q(p))^-) \leqslant p \leqslant F(Q(p))$, dla każdego $0 < p < 1$;*

(i) *$Q(F(x)) \leqslant x \leqslant Q((F(x))^+)$, dla każdego $x \in \mathbb{R}$;*

(j) *jeśli F jest ciągła w punkcie $x = Q(p)$, $p \in (0,1)$, to $F(Q(p)) = p$;*

(k) *jeśli Q jest ciągła w punkcie $p = F(x)$, $x \in \mathbb{R}$, to $Q(F(x)) = x$;*

(l) *dla $p \in (0,1)$, $F(Q(p)) = p \;\Leftrightarrow\; p \in F(\mathbb{R})$ (gdzie $F(\mathbb{R})$ oznacza zbiór wartości funkcji F);*

(m)

$$\begin{aligned} F(Q(p)) = p \text{ dla każdego } p \in (0,1) \quad &\Longleftrightarrow \quad (0,1) \subseteq F(\mathbb{R}) \\ &\Longleftrightarrow \quad F \text{ jest ciągła} \\ &\Longleftrightarrow \quad Q \text{ jest ściśle rosnąca}; \end{aligned}$$

(n) *dla dowolnego $x \in \mathbb{R}$, $Q(F(x)) = x \;\Leftrightarrow\; F(x-\varepsilon) < F(x)$ dla wszystkich $\varepsilon > 0$;*

(o) dla dowolnego $x \in \mathbb{R}$, $P(X = x) > 0 \;\Rightarrow\; Q(F(x)) = x$;

(p)

$$Q(F(x)) = x \text{ dla każdego } x \in \mathbb{R} \iff F \text{ jest ściśle rosnąca} \iff Q \text{ jest ciągła};$$

(r) funkcja F jest ciągła i ściśle rosnąca wtedy i tylko wtedy, gdy funkcja Q jest ciągła i ściśle rosnąca;

(s) $Q(F(Q(p))) = Q(p)$, *dla każdego* $p \in (0,1)$;

(t) $F(Q(F(x))) = F(x)$, *dla każdego* $x \in \mathbb{R}$.

Następne twierdzenie podaje warunki, jakie musi spełniać funkcja, aby była funkcją kwantylową pewnego rozkładu prawdopodobieństwa.

Twierdzenie 1.2. *(Dufour [15]) Jeśli G jest rzeczywistą, niemalejącą oraz lewostronnie ciągłą funkcją określoną na przedziale* $(0,1)$*, to istnieje dokładnie jedna dystrybuanta F taka, że*

$$G(u) = Q(u),$$

czyli G jest funkcją kwantylową odpowiadającą dystrybuancie F.

Warunki różniczkowalności funkcji kwantylowej podaje następujące twierdzenie.

Twierdzenie 1.3. *(Dufour [15]) Niech F będzie dystrybuantą pewnej zmiennej losowej. Jeżeli F ma dodatnią i ciągłą gęstość f w otoczeniu punktu* $Q(u_0)$*, gdzie* $0 < u_0 < 1$*, to pochodna funkcji kwantylowej Q względem u istnieje w punkcie* $u = u_0$ *i spełniona jest zależność*

$$\frac{dQ(u)}{du}\Big|_{u=u_0} = \frac{1}{f(Q(u_0))}.$$

Kolejne twierdzenie dotyczy funkcji kwantylowej zmiennej losowej, która jest funkcją innej zmiennej losowej.

Twierdzenie 1.4. *(Dufour [15]) Niech X będzie rzeczywistą zmienną losową zdefiniowaną za pomocą dystrybuanty* F_X*. Niech g będzie funkcją zmiennej rzeczywistej lewostronnie ciągłą. Wówczas:*

$$Q_{g(X)}(p) = g\big(Q_X(p)\big), \quad \text{dla wszystkich } 0 < p < 1,$$

gdzie

$$F_{g(X)}(x) = \mathbb{P}(g(X) \leqslant x) \quad \text{oraz} \quad Q_{g(X)}(p) = \inf\{x : F_{g(X)}(x) \geqslant p\}.$$

Poniższy wniosek prezentuje szczególny przypadek Twierdzenia 1.4.

Wniosek 1.1. Niech X będzie rzeczywistą zmienną losową o rozkładzie zdefiniowanym za pomocą dystrybuanty F_X oraz niech a i b będą dwoma liczbami rzeczywistymi. Jeśli $a > 0$, to

$$Q_{aX+b}(p) = aQ_X(p) + b, \quad \text{dla } 0 < p < 1.$$

Twierdzenia 1.5 oraz 1.6 regulują niektóre wybrane działania na funkcjach kwantylowych dostarczając narzędzi do dalszych rozważań w tej pracy.

Twierdzenie 1.5. *(Dufour [15]) Niech X będzie rzeczywistą zmienną losową o rozkładzie zdefiniowanym za pomocą dystrybuanty F_X, niech F_0 będzie dystrybuantą oraz $U = F_0(X)$. Wówczas, dla każdego $u \in (0,1)$*

$$U \leqslant u \iff F_0(X) \leqslant u \iff X \leqslant Q_0(u)$$

oraz

$$\mathbb{P}(U \leqslant u) = \mathbb{P}(X \leqslant Q_0(u)) = F_X(Q_0(u)).$$

Twierdzenie 1.6. *(Dufour [15]) Niech F będzie dystrybuantą, a U zmienną losową o dystrybuancie F_0. Ponadto załóżmy, że $F_0(0) = 0$ oraz $F_0(1) = 1$. Jeśli $X = Q(U)$, to dla każdego $x \in \mathbb{R}$ mamy*

$$X \leqslant x \iff Q(U) \leqslant x \iff U \leqslant F(x)$$

lub równoważnie

$$I\{X \leqslant x\} = I\{Q(U) \leqslant x\} = I\{U \leqslant F(x)\}$$

oraz

$$P(X \leqslant x) = P(Q(U) \leqslant x) = P(U \leqslant F(x)) = F_0(F(x)),$$
$$P(I\{X < x\}) = P(I\{Q(U) < x\}) = P(I\{U \leqslant F(x^-)\})$$

oraz

$$P(X < x) = P(Q(U) < x) = P(U \leqslant F(x^-)).$$

W szczególności, jeśli U ma rozkład jednostajny na przedziale $(0,1)$, to F jest dystrybuantą zmiennej losowej $Q(U)$:

$$P(Q(U) \leqslant x) = P(X \leqslant x) = P(U \leqslant F(x)) = F(x), \quad \forall x \in \mathbb{R}.$$

1.2. Estymatory kwantyla

Estymatory kwantyli odgrywają kluczową rolę w finansach, modelowaniu i statystyce. W zastosowaniach praktycznych estymatory kwantyli mają fundamentalne znaczenie, co dostrzeżono już stosunkowo wcześnie, m.in. w pracy Galtona [17]. Przez lata w literaturze opisywano liczne estymatory kwantyli. Część z nich została opracowana dla specyficznych rozkładów, podczas gdy inne zaprojektowano jako tzw. „wolne od rozkładu", czyli estymatory nieparametryczne.

Niech $(X_{(1)}, X_{(2)}, ..., X_{(n)})$ będzie uporządkowaną próbą odpowiadającą próbie $(X_1, ..., X_n)$ oraz niech $\lfloor x \rfloor$ oznacza część całkowitą liczby x. Poniżej przedstawiono listę ośmiu estymatorów nieparametrycznych kwantyla zaprezentowanych w pracy Dielmana i in. [8], które będą analizowane w dalszej części rozprawy:

— $Q_n^1(p) = (0.5 + \lfloor np + 0.5 \rfloor - np)X_{(\lfloor np+0.5 \rfloor)} + (0.5 - \lfloor np + 0.5 \rfloor + np)X_{(\lfloor np+0.5 \rfloor+1)}$, gdzie $0.5 \leqslant np \leqslant (n-0.5)$,

— $Q_n^2(p) = (1 - (np - \lfloor np \rfloor))X_{(\lfloor np \rfloor)} + (np - \lfloor np \rfloor)X_{(\lfloor np \rfloor+1)}$,

— $Q_n^3(p) = X_{(\lfloor np \rfloor)}$,

— $Q_n^4(p) = X_{(\lfloor np \rfloor+1)}$,

— $Q_n^5(p) = \begin{cases} X_{(\lfloor np \rfloor)}, & \text{if } np - \lfloor np \rfloor < 0.5 \\ X_{(\lfloor np \rfloor+1)}, & \text{if } np - \lfloor np \rfloor \geqslant 0.5 \end{cases}$,

— $Q_n^6(p) = \begin{cases} X_{(\lfloor np \rfloor)}, & \text{if } np - \lfloor np \rfloor = 0 \\ X_{(\lfloor np \rfloor+1)}, & \text{if } np - \lfloor np \rfloor > 0 \end{cases}$,

— $Q_n^7(p) = (1 - ((n+1)p - \lfloor (n+1)p \rfloor))X_{(\lfloor (n+1)p \rfloor)} + ((n+1)p - \lfloor (n+1)p \rfloor)X_{(\lfloor (n+1)p \rfloor+1)}$,

— $Q_n^8(p) = \begin{cases} \frac{X_{(\lfloor np \rfloor)} + X_{(\lfloor np \rfloor+1)}}{2}, & \text{if } np - \lfloor np \rfloor = 0 \\ X_{(\lfloor np \rfloor+1)}, & \text{if } np - \lfloor np \rfloor > 0 \end{cases}$.

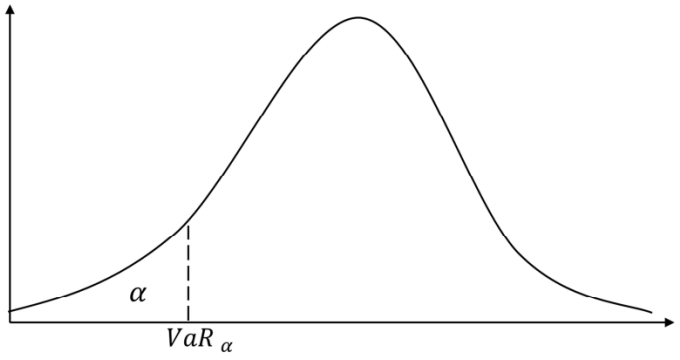
Uwaga 1.1. Dla estymatora E_1 z warunku $0.5 \leqslant np \leqslant (n-0.5)$ wynika $n \geqslant \max\left\{\frac{0.5}{1-p}, \frac{1}{2p}\right\}$.

Wybór odpowiedniego estymatora ma kluczowe znaczenie, ponieważ wpływa bezpośrednio na jakość i wiarygodność uzyskiwanych wyników. Jest to szczególnie istotne w przypadku estymacji kwantyli, które znajdują zastosowanie np. do wyznaczania miary ryzyka *Value at Risk* (VaR). Miara ta określa taką wartośc straty, która z zadanym prawodpobieństwem nie zostanie przekroczona w określonym horyzoncie czasowym.

Dla dodatniej, bliskiej zeru wartości p miarę *Value at Risk* definiuje się jako

$$\mathrm{VaR}_p = \inf\{x : F(x) \geqslant p\}.$$

Miara ta jest w istocie odpowiednim kwantylem rozkładu strat i jest powszechnie stosowana w finansach do oceny potencjalnych strat portfela inwestycyjnego. Na Rysunku 1.1 przedstawiono graficzną interpretację miary *VaR*.



Rysunek 1.1: Interpretacja graficzna miary *VaR*. **Źródło:** [7]

Na przestrzeni lat opracowano liczne techniki estymacji kwantyli. Na przykład Harrell i Davis [20] zaproponowali estymator kwantyla będący ważoną kombinacją wszystkich sta-

tystyk pozycyjnych. Zelterman [56] zaproponował gładki nieparametryczny estymator funkcji kwantylowej oparty na metodach jądrowych. Takeuchi i in. [45] przedstawili nieparametryczną metodę estymacji kwantyli warunkowych z wykorzystaniem metod jądrowych. Z kolei Han i in. [19] przedstawili metodę estymacji kwantyli w przypadku niekompletnych danych, umożliwiającą uwzględnienie różnych mechanizmów występowania brakujących obserwacji.

W ostatnich latach zagadnienie estymacji kwantyli rozwija się wyjątkowo dynamicznie, głównie ze względu na rosnącą złożoność analizowanych danych. Współczesne zastosowania – od finansów po analizę dużych zbiorów danych i modele uczenia maszynowego – wymagają narzędzi, które potrafią precyzyjnie opisywać skrajne zachowania i lokalne cechy rozkładów. Badania nad własnościami estymatorów, takich jak ich efektywność czy stabilność obliczeniowa stają się zatem niezbędnym elementem teorii estymacji. Dodatkowo badania nad granicznymi własnościami estymatorów kwantyli okazują się kluczowe przy analizie dużych prób.

1.3. Zależne zmienne losowe

W tej części rozdziału wprowadzone zostaną definicje czterech rodzajów zależności zmiennych losowych, które będą rozważane w dalszej części rozprawy. Podane zostaną również użyteczne lematy dotyczące danych zmiennych losowych.

Definicja 1.1. (Joag-Dev i Proschan [22]) Skończony zbiór zmiennych losowych $\{X_i,\, 1 \leqslant i \leqslant n\}$ nazywamy *ujemnie stowarzyszonym* (ang. *negatively associated*, NA), jeśli dla każdej pary rozłącznych podzbiorów A i B zbioru $\{1,2,\ldots,n\}$ zachodzi

$$\mathrm{Cov}(f_1(X_i, i \in A), f_2(X_j, j \in B)) \leqslant 0,$$

gdzie funkcje f_1 i f_2 są niemalejące względem każdej ze swoich współrzędnych. Nieskończony zbiór zmiennych losowych nazywamy NA, jeśli każda jego skończona podrodzina jest NA.

Uwaga 1.2. Funkcje niemalejące zdefiniowane na rozłącznych podzbiorach zbioru zmiennych ujemnie stowarzyszonych są również ujemnie stowarzyszone.

Definicja 1.2. (Rozanov i Volkonski [35]) Ciąg zmiennych losowych $\{X_n, n \geqslant 1\}$ nazywamy φ-mieszającym, jeśli

$$\varphi(n) = \sup_{m \geqslant 1} \sup_{A \in \mathscr{F}_1^m, B \in \mathscr{F}_{m+n}^{\infty}, P(A)>0} |P(B|A) - P(B)| \to 0 \text{ dla } n \to \infty,$$

gdzie $\mathscr{F}_n^m = \sigma(X_i, n \leqslant i \leqslant m)$.

Definicja 1.3. (Shao [42]) Ciąg zmiennych losowych $\{X_n, n \geqslant 1\}$ nazywamy ρ-mieszającym, jeśli

$$\rho(n) = \sup_{k\geqslant 1, X\in L^2(\mathscr{F}_1^k), Y\in L^2(\mathscr{F}_{k+n}^{\infty})} \frac{|\mathrm{Cov}(X,Y)|}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \to 0,$$

gdy $n \to \infty$, gdzie $\mathscr{F}_n^m = \sigma(X_i, n \leqslant i \leqslant m)$ oraz $L^2(\cdot)$ jest zbiorem funkcji o wartościach rzeczywistych kwadratowo całkowalnych.

Definicja 1.4. (Chen i Sung [6]) Ciąg zmiennych losowych $\{X_n, n \geqslant 1\}$ nazywamy ρ^*-mieszającym, jeśli

$$\rho^*(n) = \sup_{\substack{S,T\subset\mathbb{N} \\ \mathrm{dist}(S,T)\geqslant n}} \rho(S,T) \to 0,$$

dla $n \to \infty$, gdzie

$$\rho(S,T) = \sup_{X\in L^2(\sigma(S)), Y\in L^2(\sigma(T))} \left\{ \frac{|Cov(X,Y)|}{\sqrt{Var(X)Var(Y)}} \right\},$$

$\mathrm{dist}(S,T) = \min_{i\in S, j\in T} |j-i|$ i $\sigma(S)$, $\sigma(T)$ są σ-ciałami generowanymi przez, odpowiednio, $\{X_i, i \in S\}$ oraz $\{X_j, j \in T\}$.

Uwaga 1.3. Zauważmy, że niemalejące funkcje zdefiniowane na rozłącznych podzbiorach ρ^*-mieszającego ciała $\{X_k, k \in N^d\}$ ze współczynnikami mieszającymi $\rho^*(s)$ są także ρ^*-mieszające, ze współczynnikami nie większymi niż $\rho^*(s)$.

Wymienione typy zależności zmiennych losowych były badane przez wielu autorów zarówno np. pod kątem twierdzeń granicznych czy nierówności momentowych. Bardzo ważną grupą nierówności momentowych jest tzw. nierówność typu Rosenthala:

$$E\left\{ \max_{1\leqslant m\leqslant n} \Big| \sum_{k=1}^{m} X_k \Big|^q \right\} \leqslant C_q \left\{ \sum_{k=1}^{n} E|X_k|^q + \Big(\sum_{k=1}^{n} EX_k^2 \Big)^{q/2} \right\}, \tag{1.1}$$

gdzie $C_q > 0$, $q \geqslant 2$, $EX_n = 0$, $E|X_n|^q < \infty$ dla każdego $n \geqslant 1$.

Uwaga 1.4. Nierówność (1.1) jest prawdziwa dla zmiennych losowych ujemnie stowarzyszonych (Shao [43]), dla ρ^*-mieszających zmiennych losowych (Utev i Peligrad [46]) oraz dla φ-mieszających zmiennych losowych (Wu i in. [48]).

Dla zmiennych ρ-mieszających istnieje analogiczna nierówność podana w pracy Shao [42].

Lemat 1.1. *(Shao [42]) Niech $\{X_n, n \geqslant 1\}$ będzie ρ- mieszającym ciągiem zmiennych losowych. Załóżmy, że $EX_i = 0$, $E|X_n|^q < \infty$ oraz*

$$\sum_{n=1}^{\infty} \rho^{\frac{2}{q}}(2^n) < \infty, \quad q \geqslant 2.$$

Wtedy istnieje dodatnia stała $K = K(q, \rho(\cdot))$ *zależna jedynie od* q *i* $\rho(\cdot)$ *taka, że dla dowolnego* $k \geqslant 0$, $n \geqslant 1$,

$$E\left\{ \max_{1 \leqslant n} \left| \sum_{i=k+1}^{k+n} X_i \right|^q \right\} \leqslant K\left\{ n \max_{k<i \leqslant k+n} E|X_i|^q + \left(n \max_{k<i \leqslant k+n} EX_i^2 \right)^{\frac{q}{2}} \right\}.$$

Wspomniane wyżej nierówności odgrywają istotną rolę w rozważaniach w dalszej części pracy.

Rozdział 2

Reprezentacja Bahadura dla kwantyli z próby

W poniższym rozdziale wprowadzone zostanie pojęcie reprezentacji Bahadura estymatora kwantyla, która pokazuje jak empiryczny kwantyl różni się od rzeczywistego kwantyla w populacji.

Reprezentacja Bahadura, będąca fundamentalnym pojęciem w teorii statystyki, wywodzi się z przełomowej pracy Bahadura [2] z 1966 roku. Praca ta wprowadziła ramy teoretyczne, które umożliwiły nowe spojrzenie na kwantyle z próby poprzez pryzmat empirycznej dystrybuanty dla próby prostej. Bahadur [2] wykazał poniższe twierdzenie, w którym rozważał reprezentację estymatora kwantyla zdefiniowanego wzorem $Q_n^4(p) = X_{(\lfloor np \rfloor + 1)}$.

Twierdzenie 2.1. *(Bahadur [2]) Niech* $0 < p < 1$ *oraz* $\{X_n,\ n \geqslant 1\}$ *będzie ciągiem niezależnych zmiennych losowych o takim samym rozkładzie zdefiniowanych za pomocą dystrybuanty* F*. Załóżmy, że istnieje przynajmniej druga pochodna dystrybuanty* F *w pewnym sąsiedztwie* $Q(p)$ *oraz załóżmy, że* $F'(Q(p)) = f(Q(p)) > 0$*. Wtedy*

$$Q_n^4(p) = Q(p) - \frac{F_n(Q(p)) - p}{f(Q(p))} + O\Big(n^{-\frac{3}{4}} \log n\Big),$$

gdzie

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I[X_i \leqslant x], \quad x \in \mathbb{R},\ n \geqslant 1.$$

Reprezentacja Bahadura okazuje się użytecznym narzędziem do badania granicznych własności estymatora kwantyla oraz przy wyznaczaniu nieparametrycznego estymatora dystrybuanty wybranych zmiennych losowych. Jak wspomniano wcześniej, Bahadur [2] w swojej pracy badał jedynie przypadek próby prostej, tzn. gdy pobrana próba jest niezależna i elementy próby są pobrane z populacji o takim samym rozkładzie. W rzeczywistości, gdy badane są pewne cechy na podstawie pobranej próby, okazuje się, że istnieje w próbie pewna struktura zależności. Stąd w kolejnych latach powstawały liczne prace omawiające zagadnienie reprezentacji Bahadura, ale dla zmiennych losowych zależnych.

Sen [40] przedstawił reprezentację dla estymatora kwantyla analogiczną do reprezentacji Bahadura, ale dla zmiennych losowych φ-mieszających. Yoshihara [54] rozważał reprezentację Bahadura również dla przypadku zmiennych losowych φ-mieszających rozszerzając

wyniki przedstawione przez Sena [40]. Sun [44] wniósł swój wkład wyznaczając reprezentacje dla kwantyla z próby w przypadku zmiennych α-mieszających. Ling [27] przedstawił reprezentację analogiczną do tej zaproponowanej przez Bahadura dla zmiennych ujemnie stowarzyszonych (NA) przy pewnych ogólnych założeniach. Xing i Yang [49] przedstawili reprezentację Bahadura dla zmiennych NA i otrzymali szybszą zbieżność niż zbieżność zaprezentowana przez Linga [27]. Li i in. [26] w swojej pracy rozważali reprezentację Bahadura dla ciągu zmiennych losowych typu NOD. Wang i in. [47] skierowali uwagę na zmienne losowe α-mieszające, opracowując reprezentacje Bahadura dostosowaną do tej formy zależności. Xu i in. [51] przestawili reprezentację Bahadura dla zmiennych NA i poprawili wyniki zaprezentowane przez Xinga i Yanga [49]. Zhang i in. [57] również badali α-mieszające ciągi zmiennych losowych i przy różnych założeniach otrzymali reprezentację Bahadura z różnymi rzędami zbieżności. Xing i Yang [50] przedstawili reprezentację Bahadura dla zmiennych losowych ψ-mieszających z pewnym założeniem na współczynniki mieszające. Douge [10] przedstawił reprezentację Bahadura dla zmiennych losowych dodatnio stowarzyszonych. Yang i in. [53] rozważali reprezentację Bahadura dla zmiennych losowych φ-mieszających. Wu i in. [48] również przedstawili reprezentację Bahadura dla zmiennych losowych φ-mieszających. Dużą zaletą pracy Wu i in. [48] był brak założeń na tempo zaniku współczynników mieszających. Wszystkie wymienione wyniki potwierdzają dynamiczny rozwój badań nad reprezentacją Bahadura dla przypadku zależnych zmiennych losowych. Na uwagę zasługuje fakt, że reprezentacja Bahadura znalazła również zastosowanie w analizie innych estymatorów statystycznych. Dla przykładu Koo i in. [24] wyprowadzili reprezentację Bahadura dla estymatora liniowej maszyny wektorów nośnych, wykorzystując ją do analizy jego własności asymptotycznych. Z kolei Hu [21] wyprowadził reprezentację Bahadura dla M-estymatorów w ogólnych modelach regresji liniowej z zależnymi błędami oraz wykorzystał ją do zbadania asymptotycznej normalności i szybkości silnej zbieżności estymatorów. Warto podkreślić, że pomimo tego, że zagadnienie reprezentacji Bahadura było przedmiotem badań licznych prac, to wciąż pojawiają się nowe pozycje w literaturze ukierunkowane na to zagadnienie, co świadczy o istotności przedstawianego problemu.

2.1. Silna zgodność estymatora kwantyla i reprezentacja Bahadura dla zależnych zmiennych losowych

W niniejszej części udowodniona zostanie silna zgodność estymatora kwantyla oraz na jej podstawie wyznaczona reprezentacja Bahadura. Rozważany będzie przypadek zmiennych losowych ujemnie stowarzyszonych, φ-mieszających, ρ^*-mieszających oraz ρ-mieszających. Przedstawione wyniki dotyczą estymatorów $Q_n^1(p)$, $Q_n^2(p)$, $Q_n^3(p)$, $Q_n^4(p)$, $Q_n^5(p)$, $Q_n^6(p)$, $Q_n^7(p)$ i $Q_n^8(p)$. Szczegółowe rozważania zostaną przeprowadzone dla estymatora $Q_n^1(p)$.

Twierdzenie 2.2. *(Dudek i Kuczmaszewska [12]) Niech $\{X_n, n \geqslant 1\}$ będzie ciągiem zmiennych losowych o jednej z następujących struktur zależności:*

a) ujemnie stowarzyszone,

b) φ-mieszające przy warunku $\sum_{n=1}^{\infty} \varphi(2^n) < \infty$,

c) ρ^*-mieszające,

d) ρ-mieszające przy warunku $\sum_{n=1}^{\infty} \rho^{\frac{2}{q}}(2^n) < \infty$, dla $q > \frac{1}{\delta}$ oraz $0 < \delta < \frac{1}{2}$.

Niech $\{X_n, n \geqslant 1\}$ będzie ciągiem zmiennych losowych o jednakowym rozkładzie zdefiniowanym dystrybuantą F oraz niech $Q(p)$ będzie kwantylem rzędu p. Załóżmy, że F ma dodatnią i ciągłą funkcję gęstości f w pewnym sąsiedztwie $\mathfrak{D}_p$ punktu $Q(p)$ spełniającą warunek

$$0 < \sup\{f(x); x \in \mathfrak{D}_p\} < \infty.$$

Ponadto zakładamy, że f' w pewnym sąsiedztwie $\mathfrak{D}_p$ punktu $Q(p)$ spełnia warunek

$$|f'(x)| < M, \quad x \in \mathfrak{D}_p, \quad M \in \mathbb{R}. \tag{2.1}$$

Wtedy, dla dowolnej $0 < \delta < \frac{1}{2}$

$$P\Big(Q_n^i(p) - Q(p) = o(n^{-\frac{1}{2}+\delta}),\ \ gdy\ \ n \to \infty\Big) = 1, \quad gdzie \quad i = \{1,2,3,...,8\}. \tag{2.2}$$

Dowód. Rozważmy estymator $Q_n^1(p)$. Zauważmy, że

$$X_{\lfloor np+0.5 \rfloor} \leqslant Q_n^1(p) \leqslant X_{\lfloor np+0.5 \rfloor + 1}. \tag{2.3}$$

Wtedy $\forall\ \varepsilon > 0$

$$\left\{|Q_n^1(p) - Q(p)| \geqslant \varepsilon n^{-\frac{1}{2}+\delta}\right\} = \left\{Q_n^1(p) \geqslant Q(p) + \varepsilon n^{-\frac{1}{2}+\delta}\right\}$$

$$\cup \left\{Q_n^1(p) \leqslant Q(p) - \varepsilon n^{-\frac{1}{2}+\delta}\right\} = A_n^1 \cup A_n^2.$$

Niech $\xi_{ni} = I(X_i \leqslant Q(p) + \varepsilon n^{-\frac{1}{2}+\delta}) - F(Q(p) + \varepsilon n^{-\frac{1}{2}+\delta})$ dla $1 \leqslant i \leqslant n$. Stąd mamy

$$A_n^1 = \left\{Q_n^1(p) \geqslant Q(p) + \varepsilon n^{-\frac{1}{2}+\delta}\right\} \subset \left\{\sum_{i=1}^{n} I(X_i \leqslant Q(p) + \varepsilon n^{-\frac{1}{2}+\delta}) < \lfloor np + 0,5 \rfloor + 1\right\}$$

$$= \left\{\sum_{i=1}^{n} \xi_{ni} < \lfloor np + 0,5 \rfloor + 1 - nF(Q(p) + \varepsilon n^{-\frac{1}{2}+\delta})\right\}. \tag{2.4}$$

Wykorzystując rozwinięcie Taylora dostajemy

$$F(Q(p) + \varepsilon n^{-\frac{1}{2}+\delta}) = p + f(Q(p))\varepsilon n^{-\frac{1}{2}+\delta} + o(n^{-\frac{1}{2}+\delta}).$$

Stąd otrzymujemy, że istnieje pewna stała $c(\varepsilon) > 0$ zależna od $\varepsilon > 0$, taka że dla wystarczająco dużych n mamy

$$\left\{ \sum_{i=1}^{n} \xi_{ni} < \lfloor np + 0,5 \rfloor + 1 - np - \varepsilon f(Q(p)) n^{\frac{1}{2}+\delta} + o(n^{\frac{1}{2}+\delta}) \right\} \subset \left\{ \sum_{i=1}^{n} \xi_{ni} < -c(\varepsilon) n^{\frac{1}{2}+\delta} \right\}. \tag{2.5}$$

Na podstawie (2.4),(2.5), nierówności Markova oraz (1.1) (dla zmiennych losowych NA, φ oraz ρ^*-mieszających) lub Lematu 1.1 (dla ρ-mieszających zmiennych losowych), dla $r > \frac{1}{\delta}$, gdzie $0 < \delta < \frac{1}{2}$, dostajemy następujące oszacowanie

$$\begin{aligned}
&\sum_{n=1}^{\infty} P\left(Q_n^1(p) \geqslant Q(p) + \varepsilon n^{-\frac{1}{2}+\delta} \right) \leqslant \sum_{n=1}^{\infty} P\left(\sum_{i=1}^{n} \xi_{ni} < -c(\varepsilon) n^{\frac{1}{2}+\delta} \right) \\
&\leqslant \sum_{n=1}^{\infty} P\left(\left| \sum_{i=1}^{n} \xi_{ni} \right| > c(\varepsilon) n^{\frac{1}{2}+\delta} \right) \leqslant C \sum_{n=1}^{\infty} n^{-(\frac{1}{2}+\delta)r} E \left| \sum_{i=1}^{n} \xi_{ni} \right|^r \\
&\leqslant C \sum_{n=1}^{\infty} n^{-(\frac{1}{2}+\delta)r} \left[(nE\xi_{n1}^2)^{\frac{r}{2}} + nE|\xi_{n1}|^r \right] \leqslant C \sum_{n=1}^{\infty} n^{-\delta r} < \infty.
\end{aligned}$$

Przeprowadzając analogiczne rozważania dla A_n^2 dostajemy, że $\sum_{n=1}^{\infty} P\left[Q_n^1(p) \leqslant Q(p) - \varepsilon n^{-\frac{1}{2}+\delta} \right] < \infty$. Na podstawie lematu Borela-Cantelliego otrzymujemy tezę (2.2) dla estymatora $Q_n^1(p)$.

Zauważmy, że analogiczne wyniki można łatwo otrzymać dla poniższych estymatorów:

- $Q_n^2(p)$ – wykorzystując oszacowanie $X_{(\lfloor np \rfloor)} \leqslant Q_n^2(p) \leqslant X_{(\lfloor np \rfloor + 1)}$,
- $Q_n^3(p)$ – rachunki analogiczne jak w przypadku $Q_n^4(p)$,
- $Q_n^4(p)$ – rachunki przedstawione zostały w pracy Dudek i Kuczmaszewskiej [11],
- $Q_n^5(p)$ i $Q_n^6(p)$ – rachunki analogiczne jak w przypadku E_3 oraz $Q_n^4(p)$, ponieważ dowód nie zależy od wartości $np - \lfloor np \rfloor$,
- $Q_n^7(p)$ – wykorzystując oszacowanie $X_{(\lfloor (n+1)p \rfloor))} \leqslant Q_n^7(p) \leqslant X_{(\lfloor (n+1)p \rfloor + 1))}$,
- $Q_n^8(p)$ – wykorzystując fakt, że estymator jest kombinacją estymatorów $Q_n^4(p)$ oraz $Q_n^3(p)$.

□

Dwa następne twierdzenia będą stanowiły istotne narzędzia do wyznaczenia reprezentacji Bahadura.

Twierdzenie 2.3. *(Dudek i Kuczmaszewska [12]) Niech $\{X_n, n \geqslant 1\}$ będzie ciągiem zmiennych losowych o jednej z następujących struktur zależności:*

a) ujemnie stowarzyszone (NA)

b) ρ^*-mieszające.

Niech $\{X_n, n \geqslant 1\}$ będzie ciągiem zmiennych losowych o jednakowym rozkładzie zdefiniowanym dystrybuantą F oraz niech $Q(p)$ będzie kwantylem rzędu p. Załóżmy, że F ma dodatnią ciągłą funkcję gęstości f w pewnym sąsiedztwie $\mathfrak{D}_p$ punktu $Q(p)$ spełniającą warunek

$$0 < \sup\{f(x); x \in \mathfrak{D}_p\} < \infty. \tag{2.6}$$

Wtedy dla dowolnej $\delta > 0$

$$P\left(\sup_{x\in\mathfrak{I}_n} |F_n(x) - F(x) - (F_n(Q(p)) - p)| = O(n^{-\frac{1}{2}+\delta}), \quad n \to \infty\right) = 1,$$

gdzie $\mathfrak{I}_n = \left[Q(p) - c_0 n^{-\frac{1}{4}+\delta}, Q(p) + c_0 n^{-\frac{1}{4}+\delta}\right]$ dla pewnego $c_0 > 0$.

Dowód. Niech $\{a_n, n \geqslant 1\}$ oraz $\{b_n, n \geqslant 1\}$ będą dwoma ciągami zdefiniowanymi następująco

$$a_n = c_0 n^{-\frac{1}{4}+\delta} \quad \text{dla pewnego } \ c_0 > 0, \quad b_n = \lfloor n^{\frac{1}{4}} \rfloor + 1$$

oraz niech

$$G_n(x) = F_n(x) - F_n(Q(p)) - F(x) + p.$$

Dla dowolnego $n \in \mathbb{N}$ oraz całkowitego j wprowadźmy następujące oznaczenia

$$\eta_{j,n} = Q(p) + j a_n b_n^{-1}, \quad \alpha_{j,n} = F(\eta_{j+1,n}) - F(\eta_{j,n}) \quad \text{oraz} \quad \mathfrak{J}_{j,n} = [\eta_{j,n}, \eta_{j+1,n}].$$

Z faktu, że F_n oraz F są niemalejące, dla $x \in \mathfrak{J}_{j,n}$, dostajemy zależności

$$G_n(x) \leqslant F_n(\eta_{j+1,n}) - F_n(Q(p)) - F(\eta_{j,n}) + p = G_n(\eta_{j+1,n}) + \alpha_{j,n}$$

oraz

$$G_n(x) \geqslant F_n(\eta_{j,n}) - F_n(Q(p)) - F(\eta_{j+1,n}) + p = G_n(\eta_{j,n}) - \alpha_{j,n}.$$

Stąd mamy, że

$$\sup_{x\in\mathfrak{I}_n} |F_n(x) - F(x) - (F_n(Q(p)) - p)| \leqslant \max_{-b_n\leqslant j\leqslant b_n} \{|G_n(\eta_{j,n})|\} + \max_{-b_n\leqslant j\leqslant b_n-1} \{\alpha_{j,n}\}.$$

Wykorzystując twierdzenie o wartości średniej oraz (2.6) dostajemy

$$\alpha_{j,n} = F(\eta_{j+1,n}) - F(\eta_{j,n}) \leqslant C(\eta_{j+1,n} - \eta_{j,n}) = C a_n b_n^{-1} \leqslant C n^{-\frac{1}{2}+\delta}.$$

Zatem

$$\sum_{n=1}^{\infty} P\left(\sup_{x\in\mathfrak{I}_n} |F_n(x) - F(x) - (F_n(Q(p)) - p)| \geqslant c_0 n^{-\frac{1}{2}+\delta}\right)$$

$$\leqslant C \sum_{n=1}^{\infty} P\left(\max_{-b_n\leqslant j\leqslant b_n} |G_n(\eta_{j,n})| \geqslant \frac{c_0}{2} n^{-\frac{1}{2}+\delta}\right).$$

Zauważmy, że

$$G_n(\eta_{j,n}) = F_n(\eta_{j,n}) - F_n(Q(p)) - F(\eta_{j,n}) + p = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{Q(p)} - Y_i^{(j,n)}\right),$$

gdzie $Y_i^{Q(p)} = E(I[X_i \leqslant Q(p)]) - I[X_i \leqslant Q(p)]$ oraz $Y_i^{(j,n)} = E(I[X_i \leqslant \eta_{j,n}]) - I[X_i \leqslant \eta_{j,n}]$, $-b_n \leqslant j \leqslant b_n$ są zmiennymi losowymi NA lub ρ^*-mieszającymi.

Z nierówności Markova oraz (1.1) dla $r > \max\left\{2, \frac{5}{4\delta}\right\}$, gdzie $\delta > 0$ otrzymujemy

$$\sum_{n=1}^{\infty} P\left(\sup_{x\in\mathfrak{I}_n} |F_n(x) - F(x) - (F_n(Q(p)) - p)| \geqslant c_0 n^{-\frac{1}{2}+\delta}\right)$$

$$\leqslant C\sum_{n=1}^{\infty}\sum_{j=-b_n}^{b_n} P\left(|G_n(\eta_{j,n})| \geqslant \frac{c_0}{2} n^{-\frac{1}{2}+\delta}\right)$$

$$= C\sum_{n=1}^{\infty}\sum_{j=-b_n}^{b_n} P\left(|\sum_{i=1}^{n}\left(Y_i^{Q(p)} - Y_i^{(j,n)}\right)| \geqslant \frac{c_0}{2} n^{\frac{1}{2}+\delta}\right)$$

$$\leqslant C\sum_{n=1}^{\infty}\sum_{j=-b_n}^{b_n} P\left(|\sum_{i=1}^{n} Y_i^{Q(p)}| + |\sum_{i=1}^{n} Y_i^{(j,n)}| \geqslant \frac{c_0}{2} n^{\frac{1}{2}+\delta}\right)$$

$$\leqslant C\sum_{n=1}^{\infty}\sum_{j=-b_n}^{b_n}\left(P\left(|\sum_{i=1}^{n} Y_i^{Q(p)}| \geqslant \frac{c_0}{4} n^{\frac{1}{2}+\delta}\right) + P\left(|\sum_{i=1}^{n} Y_i^{(j,n)}| \geqslant \frac{c_0}{4} n^{\frac{1}{2}+\delta}\right)\right)$$

$$\leqslant C\sum_{n=1}^{\infty}\sum_{j=-b_n}^{b_n}\left[\frac{E\left(|\sum_{i=1}^{n} Y_i^{Q(p)}|\right)^r}{(n^{\frac{1}{2}+\delta})^r} + \frac{E\left(|\sum_{i=1}^{n} Y_i^{(j,n)}|\right)^r}{(n^{\frac{1}{2}+\delta})^r}\right]$$

$$\leqslant C\sum_{n=1}^{\infty} 2b_n n^{-\delta r} \leqslant C\sum_{n=1}^{\infty} n^{\frac{1}{4}-\delta r} < \infty.$$

Z lematu Borella-Cantelliego dostajemy tezę. □

Twierdzenie 2.4. *(Dudek i Kuczmaszewska [12]) Niech $\{X_n, n \geqslant 1\}$ będzie ciągiem zmiennych losowych o jednej z następujących struktur zależności:*

a) φ-mieszające przy warunku $\sum_{n=1}^{\infty}\varphi(2^n) < \infty$,

b) ρ-mieszające przy warunku $\sum_{n=1}^{\infty}\rho^{\frac{2}{q}}(2^n) < \infty$, dla $q > \max\left\{2, \frac{5}{2\delta}\right\}$ dla $\delta > 0$.

Niech $\{X_n, n \geqslant 1\}$ będzie ciągiem zmiennych losowych o jednakowym rozkładzie zdefiniowanym dystrybuantą F oraz niech $Q(p)$ będzie kwantylem rzędu p. Załóżmy, że F ma dodatnią ciągłą funkcję gęstości f w pewnym sąsiedztwie $\mathfrak{D}_p$ punktu $Q(p)$ spełniającą warunek

$$0 < \sup\{f(x); x \in \mathfrak{D}_p\} < \infty. \tag{2.7}$$

Wtedy dla dowolnej $\delta > 0$

$$P\left(\sup_{x\in\mathfrak{I}_n}|F_n(x)-F(x)-(F_n(Q(p))-p)| = O(n^{-\frac{3}{4}+\delta}), \quad n\to\infty\right) = 1,$$

gdzie $\mathfrak{I}_n = \left[Q(p)-c_0n^{-\frac{1}{2}+\delta}, Q(p)+c_0n^{-\frac{1}{2}+\delta}\right]$ *dla pewnego* $c_0 > 0$.

Dowód. Dowód Twierdzenia 2.4 będzie analogiczny do dowodu Twierdzenia 2.3. Ponownie wprowadźmy ciągi $\{a_n, n \geqslant 1\}$ oraz $\{b_n, n \geqslant 1\}$ zdefiniowane następująco

$$a_n = c_0n^{-\frac{1}{2}+\delta} \quad \text{dla pewnego } \ c_0 > 0, \quad b_n = \lfloor n^{\frac{1}{4}} \rfloor + 1$$

oraz funkcję

$$G_n(x) = F_n(x) - F_n(Q(p)) - F(x) + p.$$

Niech $\eta_{j,n}$, $\alpha_{j,n}$ oraz $\mathfrak{J}_{j,n}$ będą zdefiniowane tak jak w Twierdzeniu 2.3.

Stąd dostajemy

$$\sum_{n=1}^{\infty} P\left(\sup_{x\in\mathfrak{I}_n}|F_n(x)-F(x)-(F_n(Q(p))-p)| \geqslant c_0n^{-\frac{3}{4}+\delta}\right)$$
$$\leqslant C\sum_{n=1}^{\infty} P\left(\max_{-b_n\leqslant j\leqslant b_n}|G_n(\eta_{j,n})| \geqslant \frac{c_0}{2}n^{-\frac{3}{4}+\delta}\right).$$

Dalej mamy

$$G_n(\eta_{j,n}) = F_n(\eta_{j,n}) - F_n(Q(p)) - F(\eta_{j,n}) + p = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{(j,n)} - EY_i^{(j,n)}\right),$$

gdzie $Y_i^{(j,n)} = I[Q(p) < X_i \leqslant \eta_{j,n}]$ dla $0 \leqslant j \leqslant b_n$ oraz $Y_i^{(j,n)} = -I[\eta_{j,n} < X_i \leqslant Q(p)]$ dla $-b_n \leqslant j < 0$.

Z nierówności Markova oraz (1.1) dla φ-mieszających zmiennych losowych i Lematu 1.1 dla ρ-mieszających zmiennych losowych wynika, że dla dowolnego $0 < \varepsilon < \delta$ i dla $r > \max\{2, \frac{5}{2(\delta-\varepsilon)}\}$ mamy,

$$\sum_{n=1}^{\infty} P\left(\sup_{x\in\mathfrak{I}_n}|F_n(x)-F(x)-(F_n(Q(p))-p)| \geqslant 2c_0n^{-\frac{3}{4}+\delta}\right)$$
$$\leqslant C\sum_{n=1}^{\infty}\sum_{j=-b_n}^{b_n} P\left(|G_n(\eta_{j,n})| \geqslant c_0n^{-\frac{3}{4}+\delta}\right)$$
$$\leqslant C\sum_{n=1}^{\infty}\sum_{j=-b_n}^{b_n} n^{-(\frac{1}{4}+\delta)r}E\left|\sum_{i=1}^{n}\left(Y_i^{(j,n)} - EY_i^{(j,n)}\right)\right|^r$$
$$\leqslant C\sum_{n=1}^{\infty}\sum_{j=-b_n}^{b_n} n^{-(\frac{1}{4}+\delta)r}\left[\left(n^{1+\varepsilon}E(Y_1^{(j,n)})^2\right)^{\frac{r}{2}} + nE|Y_1^{(j,n)}|^r\right]$$
$$\leqslant C\sum_{n=1}^{\infty}\sum_{j=-b_n}^{b_n} n^{-(\frac{1}{4}+\delta)r}\left(n^{1+\varepsilon}|j|a_nb_n^{-1}\right)^{\frac{r}{2}} \leqslant C\sum_{n=1}^{\infty} n^{-\frac{(\delta-\varepsilon)r}{2}+\frac{1}{4}} < \infty.$$

Na podstawie lematu Borella-Cantelliego dostajemy tezę. □

Poniżej wyznaczono reprezentację Bahadura dla czterech różnych typów zależności zmiennych losowych.

Twierdzenie 2.5. *(Dudek i Kuczmaszewska [12]) Załóżmy, że spełnione są warunki Twierdzenia 2.2. Wtedy dla dowolnej* $0 < \delta < \frac{1}{2}$

$$P\left(Q_n^i(p) = Q(p) - \frac{F_n(Q(p)) - p}{f(Q(p))} + O(n^{-\frac{1}{2}+\delta}), \;\; gdy \;\; n \to \infty\right) = 1, \tag{2.8}$$

gdzie $i = \{1,2,3,...,8\}$ *dla zmiennych losowych NA oraz* ρ^**-mieszających.*

Dowód. Ponownie rozważmy estymator $Q_n^1(p)$. Najpierw zauważmy, że

$$F_n(Q_n^1(p)) \leqslant n^{-1}(\lfloor np + 0{,}5 \rfloor + 1) = p + O(n^{-1}).$$

Z rozwinięcia Taylora dla $0 < \theta < 1$ dostajemy

$$F(Q_n^1(p)) = p + f(Q(p))(Q_n^1(p) - Q(p)) + \frac{1}{2}f'(Q(p) + \theta(Q_n^1(p) - Q(p)))(Q_n^1(p) - Q(p))^2.$$

Na podstawie (2.1) i Twierdzenia 2.2 dla $0 < \delta < \frac{1}{2}$ mamy, że

$$\begin{aligned} &|F_n(Q_n^1(p)) - F(Q_n^1(p)) + f(Q(p))(Q_n^1(p) - Q(p))| \\ \leqslant \frac{1}{2}&|f'(Q(p) + \theta(Q_n^1(p) - Q(p)))|(Q_n^1(p) - Q(p))^2 + O(n^{-1}) = o(n^{-1+2\delta}). \end{aligned} \tag{2.9}$$

Z (2.9) i Twierdzenia 2.3 dla $0 < \delta < \frac{1}{2}$ otrzymujemy, że z prawdopodobieństwem 1 zachodzi

$$\begin{aligned} &|f(Q(p))(Q_n^1(p) - Q(p)) + F_n(Q(p)) - p| \\ &\leqslant |F_n(Q_n^1(p)) - F(Q_n^1(p)) + f(Q(p))(Q_n^1(p) - Q(p))| \\ &\quad + |F_n(Q_n^1(p)) - F(Q_n^1(p)) - (F_n(Q(p)) - p)| \\ &\leqslant o(n^{-1+2\delta}) + \sup_{x \in \mathfrak{I}_n} |F_n(x) - F(x) - (F_n(Q(p)) - p)| = O(n^{-\frac{1}{2}+\delta}). \end{aligned}$$

Ponadto dla estymatorów $Q_n^2(p), Q_n^4(p), Q_n^5(p), Q_n^6(p), Q_n^7(p)$ oraz $Q_n^8(p)$ rachunki przebiegają analogicznie ze względu na fakt, że $F_n(E_i) = p + O(n^{-1})$, gdzie $i = 2,4,5,6,7,8$. Z warunku $F_n(Q_n^3(p)) \leqslant p$, szybko dostajemy analogiczne rezultaty dla $Q_n^3(p)$. Co ostatecznie daje nam (2.8). □

Twierdzenie 2.6. *(Dudek i Kuczmaszewska [12]) Załóżmy, że spełnione są warunki Twierdzenia 2.2. Wtedy dla dowolnej* $0 < \delta < \frac{1}{4}$

$$P\left(Q_n^i(p) = Q(p) - \frac{F_n(Q(p)) - p}{f(Q(p))} + O(n^{-\frac{3}{4}+\delta}), \;\; gdy \;\; n \to \infty\right) = 1, \tag{2.10}$$

gdzie $i = \{1,2,3,...,8\}$ *dla* φ *oraz* ρ*-mieszających zmiennych losowych.*

Dowód. Analogicznie jak w Twierdzeniu 2.5, na podstawie (2.9) i Twierdzenia 2.4 dla $0 < \delta < \frac{1}{4}$ dostajemy, że z prawdopodobieństwem 1 zachodzi

$$|f(Q(p))(Q_n^i(p) - Q(p)) + F_n(Q(p)) - p| = O(n^{-\frac{3}{4}+\delta}),$$

co ostatecznie daje (2.10).

□

2.2. Jednostajna asymptotyczna normalność kwantyli z próby

Badanie jednostajnej asymptotycznej normalności stanowi istotny element analizy własności estymatorów. Pozwala ono ocenić, czy rozkład estymatora dąży do rozkładu normalnego w miarę wzrostu liczebności próby, co z kolei umożliwia wiarygodne wnioskowanie statystyczne w przypadku dużych prób. W praktyce, zapewnienie jednostajnej asymptotycznej normalności gwarantuje stabilność i zbieżność wyników testów dla szerokiej klasy rozkładów, a także uzasadnia stosowanie przybliżeń normalnych w konstrukcji przedziałów ufności. Analiza jednostajnej asymptotycznej normalności jest więc nie tylko zagadnieniem teoretycznym, ale również kluczowym narzędziem w ocenie jakości metod statystycznych stosowanych w badaniach empirycznych.

W tej części pracy wykazana zostanie jednostajna asymptotyczna normalność kwantyli z próby pobranej z populacji o strukturze ρ-mieszającej. W pierwszym kroku konieczne jest wprowadzenie nowych oznaczeń.

Niech $\{p_n\}$ oraz $\{q_n\}$ będą ciągami liczbowymi o wyrazach całkowitych nieujemnych, rozbieżnymi do nieskończoności, takimi, że dla wystarczająco dużych n

$$p_n + q_n \leqslant n, \quad 0 < q_n p_n^{-1} \leqslant c < \infty. \tag{2.11}$$

Ponadto dla dowolnego $\varepsilon > 0$

$$n^{\varepsilon}(p_n^{-1}q_n)^{1+\varepsilon} \to 0, \;\; n^{-\frac{1}{2}} p_n^{\frac{1+3\varepsilon}{2}} \to 0, \;\; \rho^{\frac{1}{4}}(q_n) n^{\frac{1}{4}} p_n^{\frac{\varepsilon-1}{4}} \to 0, \tag{2.12}$$

gdy $n \to \infty$.

Oznaczmy

$$\sigma_p^2 := Var[I(X_1 \leqslant Q(p))] + 2\sum_{j=1}^{\infty} Cov[I(X_1 \leqslant Q(p)), I(X_j \leqslant Q(p))] > 0,$$

$$S_n = \frac{\sqrt{n}(p - F_n(Q(p)))}{\sigma_p} = \sum_{i=1}^{n} Y_{ni} = \sum_{i=1}^{n} \frac{P(X_i \leqslant Q(p)) - I(X_i \leqslant Q(p))}{\sqrt{n}\sigma_p},$$

oraz

$$k_n = \left\lfloor \frac{n}{p_n + q_n} \right\rfloor + 1.$$

Wtedy

$$S_n = \sum_{m=1}^{k_n} y_{nm} + \sum_{m=1}^{k_n} y'_{nm} := S'_n + S''_n,$$

gdzie $y_{nm} = \sum_{i=m(p_n+q_n)-p_n-q_n+1}^{m(p_n+q_n)-q_n} Y_{ni}$, oraz $y'_{nm} = \sum_{i=m(p_n+q_n)-q_n+1}^{m(p_n+q_n)} Y_{ni}$.

Dodatkowo wprowadźmy oznaczenia

$$\gamma_{1n} := n^{\varepsilon}(p_n^{-1}q_n)^{1+\varepsilon}, \quad \gamma_{2n} := \sum_{j=1}^{n-1} \frac{j}{n}\rho(j) + \sum_{j=n}^{\infty} \rho(j), \quad \gamma_{3n} := \sum_{t=q_n}^{\infty} \rho(t),$$

$$\gamma_{4n} := n^{-\frac{1}{2}} p_n^{\frac{1+3\varepsilon}{2}}, \quad \gamma_{5n} := \rho^{\frac{1}{4}}(q_n) n^{\frac{1}{4}} p_n^{\frac{\varepsilon-1}{4}}.$$

Następujące lematy będą kluczowe w dalszych rozważaniach.

Lemat 2.1. *(Dudek i Kuczmaszewska [13]) Niech $\{X_n, n \geqslant 1\}$ będzie ρ-mieszającym ciągiem zmiennych losowych o rozkładzie zdefiniowanym za pomocą dystrybuanty F o funkcji gęstości f ciągłej w pewnym sąsiedztwie $\mathfrak{D}_p$ punktu $Q(p)$ spełniającej (2.1) i*

$$0 < \sup\{f(x); x \in \mathfrak{D}_p\} < \infty.$$

Niech $\{p_n\}$ oraz $\{q_n\}$ spełniają warunki (2.11)-(2.12). Wtedy dla dowolnego małego $\varepsilon > 0$ i dowolnego $r \geqslant 2$

$$E|S''_n|^r \leqslant C[n^{\varepsilon}(p_n^{-1}q_n)^{1+\varepsilon}]^{\frac{r}{2}} = C(\gamma_{1n})^{\frac{r}{2}} \tag{2.13}$$

oraz

$$P\left(|S''_n| > (\gamma_{1n})^{\frac{r}{2(1+r)}}\right) \leqslant C(\gamma_{1n})^{\frac{r}{2(1+r)}}. \tag{2.14}$$

Dowód. Zauważmy, że

$$E|Y_{ni}|^r = \frac{E|P(X_i \leqslant Q_p) - I(X_i \leqslant Q_p)|^r}{n^{\frac{r}{2}}\sigma_p^r} \leqslant C\frac{1}{n^{\frac{r}{2}}}.$$

Zatem z Lematu 1.1 i faktu, że $k_n q_n > 1$ mamy

$$E|S''_n|^r = E\left|\sum_{m=1}^{k_n} \sum_{i=m(p_n+q_n)-q_n+1}^{m(p_n+q_n)} Y_{ni}\right|^r \leqslant C\left\{\left[(k_n q_n)^{1+\varepsilon} EY_{n1}^2\right]^{\frac{r}{2}} + k_n q_n E|Y_{n1}|^r\right\}$$

$$\leqslant C\left[(np_n^{-1}q_n)^{1+\varepsilon} n^{-1}\right]^{\frac{r}{2}} = C(\gamma_{1n})^{\frac{r}{2}},$$

co dowodzi (2.13). Prawdziwość (2.14) dowodzimy korzystając z nierówności Markova oraz (2.13) w następujący sposób

$$P\left(|S''_n| > (\gamma_{1n})^{\frac{r}{2(1+r)}}\right) \leqslant (\gamma_{1n})^{-\frac{r^2}{2(1+r)}} E|S''_n|^r \leqslant C(\gamma_{1n})^{\frac{r}{2(1+r)}}.$$

□

W dowodzie Lematu 2.3 wykorzystamy następującą klasyczną nierówność kowariancyjną.

Lemat 2.2. *(Shao [42]) Niech $s,t > 1$ spełniają warunek $\frac{1}{s}+\frac{1}{t}=1$. Jeśli $X \in L^s(\mathscr{F}_1^k)$, $Y \in L^t(\mathscr{F}_{k+n}^{\infty})$, wtedy*

$$|EXY - EXEY| \leqslant 10\rho^{2\min\{\frac{1}{s},\frac{1}{t}\}}(n)(EX^s)^{\frac{1}{s}}(EY^t)^{\frac{1}{t}}.$$

Lemat 2.3. *(Dudek i Kuczmaszewska [13]) Niech $\{X_n, n \geqslant 1\}$ będzie ρ-mieszającym ciągiem zmiennych losowych o jednakowym rozkładzie zdefiniowanym za pomocą dystrybuanty F. Wtedy dla $\sigma_{np}^2 := E\{\sqrt{n}[p - F_n(Q(p))]\}^2$ mamy*

$$\left|\frac{\sigma_{np}^2}{\sigma_p^2} - 1\right| = O(\gamma_{2n}). \tag{2.15}$$

Dowód. Przyjmijmy $Z_i = I(X_i \leqslant Q(p))$. Wtedy

$$\sigma_{np}^2 = E\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[Z_i - EZ_i]\right\}^2 = VarZ_1 + 2\sum_{j=1}^{n}\left(1-\frac{j}{n}\right)Cov(Z_1,Z_j).$$

Z Lematu 2.2 wynika, że

$$|\sigma_{np}^2 - \sigma_p^2| = \left|VarZ_1 + 2\sum_{j=1}^{n}\left(1-\frac{j}{n}\right)Cov(Z_1,Z_j) - VarZ_1 - 2\sum_{j=1}^{\infty}Cov(Z_1,Z_j)\right|$$

$$\leqslant 2\sum_{j=1}^{n}\frac{j}{n}|Cov(Z_1,Z_j)| + 2\sum_{j=n+1}^{\infty}|Cov(Z_1,Z_j)| \leqslant C\left(\sum_{j=1}^{n-1}\frac{j}{n}\rho(j) + \sum_{j=n}^{\infty}\rho(j)\right) = C\gamma_{2n},$$

co dowodzi (2.15). □

Lemat 2.4. *(Dudek i Kuczmaszewska [13]) Niech $\{X_n, n \geqslant 1\}$ będzie ρ-mieszającym ciągiem zmiennych losowych o jednakowym rozkładzie zdefiniowanym za pomocą dystrybuanty F. Dla dowolnego małego $\varepsilon > 0$ zachodzi*

$$|s_n^2 - 1| = O\left(\gamma_{1n}^{\frac{1}{2}} + \gamma_{2n} + \gamma_{3n}\right),$$

gdzie $s_n^2 = \sum_{m=1}^{k_n} Var(y_{nm})$.

Dowód. W dowodzie tego lematu wykorzystamy równości

$$E(S_n')^2 = E\Big(\sum_{m=1}^{k_n} y_{nm}\Big)^2 = s_n^2 + 2\sum_{1\leqslant i<j\leqslant k_n} Cov(y_{ni}, y_{nj}) \tag{2.16}$$

oraz

$$ES_n^2 = \frac{\sigma_{np}^2}{\sigma_p^2}. \tag{2.17}$$

Z równości (2.16) dostajemy

$$\begin{aligned}\left|s_n^2-1\right| &= \left|E(S_n')^2 - 2\sum_{1\leqslant i<j\leqslant k_n} Cov(y_{ni},y_{nj}) - 1\right| \\ &\leqslant \left|E(S_n')^2-1\right| + 2\sum_{1\leqslant i<j\leqslant k_n}\left|Cov(y_{ni},y_{nj})\right| = I_1+I_2. \end{aligned} \tag{2.18}$$

Z Lematu 2.3 oraz równości (2.17) otrzymujemy

$$E(S_n')^2 = ES_n^2 - 2E(S_nS_n'') + E(S_n'')^2 = E(S_n'')^2 - 2E(S_nS_n'') + 1 + O(\gamma_{2n}). \tag{2.19}$$

Dalej, z równości (2.19), nierówności Höldera oraz Lematu 2.1 dostajemy oszacowanie

$$I_1 \leqslant E(S_n'')^2 + 2(ES_n^2)^{\frac{1}{2}}(E(S_n'')^2)^{\frac{1}{2}} + O(\gamma_{2n}) = O(\gamma_{1n}^{\frac{1}{2}} + \gamma_{2n}). \tag{2.20}$$

Następnie, z Lematu 2.2 mamy

$$\begin{aligned} I_2 &\leqslant \sum_{1\leqslant i<j\leqslant k_n}\ \sum_{s=i(p_n+q_n)-p_n-q_n+1}^{m(p_n+q_n)-q_n}\ \sum_{t=j(p_n+q_n)-p_n-q_n+1}^{j(p_n+q_n)-q_n} |Cov(Y_{ns},Y_{nt})| \\ &\leqslant Cn^{-1}\sum_{1\leqslant i<j\leqslant k_n}\ \sum_{s=i(p_n+q_n)-p_n-q_n+1}^{i(p_n+q_n)-q_n}\ \sum_{t=j(p_n+q_n)-p_n-q_n+1}^{j(p_n+q_n)-q_n} \rho(t-s) \\ &\leqslant Cn^{-1}\sum_{i=1}^{k_n-1}\ \sum_{s=i(p_n+q_n)-p_n-q_n+1}^{i(p_n+q_n)-q_n}\ \sum_{t=q_n}^{\infty}\rho(t) \leqslant Cn^{-1}k_np_n\sum_{t=q_n}^{\infty}\rho(t) = C\gamma_{3n}. \end{aligned} \tag{2.21}$$

Z (2.18), (2.20) i (2.21) mamy $|s_n^2-1| = O(\gamma_{1n}^{\frac{1}{2}} + \gamma_{2n} + \gamma_{3n})$. □

Niech $\{z_{nm}, 1\leqslant m\leqslant k_n\}$ będą niezależnymi kopiami $\{y_{nm}, 1\leqslant m\leqslant k_n\}$ oraz niech

$$T_n := \sum_{m=1}^{k_n} z_{nm} \quad \text{i} \quad B_n := \sum_{m=1}^{k_n} Var(z_{nm}).$$

Zauważmy, że $F_{T_n}(u) = F_{\frac{T_n}{\sqrt{B_n}}}\left(\dfrac{u}{s_n}\right)$, gdzie F_A oznacza dystrybuantę zmiennej losowej A.

Lemat 2.5. *(Dudek i Kuczmaszewska [13]) Niech $\{X_n, n\geqslant 1\}$ będzie ρ-mieszającym ciągiem zmiennych losowych o jednakowym rozkładzie zdefiniowanym za pomocą dystrybuanty F. Dla dowolnego małego $\varepsilon > 0$ dostajemy*

$$\sup_{-\infty<u<\infty}\left|F_{\frac{T_n}{\sqrt{B_n}}}(u) - \Phi(u)\right| = O(\gamma_{4n}) \tag{2.22}$$

oraz

$$\sup_{-\infty<u<\infty}\left|F_{S_n'}(u) - F_{T_n}(u)\right| = O(\gamma_{4n}+\gamma_{5n}), \tag{2.23}$$

gdzie $\Phi(u)$ jest dystrybuantą zmiennej losowej o rozkładzie normalnym $N(0,1)$.

Dowód. Najpierw wykażemy prawdziwość (2.22). Z Lematu 1.1 mamy

$$E|z_{nm}|^3 = E\left|\sum_{i=m(p_n+q_n)-p_n-q_n+1}^{m(p_n+q_n)-q_n} Y_{ni}\right|^3 \leqslant Cn^{-\frac{3}{2}} p_n^{\frac{3(1+\varepsilon)}{2}}. \tag{2.24}$$

Wykorzystując Lemat 2.4 oraz oszacowanie (2.24) zauważamy, że

$$B_n^{-\frac{3}{2}} \sum_{m=1}^{k_n} E|z_{nm}|^3 \leqslant C\frac{n}{p_n} n^{-\frac{3}{2}} p_n^{\frac{3(1+\varepsilon)}{2}} \leqslant Cn^{-\frac{1}{2}} p_n^{\frac{1+3\varepsilon}{2}} = O(\gamma_{4n}).$$

Stosując twierdzenie Berry'ego-Esseena dostajemy

$$\sup_{-\infty<u<\infty}\left|F_{\frac{T_n}{\sqrt{B_n}}}(u) - \Phi(u)\right| \leqslant CB_n^{-\frac{3}{2}} \sum_{m=1}^{k_n} E|z_{nm}|^3 = O(\gamma_{4n}),$$

co dowodzi prawdziwości (2.22).

Aby udowodnić prawdziwość (2.23) wykorzystamy następującą nierówność Esseena z pracy Fainleiba [16].

$$\sup_u |F_{S_n'}(u) - F_{T_n}(u)| \leqslant \int_{-T}^{T}\left|\frac{\chi(t)-\psi(t)}{t}\right| dt + T\sup_u \int\limits_{|y|\leqslant\frac{c}{T}} |F_{T_n}(u+y) - F_{T_n}(u)|dy, \tag{2.25}$$

gdzie T, c są stałymi dodatnimi, a χ i ψ są funkcjami charakterystycznymi, odpowiednio, S_n' oraz T_n.

Z Lematu 3.1 z pracy Dinga i Li [9] i Lematu 1.1 mamy

$$|\chi(t)-\psi(t)| \leqslant C|t|\rho^{\frac{1}{2}}(q_n)k_n[p_n^{1+\varepsilon}EY_{1n}^2 + p_nEY_{1n}^2]^{\frac{1}{2}} \leqslant C|t|\rho^{\frac{1}{2}}(q_n)n^{\frac{1}{2}}p_n^{\frac{\varepsilon-1}{2}}. \tag{2.26}$$

Ponadto

$$\sup_u |F_{T_n}(u+y) - F_{T_n}(u)| \leqslant \sup_u\left|F_{\frac{T_n}{\sqrt{B_n}}}\left(\frac{u+y}{s_n}\right) - \Phi\left(\frac{u+y}{s_n}\right)\right| \tag{2.27}$$

$$+\sup_u\left|\Phi\left(\frac{u+y}{s_n}\right) - \Phi\left(\frac{u}{s_n}\right)\right| + \sup_u\left|F_{\frac{T_n}{\sqrt{B_n}}}\left(\frac{u}{s_n}\right) - \Phi\left(\frac{u}{s_n}\right)\right| \leqslant C\left(n^{-\frac{1}{2}}p_n^{\frac{1+3\varepsilon}{2}} + |y|\right).$$

Z rozważań (2.25)-(2.27) otrzymujemy

$$\sup_u |F_{S_n'(u)} - F_{T_n}(u)| \leqslant C\left(T\rho^{\frac{1}{2}}(q_n)n^{\frac{1}{2}}p_n^{\frac{\varepsilon-1}{2}} + n^{-\frac{1}{2}}p_n^{\frac{1+3\varepsilon}{2}} + \frac{1}{T}\right). \tag{2.28}$$

Przyjmując $T = \rho^{-\frac{1}{4}}(q_n)n^{-\frac{1}{4}}p_n^{\frac{1-\varepsilon}{4}}$ dostajemy tezę (2.23). □

Korzystając z wyżej udowodnionych lematów zostanie udowodniona jednostajna asymptotyczna normalność kwantyli z próby pochodzącej z populacji o strukturze ρ-mieszającej.

Twierdzenie 2.7. *(Dudek i Kuczmaszewska [13]) Niech $\{X_n, n \geqslant 1\}$ będą ρ-mieszającymi zmiennymi losowymi o rozkładzie zdefiniowanym za pomocą dystrybuanty F, o współczynnikach mieszających $\{\rho(n), n \geqslant 1\}$ spełniających warunek $\sum_{n=1}^{\infty} \rho(n) < \infty$. Załóżmy ponadto, że spełniony jest warunek (2.1) oraz warunek*

$$0 < \sup\{f(x) = F'(x); x \in \mathfrak{D}_p\} < \infty.$$

Niech ciągi $\{p_n\}$ oraz $\{q_n\}$ będą zdefiniowane jak w (2.11)-(2.12). Wtedy dla dowolnego $r \geqslant 2$ mamy

$$\sup_{-\infty<u<\infty} \left| P\left(\frac{\sqrt{n}(Q_n^i(p) - Q(p))}{\frac{\sigma_p}{f(Q(p))}} \leqslant u \right) - \Phi(u) \right| = O\Big((\gamma_{1n})^{\frac{r}{2(1+r)}} + \gamma_{2n} + \gamma_{3n} + \gamma_{4n} + \gamma_{5n} \Big),$$

gdzie $i = \{1, 2, 3, ..., 8\}$.

Dowód. Z warunku $\sum_{n=1}^{\infty} \rho(n) < \infty$ i Twierdzenia 2.6 wynika, że

$$\lim_{n\to\infty} \sup_u \left| P\left(\frac{\sqrt{n}(Q_n^i(p) - Q(p))}{\frac{\sigma_p}{f(Q(p))}} \leqslant u \right) - \Phi(u) \right| = \lim_{n\to\infty} \sup_u |F_{S_n}(u) - \Phi(u)|.$$

Dlatego, aby udowodnić tezę, wystarczy pokazać, że

$$\sup_u |F_{S_n}(u) - \Phi(u)| = O\Big((\gamma_{1n})^{\frac{r}{2(1+r)}} + \gamma_{2n} + \gamma_{3n} + \gamma_{4n} + \gamma_{5n} \Big).$$

Na podstawie Lematu 2.5 i Lematu 2.4 dostajemy, że

$$\sup_u |F_{S_n'}(u) - \Phi(u)| \leqslant \sup_u (|F_{S_n'}(u) - F_{T_n}(u)| + |F_{T_n}(u) - \Phi\Big(\frac{u}{s_n}\Big)| + |\Phi\Big(\frac{u}{s_n}\Big) - \Phi(u)|)$$

$$\leqslant \sup_u |F_{S_n'}(u) - F_{T_n}(u)| + \sup_u |F_{\frac{T_n}{\sqrt{B_n}}}\Big(\frac{u}{s_n}\Big) - \Phi\Big(\frac{u}{s_n}\Big)| + C\frac{|s_n - 1|}{s_n} \sup_u |u| e^{-\frac{[u+\theta(\frac{u}{s_n}-u)]^2}{2}}$$

$$\leqslant \sup_u |F_{S_n'}(u) - F_{T_n}(u)| + \sup_u |F_{\frac{T_n}{\sqrt{B_n}}}(u) - \Phi(u)| + C|s_n^2 - 1| = C(\gamma_{1n}^{\frac{1}{2}} + \gamma_{2n} + \gamma_{3n} + \gamma_{4n} + \gamma_{5n}).$$

Stosując Lemat 3.7 z pracy Yanga [52], oszacowania (2.26) oraz (2.14) otrzymujemy tezę

$$\sup_u |F_{S_n}(u) - \Phi(u)| \leqslant \sup_u |P(S_n' + S_n'' \leqslant u) - \Phi(u)|$$

$$\leqslant \gamma_{4n} + \gamma_{5n} + \gamma_{1n}^{\frac{1}{2}} + \gamma_{2n} + \gamma_{3n} + (\gamma_{1n})^{\frac{r}{2(1+r)}} + P\Big(|S_n''| > (\gamma_{1n})^{\frac{r}{2(1+r)}} \Big)$$

$$\leqslant C\Big[\gamma_{4n} + \gamma_{5n} + \gamma_{1n}^{\frac{1}{2}} + \gamma_{2n} + \gamma_{3n} + (\gamma_{1n})^{\frac{r}{2(1+r)}} + (\gamma_{1n})^{\frac{r}{2(1+r)}} \Big].$$

□

Uwaga 2.1. Na podstawie Twierdzenia 2.7 obierając $q_n = \lfloor n^a \rfloor$, $p_n = \lfloor n^{\frac{1+a}{2}} \rfloor$ oraz $0 < \varepsilon < \min\left\{\frac{1-a}{1+a}, \frac{3a-1}{1+a}\right\}$, gdzie $\frac{1}{3} < a < 1$ wynika, że

$$\frac{\sqrt{n}(Q_n^i(p) - Q(p))}{\frac{\sigma_p}{f(Q(p))}} \xrightarrow{d} N(0,1), \qquad \text{gdy } n \to \infty.$$

Ponadto z Twierdzenia 2.7 wynikają poniższe dwa wnioski.

Wniosek 2.1. Niech $\{X_n, n \geqslant 1\}$ będzie ρ-mieszającym ciągiem zmiennych losowych o jednakowym rozkładzie zdefiniowanym za pomocą dystrybuanty F oraz niech $\rho(n) = O(n^{-\alpha})$, $\alpha \geqslant \frac{3}{2}$. Załóżmy, że spełnione są warunki (2.1) oraz (2.7). Wtedy dla dowolnego $0 < \kappa < \frac{1}{6}$ mamy

$$\sup_{-\infty<u<\infty} \left| P\left(\frac{\sqrt{n}(Q_n^i(p) - Q(p))}{\frac{\sigma_p}{f(Q(p))}} \leqslant u \right) - \Phi(u) \right| = O\left(n^{-\frac{1}{6}+\kappa}\right).$$

Dowód. Niech $p_n = \lfloor n^{\frac{2}{3}} \rfloor$, $q_n = \lfloor n^{\frac{1}{3}} \rfloor$, $\alpha \geqslant \frac{3}{2}$ oraz $0 < \varepsilon < \kappa$. Zauważmy, że dla wystarczająco dużego $r \geqslant 2$ mamy $-\left(\frac{1}{6} - \frac{\varepsilon}{3}\right)\frac{r}{r+1} < -\frac{1}{6} + \kappa$. Zatem

$$(\gamma_{1n})^{\frac{r}{2(1+r)}} \leqslant C\left[n^{\varepsilon}(n^{-\frac{1}{3}})^{1+\varepsilon}\right]^{\frac{r}{2(r+1)}} \leqslant Cn^{-\left(\frac{1}{6}-\frac{\varepsilon}{3}\right)\frac{r}{r+1}} = O\left(n^{-\frac{1}{6}+\kappa}\right).$$

Ponadto

$$\gamma_{2n} \leqslant C\left(\frac{1}{n}\sum_{j=1}^{n} j^{-\frac{1}{2}} + \sum_{j=n}^{\infty} j^{-\frac{3}{2}}\right) \leqslant Cn^{-\frac{1}{2}} = O\left(n^{-\frac{1}{6}+\kappa}\right),$$

$$\gamma_{3n} = \sum_{t=q_n}^{\infty} \rho(t) \leqslant Cn^{\frac{1-\alpha}{3}} = O\left(n^{-\frac{1}{6}+\kappa}\right),$$

$$\gamma_{4n} = n^{-\frac{1}{2}} p_n^{\frac{1+3\varepsilon}{2}} \leqslant Cn^{-\frac{1}{2}} n^{\frac{2}{3}\frac{1+3\varepsilon}{2}} \leqslant Cn^{-\frac{1}{6}+\varepsilon} = O\left(n^{-\frac{1}{6}+\kappa}\right).$$

Przyjmując ε spełniające warunek $\frac{\varepsilon}{6} + \frac{1}{8} < \kappa$ dostajemy

$$\gamma_{5n} = \rho^{\frac{1}{4}}(q_n) n^{\frac{1}{4}} p_n^{\frac{\varepsilon-1}{4}} \leqslant Cn^{-\frac{\alpha}{12}} n^{\frac{1}{4}} n^{\frac{2}{3}\frac{\varepsilon-1}{4}} = Cn^{-\frac{\alpha}{12}+\frac{\varepsilon}{6}+\frac{1}{12}} = O\left(n^{-\frac{1}{6}+\kappa}\right).$$

□

Wniosek 2.2. Niech $\{X_n, n \geqslant 1\}$ będzie ρ-mieszającym ciągiem zmiennych losowych o jednakowym rozkładzie zdefiniowanym za pomocą dystrybuanty F oraz niech $\rho(n) = O(e^{-sn})$, $s > \frac{3}{2}$. Załóżmy, że spełnione są warunki (2.1) oraz (2.7). Wtedy dla dowolnego $0 < \tau < \frac{1}{4}$ dostajemy

$$\sup_{-\infty<u<\infty} \left| P\left(\frac{\sqrt{n}(Q_n^i(p) - Q(p))}{\frac{\sigma_p}{f(Q(p))}} \leqslant u \right) - \Phi(u) \right| = O\left(n^{-\frac{1}{4}+\tau}\right).$$

Dowód. Niech $p_n = \lfloor n^{\frac{1}{2}} \rfloor$ oraz $q_n = \lfloor \log n \rfloor$. Wybierając ε spełniające warunek $0 < \varepsilon < \frac{4\tau}{3}$ i wystarczająco duże r, takie że $-\left(\frac{1}{4} - \frac{\varepsilon}{4}\right)\frac{r}{r+1} < -\frac{1}{4} + \tau$, dostajemy

$$(\gamma_{1n})^{\frac{r}{2(1+r)}} = O(n^{-\frac{1}{4}+\tau}) \leqslant C\left(n^{\varepsilon}(n^{-\frac{1}{2}}\log n)^{1+\varepsilon}\right)^{\frac{r}{2(1+r)}} \leqslant Cn^{-(\frac{1}{4}-\frac{\varepsilon}{4})\frac{r}{r+1}}(\log n)^{\frac{r(1+\varepsilon)}{2(r+1)}} = O\left(n^{-\frac{1}{4}+\tau}\right),$$

$$\gamma_{2n} = C\left(\sum_{j=1}^{n-1}\frac{j}{n}e^{-sj} + \sum_{j=n}^{\infty}e^{-sj}\right) \leqslant Ce^{-sn} = O\left(n^{-\frac{1}{4}+\tau}\right),$$

$$\gamma_{3n} \leqslant Ce^{-s\log n} = Cn^{-s} = O\left(n^{-\frac{1}{4}+\tau}\right),$$

$$\gamma_{4n} \leqslant Cn^{-\frac{1}{2}}n^{\frac{1+3\varepsilon}{4}} = n^{-\frac{1}{4}+\frac{3\varepsilon}{4}} = O\left(n^{-\frac{1}{4}+\tau}\right),$$

$$\gamma_{5n} \leqslant C(e^{-sq_n})^{\frac{1}{4}}n^{\frac{1}{4}}n^{\frac{\varepsilon-1}{8}} \leqslant Cn^{-\frac{s}{4}+\frac{\varepsilon+1}{8}} = O\left(n^{-\frac{1}{4}+\tau}\right),$$

co kończy dowód. □

2.2.1. Symulacje

W tym podrozdziale przedstawione zostaną wyniki symulacji, której celem było numeryczne zilustrowanie silnej zgodności oraz asymptotycznej normalności estymatora kwantyla dla prób z populacji ρ-mieszających.

W tym celu rozważano następujący ciąg (Peligrad [36])

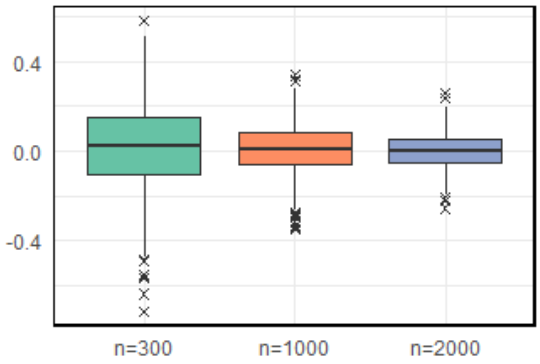
$$X_k^{(\alpha)} = Y_k - Y_{k-1} + \sqrt{\alpha}Z_k,$$

gdzie $k \geqslant 0$, $\alpha > 0$, $\{Y_k\}$, $\{Z_k\}$ niezależne zmienne losowe o jednakowych rozkładach normalnych $N(0,1)$. Zauważmy, że dla tak zdefiniowanych zmiennych losowych współczynniki mieszające $\{\rho(n)\}$ spełniają warunek $\rho(n) = 0$ dla $n \geqslant 2$. W związku z czym powyższy ciąg jest ρ–mieszający spełnia warunki

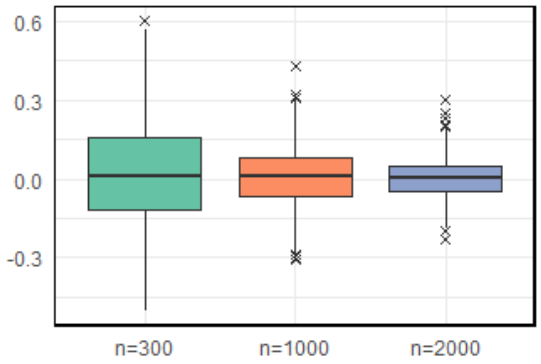
$$E\left(X_k^{(\alpha)}\right)^2 = 2+\alpha, \quad \sigma_n^2 = 2+n\alpha, \quad \inf_n \frac{\sigma_n^2}{n} = \alpha,$$

gdzie $\sigma_n^2 = Var(X_1 + ... + X_n)$.

W celu zilustrowania rozkładu estymatora kwantyla wygenerowano 1000 próbek (po 100 000 elementów każda) ciągu $\{X_k^{(\alpha)}, k \geqslant 1, \alpha = 1\}$ przy użyciu oprogramowania *R* studio. Empiryczne przybliżenie kwantyla $Q(p)$, dla $p \in \{0.05, 0.95\}$, zostało wyznaczone dla każdego z wygenerowanych ciągów. Następnie, wykorzystując pierwsze $n \in \{300, 1000, 2000\}$ obserwacji tej samej realizacji, obliczono estymator $Q_n^4(p)$. Na Rysunkach 2.1 oraz 2.2 przedstawiono wykresy boxplot dla różnic $Q_n^4(p) - Q(p)$. Wykresy pokazują jak rozmiar próby wpływa na rozkład estymatora kwantyla.

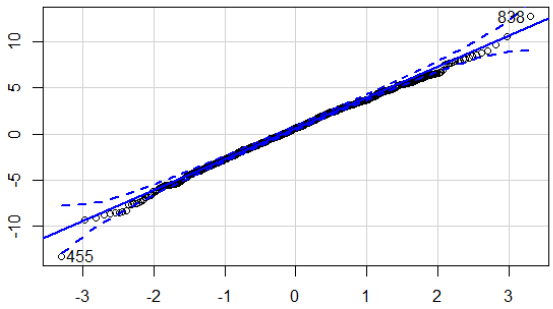


Rysunek 2.1: Wykres boxplot dla $Q_n^4(p) - Q(p)$ dla $p = 0,05$.

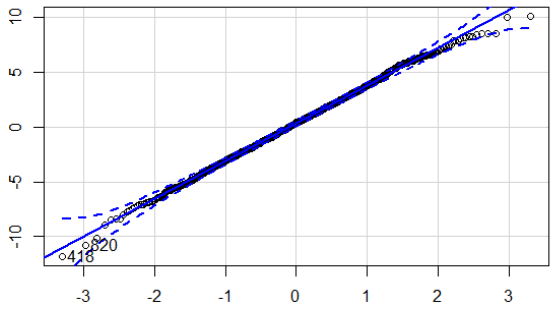


Rysunek 2.2: Wykres boxplot dla $Q_n^4(p) - Q(p)$ dla $p = 0,95$.

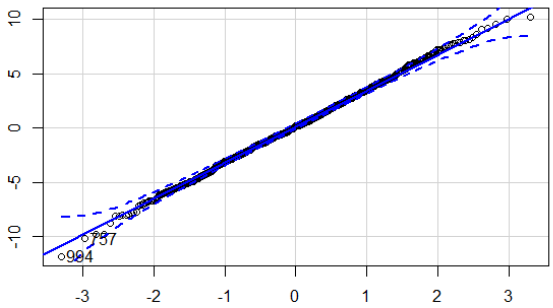
Następnie wyznaczono statystykę $U_n = \sqrt{n}(Q_n^4(p) - Q(p))$ dla $p = 0,05$ oraz $p = 0,95$. W celu oceny zgodności rozkładu statystyki U_n z rozkładem normalnym sporządzono wykresy kwantyl-kwantyl, przedstawione na Rysunkach 2.3-2.8. Pozwalają one zweryfikować, czy wraz ze wzrostem liczebności próby rozkład statystyki zbliża się do rozkładu normalnego.



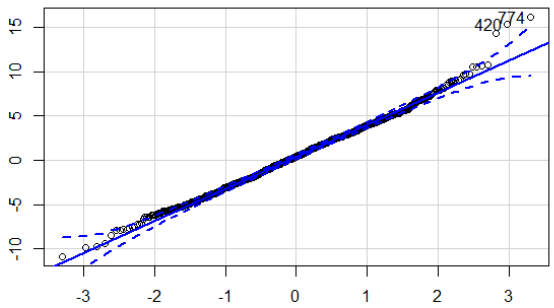
Rysunek 2.3: Próbka: $n = 300$, $p = 0,05$.
Źródło: [13]



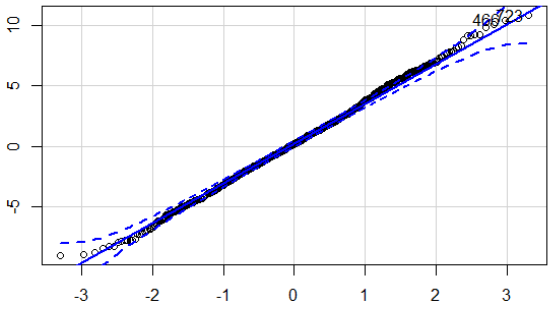
Rysunek 2.4: Próbka: $n = 1000$, $p = 0,05$.
Źródło: [13]



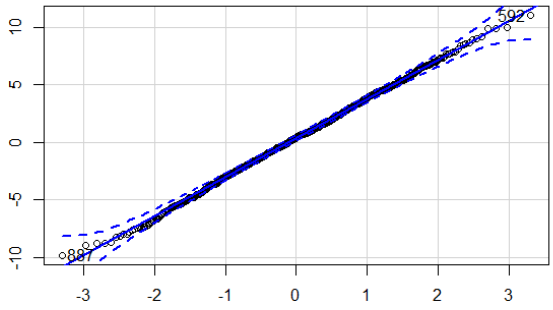
Rysunek 2.5: Próbka: $n = 2000$, $p = 0,05$.
Źródło: [13]



Rysunek 2.6: Próbka: $n = 300$, $p = 0,95$.
Źródło: [13]



Rysunek 2.7: Próbka: $n = 1000$, $p = 0.95$.
Źródło: [13]



Rysunek 2.8: Próbka: $n = 2000$, $p = 0.95$.
Źródło: [13]

Z Uwagi 2.1 mamy, że $U_n \xrightarrow{d} N\left(0, \frac{\sigma_p^2}{f(Q(p))^2}\right)$. Zgodność tego wyniku z symulacjami potwierdzają Rysunki 2.3-2.8. Wraz ze wzrostem liczebności próby punkty na wykresach kwantyl-kwantyl układają się coraz bliżej prostej odniesienia, co wskazuje na coraz lepsze przybliżenie rozkładu statystyki U_n rozkładem normalnym.

W celu weryfikacji silnej zgodności estymatora kwantyla obliczono również błąd średniokwadratowy (MSE) oraz obciążenie (bias). Wyniki zebrano w Tabeli 2.1.

Tabela 2.1: Obciążenie i MSE dla estymatora kwantyla dla $p = 0.05$ oraz $p = 0.95$. **Źródło:** [13]

	$p = 0.05$			$p = 0.95$		
	$n = 300$	$n = 1000$	$n = 2000$	$n = 300$	$n = 1000$	$n = 2000$
MSE	0.0377	0.0119	0.0057	0.0437	0.0119	0.0059
Bias	0.0310	0.0101	0.0030	0.0256	0.0080	0.0074

Powyższe wyniki pokazują, że zarówno MSE jak i obciążenie maleją wraz ze wzrostem liczności próby. Symulacje przedstawione w tej sekcji dobrze ilustrują wyniki teoretyczne uzyskane w Rozdziale 2.

Uwaga 2.2. W pracy Dudek i Kuczmaszewskiej [11] zaprezentowano analogiczne wyniki jak w podrozdziale 2.2, ale dla zmiennych losowych ρ^*-mieszających. We wspomnianej pracy, z wykorzystaniem reprezentacji Bahadura, została wykazana jednostajna asymptotyczna normalność estymatora kwantyla danego wzorem $Q_n^4(p) = X_{(\lfloor np \rfloor + 1)}$.

Rozdział 3

Reprezentacja Bahadura w przestrzeni podliniowej wartości oczekiwanej

Modelowanie systemów technicznych oraz procesów ekonomicznych – takich jak prognozowanie rynków finansowych czy ocena ryzyka kredytowego – wymaga zastosowania odpowiednich narzędzi matematycznych. Klasyczna teoria prawdopodobieństwa okazuje się w wielu sytuacjach niewystarczająca, ponieważ wiele rzeczywistych zjawisk przebiega w warunkach niepewności, której tradycyjne modele nie są w stanie dobrze opisywać. W związku z tym coraz większe zainteresowanie koncentruje się na nieaddytywnych miarach prawdopodobieństwa oraz nieaddytywnych wartościach oczekiwanych. W ostatnich latach obserwuje się znaczny rozwój teorii przestrzeni nieliniowych wartości oczekiwanych.

W pracy Penga [38] wprowadzone zostało pojęcie przestrzeni podliniowej wartości oczekiwanej jako naturalne uogólnienie klasycznej przestrzeni probabilistycznej. Prace te stały się podstawą nowych narzędzi do modelowania zjawisk w warunkach niepewności.

Konstrukcja tej przestrzeni opiera się na zastąpieniu liniowej wartości oczekiwanej odpowiednim funkcjonałem. Podliniową wartość oczekiwaną można wyrazić jako supremum po pewnej rodzinie liniowych wartości oczekiwanych $\{E_\theta, \theta \in \Theta\}$, tj.

$$\mathbb{E}[X] = \sup_{\theta \in \Theta} E_\theta[X],$$

gdzie odpowiadający zbiór miar prawdopodobieństwa $\{P_\theta, \theta \in \Theta\}$ reprezentuje niepewność modelu.

Rozszerzenie klasycznej przestrzeni probabilistycznej wymaga ponownego zdefiniowania podstawowych pojęć tak, aby odgrywały one analogiczną rolę w nowej przestrzeni. Do pojęć tych należą m.in. wartość oczekiwana, wariancja, rozkład zmiennej losowej, prawdopodobieństwo, niezależność zdarzeń, prawa wielkich liczb (Marinacci [31], Zhang [59]), centralne twierdzenie graniczne (Peng [37]), zupełna zbieżność oraz zupełna zbieżność momentów (Zhong [61]). Innym parametrem rozkładu, który wymaga nowej definicji, jest kwantyl, gdyż jego klasyczna interpretacja opiera się na jednej ustalonej mierze prawdopodobieństwa. W pracy Kuczmaszewskiej [25] zaproponowano nową definicję kwantyla w przestrzeni podliniowej wartości oczekiwanej.

Wiele problemów z klasycznej liniowej przestrzeni probabilistycznej zostało przeniesionych do przestrzeni podliniowej wartości oczekiwanej. Celem tego rozdziału jest wyznaczenie reprezentacji Bahadura dla estymatora kwantyla dla niezależnych zmiennych losowych w przestrzeni podliniowej wartości oczekiwanej.

W całym rozdziale symbol c oznacza pewną dodatnią stałą, która może przyjmować różne wartości w różnych miejscach, $I(A)$ oznacza indykator zbioru A, natomiast $\lfloor x \rfloor$ oznacza część całkowitą liczby x.

3.1. Wprowadzenie i podstawowe założenia

Niech Ω będzie ustalonym zbiorem oraz niech $\mathscr{H}$ będzie przestrzenią liniową funkcji rzeczywistych określonych na Ω. Zgodnie z pracą Penga [38] zakładamy, że $\mathscr{H}$ spełnia warunki: $c \in \mathscr{H}$ dla każdego $c \in \mathbb{R}$ oraz $|X| \in \mathscr{H}$ jeśli $X \in \mathscr{H}$.

Podliniowa wartość oczekiwana $\mathbb{E}$ jest zdefiniowana następująco.

Definicja 3.1. (Peng [38]) Podliniowa wartość oczekiwana $\mathbb{E}$ jest funkcjonałem $\mathbb{E}\colon \mathscr{H} \mapsto \mathbb{R}$ spełniającym następujące własności dla wszystkich $X, Y \in \mathscr{H}$, $c \in \mathbb{R}$:

a) monotoniczność
$$X \geqslant Y \implies \mathbb{E}(X) \geqslant \mathbb{E}(Y),$$

b) zachowanie stałych
$$\mathbb{E}(c) = c,$$

c) podaddytywność
$$\mathbb{E}(X) + \mathbb{E}(Y) \geqslant \mathbb{E}(X+Y),$$

d) dodatnia jednorodność
$$\mathbb{E}(\lambda X) = \lambda \mathbb{E}(X), \ \ \forall \lambda \geqslant 0.$$

Trójka $(\Omega, \mathscr{H}, \mathbb{E})$ nazywana jest przestrzenią podliniowej wartości oczekiwanej.

Dla danej podliniowej wartości oczekiwanej $\mathbb{E}$ sprzężona wartość oczekiwana $\mathscr{E}$ zdefiniowana jest następującym wzorem

$$\mathscr{E}(X) = -\mathbb{E}(-X) \quad \forall X \in \mathscr{H}.$$

Podliniowe wartości oczekiwane $\mathbb{E}$ oraz $\mathscr{E}$ nazywane są, odpowiednio, górną i dolną wartością oczekiwaną zmiennej $X \in \mathscr{H}$.

Klasyczne pojęcie prawdopodobieństwa z przestrzeni probabilistycznej $(\Omega, \mathscr{F}, P)$ zastępujemy pojęciem pojemności. Niech $\mathscr{G} \subset \mathscr{F}$. Pojemnością nazywamy funkcję $V\colon \mathscr{G} \to [0,1]$ spełniającą warunki

$$V(\emptyset) = 0, \quad V(\Omega) = 1 \quad \text{oraz} \quad V(A) \leqslant V(B) \quad \text{dla każdego} \quad A \subset B, \quad \text{gdzie} \quad A, B \in \mathscr{G}.$$

Jeśli pojemność V spełnia warunek $V(A\cup B)\leqslant V(A)+V(B)$ dla każdego $A,B\in\mathscr{G}$, gdzie $A\cup B\in\mathscr{G}$, wtedy V jest nazywana podaddytywną pojemnością.

Dla ustalonej trójki $(\Omega,\mathscr{H},\mathbb{E})$ definiujemy parę pojemności $(\mathbb{V},\mathscr{V})$, wykorzystując definicje, odpowiednio, górnej i dolnej podliniowej wartości oczekiwanej:

$$\mathbb{V}(A):=\inf\{\mathbb{E}|\xi|\colon I_A\leqslant\xi,\xi\in\mathscr{H}\},\quad \mathscr{V}(A)=1-\mathbb{V}(A^c)\quad \forall A\in\mathscr{F}.$$

Wtedy jeśli $I_A\in\mathscr{H}$ mamy

$$\mathbb{V}(A)=\mathbb{E}(I_A)\quad\text{oraz}\quad \mathscr{V}(A)=\mathscr{E}(I_A).$$

Dla tak zdefiniowanych pojemności możemy określić pojęcia przeliczalnej podaddytywności i ciągłości.

Definicja 3.2. (Zhang [58]) Funkcja $V\colon\mathscr{F}\to[0,1]$ jest:

1. przeliczalnie podaddytywna, jeśli
$$V\left(\bigcup_{n=1}^{\infty}A_n\right)\leqslant\sum_{n=1}^{\infty}V(A_n),\quad \forall A_n\in\mathscr{F};$$
2. ciągła, jeśli spełnione są jednocześnie warunki:
   (a) ciągłości od dołu: $V(A_n)\uparrow V(A)$, gdy $A_n\uparrow A$, gdzie $A_n,A\in\mathscr{F}$ oraz
   (b) ciągłości od góry: $V(A_n)\downarrow V(A)$, gdy $A_n\downarrow A$, gdzie $A_n,A\in\mathscr{F}$.

Nowego zdefiniowania w tej przestrzeni wymaga również pojęcie rozkładu zmiennej losowej. Niech $X=(X_1,...,X_n)$ będzie ustalonym n-wymiarowym wektorem losowym z podliniowej przestrzeni $(\Omega,\mathscr{H},\mathbb{E})$ oraz niech $C_{l.Lip}(\mathbb{R}^n)$ oznacza przestrzeń funkcji φ spełniających warunek

$$|\varphi(x)-\varphi(y)|\leqslant C(1+|x|^m+|y|^m)|x-y|$$

dla $x,y\in\mathbb{R}^n$ i stałych $C>0$, $m\in\mathbb{N}$ zależnych od φ.

Definicja 3.3. (Peng [38]) Funkcjonał $\mathbb{F}_X$ zdefiniowany następująco

$$\mathbb{F}_X(\varphi):=\mathbb{E}(\varphi(X))\colon\ \varphi\in C_{l.Lip}(\mathbb{R}^n),$$

nazywa się rozkładem zmiennej losowej X.

Trójka $(\mathbb{R}^n,C_{l.Lip}(\mathbb{R}^n),\mathbb{F}_X)$ tworzy przestrzeń podliniowej wartości oczekiwanej, gdzie $\mathbb{F}_X$ jest podliniową wartością oczekiwaną. Ponadto, na podstawie Twierdzenia 1.2.2 w pracy Penga [38], wiemy, że istnieje pewna rodzina miar prawdopodobieństwa $\{F_X^\theta,\theta\in\Theta\}$ zdefiniowana w przestrzeni $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ taka, że

$$\mathbb{F}_X(\varphi)=\sup_{\theta\in\Theta}\int_{\mathbb{R}^n}\varphi(x)F_X^\theta(dx),\quad\text{dla każdej}\quad\varphi\in C_{l.Lip}(\mathbb{R}^n).$$

Kolejna definicja wprowadza pojęcie niezależności zmiennych losowych wyrażone w terminach przestrzeni podliniowej wartości oczekiwanej.

Definicja 3.4. (Peng [38]) Rozważmy przestrzeń podliniowej wartości oczekiwanej $(\Omega, \mathscr{H}, \mathbb{E})$. Wektor losowy $Y \in \mathscr{H}^n$ nazywany jest niezależnym od innego wektora losowego $X \in \mathscr{H}^m$, jeśli dla każdej funkcji $\varphi \in C_{l.Lip}(\mathbb{R}^{m+n})$ mamy

$$\mathbb{E}(\varphi(X,Y)) = \mathbb{E}\left(\mathbb{E}(\varphi(x,Y))|_{x=X}\right).$$

Ciąg $\{X_n, n \geqslant 1\}$ nazywany jest ciągiem niezależnych zmiennych losowych, jeśli zmienna losowa X_{i+1} jest niezależna od wektora losowego $(X_1, X_2, ..., X_i)$ dla każdego $i \geqslant 1$.

Poniżej przedstawiono dwa ważne lematy związane z przestrzenią podliniowej wartości oczekiwanej.

Lemat 3.1. *(Nierówność typu Rosenthala, Zhang [59]) Niech $X_1, ..., X_n$ będzie ciągiem niezależnych zmiennych losowych z przestrzeni $(\Omega, \mathscr{H}, \mathbb{E})$. Dodatkowo niech $EX_k \leqslant 0$ oraz $S_k = X_1 + X_2 + ... + X_k$, $k = 1, ..., n$. Wtedy dla $p \geqslant 2$*

$$\mathbb{E}\left(\left|\max_{k\leqslant n} S_k\right|^p\right) \leqslant C_p\left\{\left(\sum_{i=1}^{n} \mathbb{E}X_i^2\right)^{\frac{p}{2}} + \sum_{i=1}^{n} \mathbb{E}|X_i|^p\right\}.$$

Lemat 3.2. *(Lemat Borela-Cantelliego, Zhang [59]) Niech $\{A_n, n > 1\}$ będzie ciągiem zdarzeń z σ-ciała $\mathscr{F}$. Niech V będzie ciągłą podaddytywną pojemnością. Jeśli $\sum_{n=1}^{\infty} V(A_n) < \infty$, wtedy*

$$V\left(\limsup_{n\to\infty} A_n\right) = 0, \quad gdzie \quad \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i.$$

Kolejnym kluczowym pojęciem jest pojęcie kwantyla w przestrzeni podliniowej wartości oczekiwanej.

Definicja 3.5. (Kuczmaszewska [25]) Niech $(\Omega, \mathscr{H}, \mathbb{E})$ będzie przestrzenią podliniowej wartości oczekiwanej. Dolnym kwantylem rzędu p, $0 < p < 1$, zmiennej losowej $X \in \mathscr{H}$ nazywamy liczbę $\underline{q}_p$ zdefiniowaną wzorem

$$\underline{q}_p = \inf\{x\colon \mathbb{V}(X \leqslant x) \geqslant p\}.$$

Górnym kwantylem rzędu p, $0 < p < 1$, zmiennej losowej $X \in \mathscr{H}$ nazywamy liczbę $\overline{q}_p$ zdefiniowaną wzorem

$$\overline{q}_p = \inf\{x\colon \mathscr{V}(X \leqslant x) \geqslant p\}.$$

Rozważmy próbę losową $X_1, X_2, ..., X_n$ i jej uporządkowaną, niemalejącą postać $X_{n,1}, X_{n,2}, ..., X_{n,n}$. W dalszych rozważaniach w tym rozdziale będziemy przyjmować kwantyl empiryczny rzędu p zdefiniowany wzorem:

$$Q_n^4(p) = X_{n,\lfloor np\rfloor+1}.$$

W nawiązaniu do pracy Bahadura [2], gdzie kwantyl empiryczny został wyrażony za pomocą dystrybuanty empirycznej dla niezależnych zmiennych losowych o jednakowym rozkładzie w klasycznej przestrzeni $(\Omega, \mathscr{F}, P)$ będziemy zajmować się analogiczną reprezentacją w przestrzeni $(\Omega, \mathscr{H}, \mathbb{E})$.

3.2. Reprezentacja Bahadura

W pierwszym kroku wprowadźmy założenia, które będą kluczowe w Rozdziale 3.

A. Niech $\{X_n, n \geqslant 1\}$ będzie ciągiem niezależnych zmiennych losowych z przestrzeni podliniowej wartości oczekiwanej $(\Omega, \mathscr{H}, \mathbb{E})$ o tym samym rozkładzie $\mathbb{F}_X$ zdefiniowanym pewną rodziną rozkładów $\{F_X^\theta, \theta \in \Theta\}$.

B. Załóżmy, że pojemność $\mathbb{V}$ jest ciągła, przeliczalnie podaddytywna oraz dystrybuanta F_X^θ, dla każdego $\theta \in \Theta$, posiada dodatnią, ciągłą gęstość f_X^θ w pewnym sąsiedztwie kwantyla górnego $\overline{q}_p$ oraz kwantyla dolnego $\underline{q}_p$ taką, że

$$\sup_{\theta\in\Theta} f_X^\theta(\overline{q}_p) < c \quad \text{oraz} \quad \sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p) < c.$$

C. Załóżmy, że $\sup_{\theta\in\Theta}(f_X^\theta)'(x)$ jest ograniczone w pewnym sąsiedztwie $\underline{q}_p$ oraz $\inf_{\theta\in\Theta}(f_X^\theta)'(x)$ jest ograniczone w pewnym sąsiedztwie $\overline{q}_p$.

Następujące twierdzenie przedstawia reprezentację Bahadura dla estymatora kwantyla dla niezależnych zmiennych losowych w terminach przestrzeni podliniowej wartości oczekiwanej.

Twierdzenie 3.1. *Załóżmy, że spełnione są założenia A, B oraz C. Wtedy dla dowolnej* $0 < \delta < \frac{1}{2}$

$$\mathbb{V}\Bigg(Q_n^4(p) < \underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} - cn^{-\frac{1}{2}+\delta}$$

$$\vee \quad Q_n^4(p) > \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\inf_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-\frac{1}{2}+\delta}, \quad \textit{nieskończenie wiele razy} \Bigg) = 0.$$

Dowód. Niech

$$A_n = \left\{ Q_n^4(p) < \underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} - cn^{-\frac{1}{2}+\delta} \right\}$$

oraz

$$B_n = \left\{ Q_n^4(p) > \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\inf_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-\frac{1}{2}+\delta} \right\}.$$

Aby wykazać tezę musimy pokazać, że

$$\sum_{n=1}^{\infty} \mathbb{V}(A_n) < \infty \tag{3.1}$$

oraz

$$\sum_{n=1}^{\infty} \mathbb{V}(B_n) < \infty. \tag{3.2}$$

Zauważmy, że

$$\begin{aligned}
A_n = &\left\{ Q_n^4(p) < \underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} - cn^{-\frac{1}{2}+\delta} \quad \wedge \quad Q_n^4(p) \leqslant \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} \right\} \\
&\cup \left\{ Q_n^4(p) < \underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} - cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} < Q_n^4(p) < \underline{q}_p \right\} \\
&\cup \left\{ Q_n^4(p) < \underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} - cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \underline{q}_p \leqslant Q_n^4(p) \leqslant \overline{q}_p \right\} \\
&\cup \left\{ Q_n^4(p) < \underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} - cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \overline{q}_p < Q_n^4(p) < \overline{q}_p + \varepsilon n^{-\frac{1}{2}+\delta} \right\} \\
&\cup \left\{ Q_n^4(p) < \underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} - cn^{-\frac{1}{2}+\delta} \quad \wedge \quad Q_n^4(p) \geqslant \overline{q}_p + \varepsilon n^{-\frac{1}{2}+\delta} \right\} \\
&= A_n^1 \cup A_n^2 \cup A_n^3 \cup A_n^4 \cup A_n^5.
\end{aligned}$$

Łatwo widać, że $A_n^1 \subset \{Q_n^4(p) \leqslant \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta}\}$ oraz $A_n^5 \subset \{Q_n^4(p) \geqslant \overline{q}_p + \varepsilon n^{-\frac{1}{2}+\delta}\}$. Stąd na podstawie pracy Kuczmaszewskiej [25] dostajemy

$$\sum_{n=1}^{\infty} \mathbb{V}\left(A_n^1\right) \leqslant \sum_{n=1}^{\infty} \mathbb{V}\left(Q_n^4(p) \leqslant \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta}\right) < \infty \tag{3.3}$$

oraz

$$\sum_{n=1}^{\infty} \mathbb{V}\left(A_n^5\right) \leqslant \sum_{n=1}^{\infty} \mathbb{V}\left(Q_n^4(p) \geqslant \overline{q}_p + \varepsilon n^{-\frac{1}{2}+\delta}\right) < \infty. \tag{3.4}$$

Teraz rozważymy zbiór A_n^3. Z faktu, że $Q_n^4(p) \leqslant \overline{q}_p$ dostajemy

$$F_n(Q_n^4(p)) = p + cn^{-1} \leqslant F_n(\overline{q}_p),$$

co, na podstawie założenia B, jest równoważne nierówności

$$-\frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} \leqslant -cn^{-1}.$$

Ponadto, dla $\underline{q}_p \leqslant Q_n^4(p)$ otrzymujemy

$$\underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} + cn^{-1} \leqslant Q_n^4(p) \tag{3.5}$$

oraz

$$F_n(\underline{q}_p) \leqslant F_n\left(Q_n^4(p)\right) = p + cn^{-1}. \tag{3.6}$$

Na podstawie założenia B oraz (3.6) mamy

$$\frac{F_n(\underline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} \leqslant cn^{-1}.$$

Zatem dla $\underline{q}_p \leqslant Q_n^4(p) \leqslant \overline{q}_p$ dostajemy

$$Q_n^4(p) \leqslant \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-1} \tag{3.7}$$

i ostatecznie z (3.5) oraz (3.7) wynika, że

$$\underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} + cn^{-1} \leqslant Q_n^4(p) \leqslant \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-1}. \tag{3.8}$$

Stąd mamy

$$A_n^3 = \left\{ Q_n^4(p) < \underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} - cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \underline{q}_p \leqslant Q_n^4(p) \leqslant \overline{q}_p \right\} = \emptyset$$

i wtedy

$$\sum_{n=1}^{\infty} \mathbb{V}\left(A_n^3\right) = 0. \tag{3.9}$$

W kolejnym kroku rozważymy zbiór A_n^2. Aby pokazać, że $\sum_{n=1}^{\infty} \mathbb{V}\left(A_n^2\right) < \infty$ konieczne jest wprowadzenie rozszerzenia funkcjonału $\mathbb{E}$ na zbiór dowolnych zmiennych losowych. Na podstawie pracy Zhanga [60] definiujemy rozszerzenie w następujący sposób

$$\mathbb{E}^*(X) = \inf\{\mathbb{E}(Y) \colon X \leqslant Y,\ Y \in \mathscr{H}\}.$$

Tak zdefiniowane $\mathbb{E}^*$ jest podliniową wartością oczekiwaną w przestrzeni wszystkich zmiennych losowych oraz

$$\mathbb{E}^*(X) = \mathbb{E}(X), \quad \forall X \in \mathscr{H}, \text{ a } \quad \mathbb{V}(A) = \mathbb{E}^*(I_A), \quad \forall A \in \mathscr{F}.$$

W kolejnym etapie wprowadzimy funkcję h_a^μ, która będzie niezbędna w dalszych rozważaniach.

Dla $a \in \mathbb{R}$ oraz $0 < \mu < 1$, niech $h_a^\mu \in C_{l,Lip}(\mathbb{R}^n)$ będzie nierosnącą funkcją taką, że $0 \leqslant h_a^\mu(x) \leqslant 1$ dla każdego $x \in \mathbb{R}$ oraz

$$h_a^\mu(x) = 1 \quad \text{gdy} \quad x \in (-\infty, a-\mu], \quad h_a^\mu(x) = 0 \quad \text{gdy} \quad x > a.$$

Zauważmy, że

$$\mathbb{E}^*(I(X \leqslant a)) = \lim_{\varepsilon\to 0} \mathbb{E}(h^{\varepsilon}_{a+\varepsilon}(X)) = \lim_{\varepsilon\to 0} \sup_{\theta\in\Theta} E^{\theta}(h^{\varepsilon}_{a+\varepsilon}(X))$$

$$= \lim_{\varepsilon\to 0} \sup_{\theta\in\Theta} \int_{-\infty}^{a+\varepsilon} h^{\varepsilon}_{a+\varepsilon}(x) f^{\theta}(x)dx = \lim_{\varepsilon\to 0} \sup_{\theta\in\Theta} \left[\int_{-\infty}^{a} f^{\theta}(x)dx + \int_{a}^{a+\varepsilon} h^{\varepsilon}_{a+\varepsilon}(x) f^{\theta}(x)dx \right]$$

$$= \lim_{\varepsilon\to 0} \sup_{\theta\in\Theta} \left[F^{\theta}(a) + \int_{a}^{a+\varepsilon} h^{\varepsilon}_{a+\varepsilon}(x) f^{\theta}(x)dx \right] = \sup_{\theta\in\Theta} F^{\theta}_X(a).$$

Zatem mamy

$$\mathbb{V}(X \leqslant a) = \mathbb{E}^*(I(X \leqslant a)) = \sup_{\theta\in\Theta} F^{\theta}_X(a).$$

Wykorzystując rozwinięcie Taylora funkcji F^{θ}, dla dostatecznie małych $0 < \tau_1 < 1$, dostajemy

$$\mathbb{V}\left(X_i \leqslant Q^4_n(p)\right) = \mathbb{E}^*\left(I(X_i \leqslant Q^4_n(p))\right) = \sup_{\theta\in\Theta} F^{\theta}_X\left(Q^4_n(p)\right)$$

$$= \sup_{\theta\in\Theta} \left[F^{\theta}_X(\underline{q}_p) + f^{\theta}_X(\underline{q}_p)(Q^4_n(p) - \underline{q}_p) + (f^{\theta}_X)'(\underline{q}_p + \tau_1(Q^4_n(p) - \underline{q}_p))(Q^4_n(p) - \underline{q}_p)^2 \right]. \tag{3.10}$$

Powyższe rozważania implikują, że

$$\mathbb{V}\left(A^2_n\right) = \mathbb{V}\left(Q^4_n(p) < \underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f^{\theta}_X(\underline{q}_p)} - cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} < Q^4_n(p) \leqslant \underline{q}_p \right)$$

$$= \mathbb{V}\left(\sup_{\theta\in\Theta} \left[f^{\theta}_X(\underline{q}_p) \right] (Q^4_n(p) - \underline{q}_p) + F_n(\overline{q}_p) - p < -cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} < Q^4_n(p) \leqslant \underline{q}_p \right)$$

$$= \mathbb{V}\Bigg(F_n(Q^4_n(p)) - \mathbb{V}(X_i \leqslant Q^4_n(p)) + \sup_{\theta\in\Theta} \left[f^{\theta}_X(\underline{q}_p) \right] (Q^4_n(p) - \underline{q}_p)$$

$$- F_n(Q^4_n(p)) + \mathbb{V}(X_i \leqslant Q^4_n(p)) + F_n(\overline{q}_p) - p < -cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} < Q^4_n(p) \leqslant \underline{q}_p \Bigg)$$

$$\leqslant \mathbb{V}\Bigg(F_n(Q^4_n(p)) - \mathbb{V}(X_i \leqslant Q^4_n(p)) + \sup_{\theta\in\Theta} \left[f^{\theta}_X(\underline{q}_p) \right] (Q^4_n(p) - \underline{q}_p) < -cn^{-\frac{1}{2}+\delta}$$

$$\wedge\, \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} < Q^4_n(p) \leqslant \underline{q}_p \Bigg)$$

$$+ \mathbb{V}\left(- F_n(Q^4_n(p)) + \mathbb{V}(X_i \leqslant Q^4_n(p)) + F_n(\overline{q}_p) - p < -cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} < Q^4_n(p) \leqslant \underline{q}_p \right)$$

$$= \mathbb{V}\left(A^2_{n1}\right) + \mathbb{V}\left(A^2_{n2}\right). \tag{3.11}$$

Dla $0 < \delta < \frac{1}{2}$, na podstawie równości (3.10) i założenia C, otrzymujemy

$$\mathbb{V}\left(A_{n1}^2\right) \leqslant \mathbb{V}\Bigg(p + cn^{-1} - p - \sup_{\theta\in\Theta}\left[f_X^\theta(\underline{q}_p)\right](Q_n^4(p) - \underline{q}_p)$$

$$- \sup_{\theta\in\Theta}(f_X^\theta)'(\underline{q}_p + \tau_1(Q_n^4(p) - \underline{q}_p))(Q_n^4(p) - \underline{q}_p)^2$$

$$+ \sup_{\theta\in\Theta}\left[f_X^\theta(\underline{q}_p)\right](Q_n^4(p) - \underline{q}_p) < -cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} < Q_n^4(p) \leqslant \underline{q}_p \Bigg)$$

$$= \mathbb{V}\Bigg(cn^{-1} - \sup_{\theta\in\Theta}(f_X^\theta)'(\underline{q}_p + \tau_1(Q_n^4(p) - \underline{q}_p))(Q_n^4(p) - \underline{q}_p)^2 < -cn^{-\frac{1}{2}+\delta}$$

$$\wedge\, \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} < Q_n^4(p) \leqslant \underline{q}_p \Bigg)$$

$$\leqslant \mathbb{V}\Bigg(-M(Q_n^4(p) - \underline{q}_p)^2 < -cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} < Q_n^4(p) \leqslant \underline{q}_p \Bigg)$$

$$\leqslant \mathbb{V}\Bigg(-n^{-1+2\delta} < -cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} < Q_n^4(p) \leqslant \underline{q}_p \Bigg) = 0.$$

Stąd

$$\sum_{n=1}^{\infty} \mathbb{V}\left(A_{n1}^2\right) < \infty. \tag{3.12}$$

Teraz oszacujemy $\mathbb{V}\left(A_{n2}^2\right)$. Niech $\{a_n, n \geqslant 1\}$ oraz $\{b_n, n \geqslant 1\}$ będą dwoma ciągami zdefiniowanymi następująco

$$a_n = \varepsilon n^{-\frac{1}{4}+\delta} \quad \text{dla pewnego } \varepsilon > 0 \quad \text{oraz} \quad b_n = \lfloor n^{\frac{1}{4}} \rfloor + 1. \tag{3.13}$$

Ponadto niech

$$\mathbb{G}_n(x) = F_n(x) - F_n(\overline{q}_p) - \mathbb{V}(X_i \leqslant x) + \mathbb{V}(X_i \leqslant \overline{q}_p)$$

i

$$\eta_{j,n} = \underline{q}_p + j a_n b_n^{-1}, \quad \alpha_{j,n} = \mathbb{V}(\eta_{j+1,n}) - \mathbb{V}(\eta_{j,n}) \quad \text{oraz} \quad \mathfrak{J}_{j,n} = [\eta_{j,n}, \eta_{j+1,n}],$$

dla każdego $n \in \mathbb{N}$ i dowolnego $j \in \mathbb{Z}_-$.

Dla $x \in \mathfrak{J}_{j,n}$ dostajemy

$$\mathbb{G}_n(x) \leqslant F_n(\eta_{j+1,n}) - F_n(\overline{q}_p) - \mathbb{V}(X_i \leqslant \eta_{j,n}) + \mathbb{V}(X_i \leqslant \overline{q}_p) = \mathbb{G}_n(\eta_{j+1,n}) + \alpha_{j,n}$$

i

$$\mathbb{G}_n(x) \geqslant F_n(\eta_{j,n}) - F_n(\overline{q}_p) - \mathbb{V}(X_i \leqslant \eta_{j+1,n}) + \mathbb{V}(X_i \leqslant \overline{q}_p) = \mathbb{G}_n(\eta_{j,n}) - \alpha_{j,n}.$$

Zatem dla pewnego $\varepsilon > 0$

$$\sup_{x\in\mathfrak{I}_n} |F_n(x) - \mathbb{V}(X_i \leqslant x) - (F_n(\overline{q}_p) - \mathbb{V}(X_i \leqslant \overline{q}_p))| \leqslant \max_{-b_n\leqslant j\leqslant -1} \{|\mathbb{G}_n(\eta_{j,n})|\} + \max_{-b_n\leqslant j\leqslant -1} \{\alpha_{j,n}\},$$

gdzie $\mathfrak{I}_n = \left[\underline{q}_p - \varepsilon n^{-\frac{1}{4}+\delta}, \underline{q}_p\right]$.

Z twierdzenia o wartości średniej oraz założenia B dostajemy

$$\alpha_{j,n} = \mathbb{V}(X_i \leqslant \eta_{j+1,n}) - \mathbb{V}(X_i \leqslant \eta_{j,n}) = \mathbb{E}^*(I(X_i \leqslant \eta_{j+1,n})) - \mathbb{E}^*(I(X_i \leqslant \eta_{j,n}))$$

$$= \sup_{\theta\in\Theta} F_X^\theta(\eta_{j+1,n}) - \sup_{\theta\in\Theta} F_X^\theta(\eta_{j,n}) \leqslant \sup_{\theta\in\Theta} \left(F_X^\theta(\eta_{j+1,n}) - F_X^\theta(\eta_{j,n})\right)$$

$$\leqslant \sup_{\theta\in\Theta} f_X^\theta\left(\eta_{j,n} + \tau(\eta_{j+1,n} - \eta_{j,n})\right) \varepsilon n^{-\frac{1}{2}+\delta} \leqslant c\varepsilon n^{-\frac{1}{2}+\delta}.$$

Na podstawie powyższych rozważań otrzymujemy

$$\sum_{n=1}^{\infty} \mathbb{V}(A_{n2}^2) \leqslant \sum_{n=1}^{\infty} \mathbb{V}\left(\sup_{x\in\mathfrak{I}_n} |F_n(x) - \mathbb{V}(X_i \leqslant x) - (F_n(\overline{q}_p) - \mathbb{V}(X_i \leqslant \overline{q}_p))| > cn^{-\frac{1}{2}+\delta}\right)$$

$$\leqslant \sum_{n=1}^{\infty} \mathbb{V}\left(\max_{-b_n\leqslant j\leqslant -1} |\mathbb{G}_n(\eta_{j,n})| > cn^{-\frac{1}{2}+\delta}\right).$$

Ponadto zauważmy, że

$$\mathbb{G}_n(\eta_{j,n}) = F_n(\eta_{j,n}) - F_n(\overline{q}_p) - \mathbb{V}(X_i \leqslant \eta_{j,n}) + \mathbb{V}(X_i \leqslant \overline{q}_p) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{\eta_{j,n}} - Y_i^{\overline{q}_p}\right),$$

gdzie $Y_i^{\eta_{j,n}} = I(X_i \leqslant \eta_{j,n}) - \mathbb{E}^*(I(X_i \leqslant \eta_{j,n}))$ oraz $Y_i^{\overline{q}_p} = I(X_i \leqslant \overline{q}_p) - \mathbb{E}^*(I(X_i \leqslant \overline{q}_p))$.

Na podstawie nierówności Markova oraz Lematu 3.1 dla $r > \max\{2, \frac{5}{4\delta}\}$, dowolnej $\delta > 0$ dostajemy

$$\sum_{n=1}^{\infty} \mathbb{V}\left(\sup_{x\in\mathfrak{I}_n} |F_n(x) - \mathbb{V}(X_i \leqslant \eta_{j,n}) - (F_n(\overline{q}_p) - \mathbb{V}(X_i \leqslant \overline{q}_p))| > cn^{-\frac{1}{2}+\delta}\right)$$

$$\leqslant \sum_{n=1}^{\infty}\sum_{j=-b_n}^{-1} \mathbb{V}\left(|\mathbb{G}_n(\eta_{j,n})| > cn^{-\frac{1}{2}+\delta}\right)$$

$$= \sum_{n=1}^{\infty}\sum_{j=-b_n}^{-1} \mathbb{V}\left(\left|\sum_{i=1}^{n}\left(Y_i^{\eta_{j,n}} - Y_i^{\overline{q}_p}\right)\right| > cn^{\frac{1}{2}+\delta}\right)$$

$$\leqslant \sum_{n=1}^{\infty}\sum_{j=-b_n}^{-1} \mathbb{V}\left(\left|\sum_{i=1}^{n} Y_i^{\eta_{j,n}}\right| + \left|\sum_{i=1}^{n} Y_i^{\overline{q}_p}\right| > cn^{\frac{1}{2}+\delta}\right)$$

$$\leqslant \sum_{n=1}^{\infty}\sum_{j=-b_n}^{-1} \left[\mathbb{V}\left(\left|\sum_{i=1}^{n} Y_i^{\eta_{j,n}}\right| > \frac{c}{2}n^{\frac{1}{2}+\delta}\right) + \mathbb{V}\left(\left|\sum_{i=1}^{n} Y_i^{\overline{q}_p}\right| > \frac{c}{2}n^{\frac{1}{2}+\delta}\right)\right]$$

$$\leqslant \sum_{n=1}^{\infty}\sum_{j=-b_n}^{-1} \left\{n^{-(\frac{1}{2}+\delta)r}\left[\mathbb{E}^*\left(\left|\sum_{i=1}^{n} Y_i^{\eta_{j,n}}\right|\right)^r + \mathbb{E}^*\left(\left|\sum_{i=1}^{n} Y_i^{\overline{q}_p}\right|\right)^r\right]\right\}$$

$$\leqslant \sum_{n=1}^{\infty} \sum_{j=-b_n}^{-1} n^{-(\frac{1}{2}+\delta)r} \left\{ \left[n^{\frac{r}{2}} \left(\mathbb{E}^*(Y_i^{\eta_{j,n}})^2 \right)^{\frac{r}{2}} + n\mathbb{E}^*|Y_i^{\eta_{j,n}}|^r \right] + \left[n^{\frac{r}{2}} \left(\mathbb{E}^*(Y_i^{\overline{q}_p})^2 \right)^{\frac{r}{2}} + n\mathbb{E}^*|Y_i^{\overline{q}_p}|^r \right] \right\}$$

$$\leqslant \sum_{n=1}^{\infty} b_n n^{-\delta r} \leqslant C \sum_{n=1}^{\infty} n^{-\delta r+\frac{1}{4}} < \infty.$$

Zatem dla przeliczalnie podaddytywnej pojemności $\mathbb{V}$ mamy

$$\sum_{n=1}^{\infty} \mathbb{V}(A_{n2}^2) < \infty. \tag{3.14}$$

Na podstawie (3.11), (3.12) i (3.14) otrzymujemy

$$\sum_{n=1}^{\infty} \mathbb{V}(A_n^2) < \infty. \tag{3.15}$$

Aby zakończyć tę część dowodu należy rozważyć jeszcze przypadek zbioru A_n^4. Zauważmy, że w tym przypadku zachodzą jednocześnie dwie nierówności

$$Q_n^4(p) < \underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} - cn^{-\frac{1}{2}+\delta}$$

oraz

$$Q_n^4(p) > \overline{q}_p.$$

Wobec tego

$$\overline{q}_p < \underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} - cn^{-\frac{1}{2}+\delta}$$

oraz

$$p - F_n(\overline{q}_p) > \sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)(\overline{q}_p - \underline{q}_p) + cn^{-\frac{1}{2}+\delta} \geqslant cn^{-\frac{1}{2}+\delta}.$$

Zatem

$$A_n^4 \subseteq \left\{ p - F_n(\overline{q}_p) > cn^{-\frac{1}{2}+\delta} \right\}.$$

Z faktu, że $\mathscr{V}(X_i \leqslant \overline{q}_p) = p$ mamy

$$p - F_n(\overline{q}_p) = \frac{1}{n} \sum_{i=1}^{n} Z_i^{\overline{q}_p},$$

gdzie $Z_i^{\overline{q}_p} = \mathscr{E}^*(I(X_i \leqslant \overline{q}_p)) - I(X_i \leqslant \overline{q}_p)$, a $\mathscr{E}^*(X) = -\mathbb{E}^*(-X)$.

Stąd, na podstawie nierówności Markova oraz Lematu 3.1, dla $r \geqslant 2$ otrzymujemy

$$\mathbb{V}(A_n^4) \leqslant \mathbb{V}\left(\left| \sum_{i=1}^{n} Z_i^{\overline{q}_p} \right| > cn^{\frac{1}{2}+\delta} \right) \leqslant cn^{-(\frac{1}{2}+\delta)r} \mathbb{E}^* \left| \sum_{i=1}^{n} Z_i^{\overline{q}_p} \right|^r$$

$$\leqslant cn^{-(\frac{1}{2}+\delta)r} \left[n^{\frac{r}{2}} \left(\mathbb{E}^*|Z_1^{\overline{q}_p}|^2 \right)^{\frac{r}{2}} + n\mathbb{E}^*|Z_1^{\overline{q}_p}|^r \right] \leqslant cn^{-r\delta}.$$

Wybierając $r > \frac{1}{\delta}$ otrzymujemy

$$\sum_{n=1}^{\infty} \mathbb{V}(A_n^4) < \infty. \tag{3.16}$$

Ostatecznie na podstawie (3.3), (3.4), (3.9), (3.15) i (3.16) dostajemy

$$\sum_{n=1}^{\infty} \mathbb{V}(A_n) \leqslant \sum_{n=1}^{\infty} \mathbb{V}(A_n^1) + \sum_{n=1}^{\infty} \mathbb{V}(A_n^2) + \sum_{n=1}^{\infty} \mathbb{V}(A_n^3) + \sum_{n=1}^{\infty} \mathbb{V}(A_n^4) + \sum_{n=1}^{\infty} \mathbb{V}(A_n^5) < \infty. \tag{3.17}$$

W drugiej części dowodu rozważamy zbiór

$$B_n = \left\{ Q_n^4(p) > \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\inf_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-\frac{1}{2}+\delta} \right\}.$$

Analogicznie jak w części pierwszej zauważmy, że

$$\begin{aligned}
B_n = &\left\{ Q_n^4(p) > \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\inf_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-\frac{1}{2}+\delta} \quad \wedge \quad Q_n^4(p) \leqslant \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} \right\} \\
\cup &\left\{ Q_n^4(p) > \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\inf_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} < Q_n^4(p) < \underline{q}_p \right\} \\
\cup &\left\{ Q_n^4(p) > \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\inf_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \underline{q}_p \leqslant Q_n^4(p) \leqslant \overline{q}_p \right\} \\
\cup &\left\{ Q_n^4(p) > \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\inf_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \overline{q}_p < Q_n^4(p) < \overline{q}_p + \varepsilon n^{-\frac{1}{2}+\delta} \right\} \\
\cup &\left\{ Q_n^4(p) > \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\inf_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-\frac{1}{2}+\delta} \quad \wedge \quad Q_n^4(p) \geqslant \overline{q}_p + \varepsilon n^{-\frac{1}{2}+\delta} \right\} \\
&= B_n^1 \cup B_n^2 \cup B_n^3 \cup B_n^4 \cup B_n^5.
\end{aligned}$$

Ponownie, łatwo stwierdzić, że $B_n^1 \subset \{Q_n^4(p) \leqslant \underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta}\}$ i $B_n^5 \subset \{Q_n^4(p) \geqslant \overline{q}_p + \varepsilon n^{-\frac{1}{2}+\delta}\}$. Zatem na podstawie pracy Kuczmaszewskiej [25] mamy

$$\sum_{n=1}^{\infty} \mathbb{V}\left(B_n^1\right) < \infty \quad \text{i} \quad \sum_{n=1}^{\infty} \mathbb{V}\left(B_n^5\right) < \infty. \tag{3.18}$$

Z oszacowania (3.8) otrzymujemy

$$\sum_{n=1}^{\infty} \mathbb{V}\left(B_n^3\right) < \infty. \tag{3.19}$$

Teraz rozważymy zbiór B_n^4. W tym celu niezbędne będzie wprowadzenie funkcji g_a^μ. Dla $a \in \mathbb{R}$ oraz $0 < \mu < 1$, niech $g_a^\mu \in C_{l,Lip}(\mathbb{R}^n)$ będzie niemalejącą funkcją taką, że $0 \leqslant g_a^\mu(x) \leqslant 1$ dla każdego $x \in \mathbb{R}$ oraz

$$g_a^\mu(x) = 0 \quad \text{gdy} \quad x \in (-\infty, a-\mu), \quad g_a^\mu(x) = 1 \quad \text{gdy} \quad x \in \langle a, +\infty).$$

Stąd dla dowolnego $\varepsilon > 0$ mamy

$$\mathscr{V}(X_i \leqslant a) = 1 - \mathbb{E}^*(I(X_i > a)) = 1 - \lim_{\varepsilon \to 0} \mathbb{E}(g_a^\varepsilon(X))$$

$$= 1 - \lim_{\varepsilon \to 0} \sup_{\theta \in \Theta} E^\theta(g_a^\varepsilon(X)) = 1 - \lim_{\varepsilon \to 0} \sup_{\theta \in \Theta} \int_{a-\varepsilon}^{\infty} g_a^\varepsilon(x) f^\theta(x) dx$$

$$= 1 - \lim_{\varepsilon \to 0} \sup_{\theta \in \Theta} \left[\int_{a-\varepsilon}^{a} g_a^\varepsilon(x) f^\theta(x) dx + \int_a^\infty f^\theta(x) dx \right]$$

$$= 1 - \sup_{\theta \in \Theta} \left[1 - F^\theta(a) \right] = \inf_{\theta \in \Theta} F^\theta(a). \tag{3.20}$$

Wykorzystując rozwinięcie Taylora funkcji F^θ, dla $0 < \tau_2 < 1$, dostajemy

$$\mathscr{V}(X_i \leqslant Q_n^4(p)) = \inf_{\theta \in \Theta} \left[F_X^\theta(\overline{q}_p) + f_X^\theta(\overline{q}_p)(Q_n^4(p) - \overline{q}_p) + (f_X^\theta)'(\overline{q}_p + \tau_2(Q_n^4(p) - \overline{q}_p))(Q_n^4(p) - \overline{q}_p)^2 \right]. \tag{3.21}$$

Wówczas

$$\mathbb{V}(B_n^4) = \mathbb{V}\left(Q_n^4(p) > \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\inf_{\theta \in \Theta} f^\theta(\overline{q}_p)} + cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \overline{q}_p < Q_n^4(p) < \overline{q}_p + \varepsilon n^{-\frac{1}{2}+\delta} \right)$$

$$= \mathbb{V}\left(\inf_{\theta \in \Theta} \left[f_X^\theta(\overline{q}_p) \right] (Q_n^4(p) - \overline{q}_p) + F_n(\underline{q}_p) - p > cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \overline{q}_p < Q_n^4(p) < \overline{q}_p + \varepsilon n^{-\frac{1}{2}+\delta} \right)$$

$$= \mathbb{V}\Bigg(F_n(Q_n^4(p)) - \mathscr{V}(X_i \leqslant Q_n^4(p)) + \inf_{\theta \in \Theta} \left[f_X^\theta(\overline{q}_p) \right] (Q_n^4(p) - \overline{q}_p)$$

$$- F_n(Q_n^4(p)) + \mathscr{V}(X_i \leqslant Q_n^4(p)) + F_n(\underline{q}_p) - p > cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \overline{q}_p < Q_n^4(p) < \overline{q}_p + \varepsilon n^{-\frac{1}{2}+\delta} \Bigg)$$

$$\leqslant \mathbb{V}\Bigg(F_n(Q_n^4(p)) - \mathscr{V}(X_i \leqslant Q_n^4(p)) + \inf_{\theta \in \Theta} \left[f_X^\theta(\overline{q}_p) \right] (Q_n^4(p) - \overline{q}_p) > cn^{-\frac{1}{2}+\delta}$$

$$\wedge\, \overline{q}_p < Q_n^4(p) < \overline{q}_p + \varepsilon n^{-\frac{1}{2}+\delta} \Bigg)$$

$$+ \mathbb{V}\left(-F_n(Q_n^4(p)) + \mathscr{V}(X_i \leqslant Q_n^4(p)) + F_n(\underline{q}_p) - p > cn^{-\frac{1}{2}+\delta} \quad \wedge \quad \overline{q}_p < Q_n^4(p) < \overline{q}_p + \varepsilon n^{-\frac{1}{2}+\delta} \right)$$

$$= \mathbb{V}(B_{n1}^4) + \mathbb{V}(B_{n2}^4). \tag{3.22}$$

Na podstawie (3.21) oraz założenia C otrzymujemy

$$\mathbb{V}(B_{n1}^4)$$

$$\leqslant \mathbb{V}\left(p + cn^{-1} - p - \inf_{\theta \in \Theta} \left[f_X^\theta(\overline{q}_p) \right] (Q_n^4(p) - \overline{q}_p) - \inf_{\theta \in \Theta} (f_X^\theta)'(\overline{q}_p + \tau_2(Q_n^4(p) - \overline{q}_p))(Q_n^4(p) - \overline{q}_p)^2 \right.$$

$$+\inf_{\theta\in\Theta}\left[f_X^\theta(\overline{q}_p)\right](Q_n^4(p)-\overline{q}_p)>cn^{-\frac{1}{2}+\delta}\quad\wedge\quad\overline{q}_p<Q_n^4(p)<\overline{q}_p+\varepsilon n^{-\frac{1}{2}+\delta}\Bigg)$$

$$=\mathbb{V}\Bigg(cn^{-1}-\inf_{\theta\in\Theta}(f_X^\theta)'(\overline{q}_p+\tau_2(Q_n^4(p)-\overline{q}_p))(Q_n^4(p)-\overline{q}_p)^2>cn^{-\frac{1}{2}+\delta}$$

$$\wedge\,\overline{q}_p<Q_n^4(p)<\overline{q}_p+\varepsilon n^{-\frac{1}{2}+\delta}\Bigg)$$

$$\leqslant\mathbb{V}\Bigg(-m(Q_n^4(p)-\overline{q}_p)^2>cn^{-\frac{1}{2}+\delta}\quad\wedge\quad\overline{q}_p<Q_n^4(p)<\overline{q}_p+\varepsilon n^{-\frac{1}{2}+\delta}\Bigg)=0.$$

Zatem

$$\mathbb{V}(B_{n1}^4)<\infty. \tag{3.23}$$

W kolejnym kroku przejdziemy do oszacowania $\mathbb{V}(B_{n2}^4)$. Oznaczmy

$$\mathscr{G}_n(x)=F_n(x)-F_n(\underline{q}_p)-\mathscr{V}(X_i\leqslant x)+\mathscr{V}(X_i\leqslant\underline{q}_p),$$

$$\eta'_{j,n}=\overline{q}_p+ja_nb_n^{-1},\quad\alpha'_{j,n}=\mathscr{V}(\eta'_{j+1,n})-\mathscr{V}(\eta'_{j,n})\quad\text{oraz}\quad\mathfrak{J}'_{j,n}=[\eta'_{j,n},\eta'_{j+1,n}],$$

dla każdego $n\in\mathbb{N}$ i dowolnego $j\in\mathbb{Z}_+\cup\{0\}$ oraz $\{a_n,n\geqslant1\}$ oraz $\{b_n,n\geqslant1\}$ zdefiniowanych jak w (3.13).

Dla $x\in\mathfrak{J}'_{j,n}$ dostajemy

$$\mathscr{G}_n(x)\leqslant F_n(\eta'_{j+1,n})-F_n(\underline{q}_p)-\mathscr{V}(X_i\leqslant\eta'_{j,n})+\mathscr{V}(X_i\leqslant\underline{q}_p)=\mathscr{G}_n(\eta'_{j+1,n})+\alpha'_{j,n}$$

i

$$\mathscr{G}_n(x)\geqslant F_n(\eta'_{j,n})-F_n(\underline{q}_p)-\mathscr{V}(X_i\leqslant\eta'_{j+1,n})+\mathscr{V}(X_i\leqslant\underline{q}_p)=\mathscr{G}_n(\eta'_{j,n})-\alpha'_{j,n}.$$

Stąd

$$\sup_{x\in\mathfrak{J}'_n}|F_n(x)-\mathscr{V}(X_i\leqslant x)-(F_n(\underline{q}_p)-\mathscr{V}(X_i\leqslant\underline{q}_p))|$$

$$\leqslant\max_{0\leqslant j\leqslant b_n}\{|\mathscr{G}_n(\eta'_{j,n})|\}+\max_{0\leqslant j\leqslant b_n-1}\{\alpha'_{j,n}\}, \tag{3.24}$$

gdzie $\mathfrak{J}'_n=\left[\overline{q}_p,\overline{q}_p+\varepsilon n^{-\frac{1}{4}+\delta}\right]$.

Ponadto, dla $0<\tau<1$, otrzymujemy

$$\alpha'_{j,n}=\mathscr{V}(X_i\leqslant\eta'_{j+1,n})-\mathscr{V}(X_i\leqslant\eta'_{j,n})=\inf_{\theta\in\Theta}F^\theta(\eta'_{j+1,n})-\inf_{\theta\in\Theta}F^\theta(\eta'_{j,n})$$

$$\leqslant\sup_{\theta\in\Theta}\left(F_X^\theta(\eta'_{j+1,n})-F_X^\theta(\eta'_{j,n})\right)\leqslant\sup_{\theta\in\Theta}f_X^\theta\left(\eta'_{j,n}+\tau(\eta'_{j+1,n}-\eta'_{j,n})\right)\varepsilon n^{-\frac{1}{2}+\delta}\leqslant c\varepsilon n^{-\frac{1}{2}+\delta}.$$

Wówczas na mocy oszacowania (3.24) mamy

$$\sum_{n=1}^{\infty}\mathbb{V}\Bigg(-F_n(Q_n^4(p))+\mathscr{V}(X_i\leqslant Q_n^4(p))+F_n(\underline{q}_p)-\mathscr{V}(X_i\leqslant\underline{q}_p)>cn^{-\frac{1}{2}+\delta}\Bigg)$$

$$\leqslant \sum_{n=1}^{\infty} \mathbb{V}\left(\sup_{x\in\mathfrak{I}'_n} |F_n(x) - \mathscr{V}(X_i \leqslant x) - (F_n(\underline{q}_p) - \mathscr{V}(X_i \leqslant \underline{q}_p))| > cn^{-\frac{1}{2}+\delta} \right)$$

$$\leqslant \sum_{n=1}^{\infty} \mathbb{V}\left(\max_{0\leqslant j\leqslant b_n} |\mathscr{G}_n(\eta_{j,n})| > cn^{-\frac{1}{2}+\delta} \right). \tag{3.25}$$

Dodatkowo, z definicji $\mathscr{G}_n$ oraz F_n dostajemy

$$\mathscr{G}_n(\eta'_{j,n}) = F_n(\eta'_{j,n}) - F_n(\underline{q}_p) - \mathscr{V}(X_i \leqslant \eta'_{j,n}) + \mathscr{V}(X_i \leqslant \underline{q}_p) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{\underline{q}_p} - Y_i^{\eta'_{j,n}} \right), \tag{3.26}$$

gdzie $Y_i^{\eta'_{j,n}} = \mathscr{E}^*(I(X_i \leqslant \eta'_{j,n})) - I(X_i \leqslant \eta'_{j,n})$ oraz $Y_i^{\underline{q}_p} = \mathscr{E}^*(I(X_i \leqslant \underline{q}_p)) - I(X_i \leqslant \underline{q}_p)$, gdzie $\mathscr{E}^*(X) = -\mathbb{E}^*(-X)$.

Korzystając z oszacowań (3.25), (3.26), nierówności Markova oraz Lematu 3.1 dla $r > \max\{2, \frac{5}{4\delta}\}$ oraz dowolnej $\delta > 0$ otrzymujemy

$$\sum_{n=1}^{\infty} \mathbb{V}\left(\sup_{x\in\mathfrak{I}'_n} |F_n(x) - \mathscr{V}(X_i \leqslant x) - (F_n(\underline{q}_p) - \mathscr{V}(X_i \leqslant \underline{q}_p))| > cn^{-\frac{1}{2}+\delta} \right)$$

$$\leqslant \sum_{n=1}^{\infty}\sum_{j=0}^{b_n} \mathbb{V}\left(|\mathscr{G}_n(\eta'_{j,n})| \geqslant cn^{-\frac{1}{2}+\delta} \right)$$

$$= \sum_{n=1}^{\infty}\sum_{j=0}^{b_n} \mathbb{V}\left(\left| \sum_{i=1}^{n}\left(Y_i^{\underline{q}_p} - Y_i^{\eta'_{j,n}} \right) \right| > cn^{\frac{1}{2}+\delta} \right)$$

$$\leqslant \sum_{n=1}^{\infty}\sum_{j=0}^{b_n} \mathbb{V}\left(\left| \sum_{i=1}^{n} Y_i^{\eta'_{j,n}} \right| + \left| \sum_{i=1}^{n} Y_i^{\underline{q}_p} \right| > cn^{\frac{1}{2}+\delta} \right)$$

$$\leqslant \sum_{n=1}^{\infty}\sum_{j=0}^{b_n} \left[\mathbb{V}\left(\left| \sum_{i=1}^{n} Y_i^{\eta'_{j,n}} \right| > \frac{c}{2}n^{\frac{1}{2}+\delta} \right) + \mathbb{V}\left(\left| \sum_{i=1}^{n} Y_i^{\underline{q}_p} \right| > \frac{c}{2}n^{\frac{1}{2}+\delta} \right) \right]$$

$$\leqslant \sum_{n=1}^{\infty}\sum_{j=0}^{b_n} \left\{ n^{-(\frac{1}{2}+\delta)r} \left[\mathbb{E}^*\left(\left| \sum_{i=1}^{n} Y_i^{\eta'_{j,n}} \right| \right)^r + \mathbb{E}^*\left(\left| \sum_{i=1}^{n} Y_i^{\underline{q}_p} \right| \right)^r \right] \right\}$$

$$\leqslant \sum_{n=1}^{\infty}\sum_{j=0}^{b_n} n^{-(\frac{1}{2}+\delta)r} \left\{ \left[n^{\frac{r}{2}} \left(\mathbb{E}^*\left(Y_i^{\eta'_{j,n}} \right)^2 \right)^{\frac{r}{2}} + n\mathbb{E}^*\left| Y_i^{\eta'_{j,n}} \right|^r \right] \right.$$

$$\left. + \left[n^{\frac{r}{2}} \left(\mathbb{E}^*\left(Y_i^{\underline{q}_p} \right)^2 \right)^{\frac{r}{2}} + n\mathbb{E}^*\left| Y_i^{\underline{q}_p} \right|^r \right] \right\} \leqslant \sum_{n=1}^{\infty} b_n n^{-\delta r} \leqslant C\sum_{n=1}^{\infty} n^{-\delta r+\frac{1}{4}} < \infty.$$

Stąd mamy, że

$$\sum_{n=1}^{\infty} \mathbb{V}(B_{n2}^4) < \infty. \tag{3.27}$$

Na podstawie (3.22), (3.23) oraz (3.27) dostajemy

$$\sum_{n=1}^{\infty} \mathbb{V}(B_n^4) < \infty. \tag{3.28}$$

Do rozważenia pozostał zbiór B_n^2. Zauważmy, że w tym przypadku zachodzą jednocześnie nierówności

$$Q_n^4(p) > \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\inf_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-\frac{1}{2}+\delta}$$

oraz

$$\underline{q}_p - \varepsilon n^{-\frac{1}{2}+\delta} < Q_n^4(p) < \underline{q}_p.$$

Zatem

$$\underline{q}_p > \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\inf_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-\frac{1}{2}+\delta}.$$

Stąd

$$F_n(\underline{q}_p) - p > \inf_{\theta\in\Theta} f_X^\theta(\overline{q}_p)(\overline{q}_p - \underline{q}_p) + cn^{-\frac{1}{2}+\delta} \geqslant cn^{-\frac{1}{2}+\delta}.$$

Wobec tego

$$B_n^2 \subseteq \left\{F_n(\underline{q}_p) - p > cn^{-\frac{1}{2}+\delta}\right\}.$$

Zauważmy, że

$$F_n(\underline{q}_p) - p = \frac{1}{n}\sum_{i=1}^n Z_i^{\underline{q}_p},$$

gdzie $Z_i^{\underline{q}_p} = I(X_i \leqslant \underline{q}_p) - \mathbb{E}^*\left(I(X_i \leqslant \underline{q}_p)\right)$.

Na podstawie nierówności Markowa oraz Lematu 3.1, dla dowolnego $r \geqslant 2$, otrzymujemy

$$\mathbb{V}(B_n^2) \leqslant \mathbb{V}\left(\left|\sum_{i=1}^n Z_i^{\underline{q}_p}\right| > cn^{\frac{1}{2}+\delta}\right)$$

$$\leqslant Cn^{-r(\frac{1}{2}+\delta)}\left[\left(\sum_{i=1}^n \mathbb{E}^*\left((Z_i^{\underline{q}_p})^2\right)\right)^{\frac{r}{2}} + \sum_{i=1}^n \mathbb{E}^*\left(|Z_i^{\underline{q}_p}|^r\right)\right] \leqslant Cn^{-r\delta}.$$

Wybierając $r > \dfrac{1}{\delta}$ otrzymujemy

$$\sum_{n=1}^\infty \mathbb{V}(B_n^2) < \infty. \tag{3.29}$$

Stąd na podstawie (3.18), (3.19), (3.28) i (3.29) mamy, że

$$\sum_{n=1}^\infty \mathbb{V}(B_n) < \infty. \tag{3.30}$$

Ostatecznie, z warunku (3.17) oraz (3.30) dostajemy

$$\sum_{n=1}^\infty \mathbb{V}\left(Q_n^4(p) < \underline{q}_p - \frac{F_n(\overline{q}_p) - p}{\sup_{\theta\in\Theta} f_X^\theta(\underline{q}_p)} - cn^{-\frac{1}{2}+\delta} \vee Q_n^4(p) > \overline{q}_p - \frac{F_n(\underline{q}_p) - p}{\inf_{\theta\in\Theta} f_X^\theta(\overline{q}_p)} + cn^{-\frac{1}{2}+\delta}\right)$$

$$\leqslant \sum_{n=1}^\infty \mathbb{V}(A_n) + \sum_{n=1}^\infty \mathbb{V}(B_n) < \infty,$$

co na mocy Lematu Borela-Cantelliego dowodzi tezy. □

Rozdział 4

Reprezentacja kwantylowa dystrybuanty sumy, różnicy, iloczynu i ilorazu zmiennych losowych

Wyznaczenie rozkładu sumy, różnicy, iloczynu oraz ilorazu niezależnych zmiennych losowych stanowi jedno z klasycznych, a zarazem trudniejszych zagadnień rachunku prawdopodobieństwa. Choć dla prostych przypadków – takich jak zmienne o rozkładzie jednostajnym czy wykładniczym – możliwe jest analityczne określenie rozkładu docelowego, to w praktyce już niewielkie skomplikowanie postaci rozkładów prowadzi do znacznych trudności obliczeniowych. W wielu sytuacjach analityczne uzyskanie rozwiązania okazuje się wręcz niemożliwe, ponieważ wymagałoby obliczania skomplikowanych całek, które nie mają prostych postaci analitycznych.

Z tego względu bardzo ważnym zagadnieniem są metody estymacji, które pozwalają przybliżyć rozkład sumy, różnicy, iloczynu czy ilorazu dwóch zmiennych losowych. Jedną z dobrze znanych technik statystycznych jest metoda Monte Carlo, która umożliwia empiryczne wyznaczenie rozkładu poprzez wielokrotne próbkowanie. Takie podejście, choć przybliżone, pozwala uzyskać bardzo dobrą zgodność z rzeczywistym rozkładem nawet w przypadkach, gdy nie istnieje żadne rozwiązanie analityczne.

W niniejszym rozdziale przedstawione zostaną propozycje wzorów przybliżających dystrybuantę różnicy, iloczynu oraz ilorazu dwóch niezależnych zmiennych losowych oparte na reprezentacji kwantylowej dystrybuanty wymienionych zmiennych losowych. Następnie na podstawie wyprowadzonego wzoru aproksymującego zostanie skonstruowany nieparametryczny estymator dystrybuanty różnicy, iloczynu i ilorazu niezależnych zmiennych losowych. Dodatkowo zostanie wykazana silna zgodność każdego z tych estymatorów. Na końcu zostaną przeprowadzone symulacje ilustrujące otrzymane wyniki teoretyczne.

4.1. Suma niezależnych zmiennych losowych

Badanie rozkładu sumy zmiennych losowych stanowi jedno z fundamentalnych zagadnień teorii prawdopodobieństwa i statystyki matematycznej. Rozkład sumy zmiennych losowych ma kluczowe znaczenie m.in. w analizie ryzyka, ekonomii, teorii błędów pomiarowych czy teorii niezawodności. Badanie tego typu rozkładów umożliwia zarówno głębsze

zrozumienie teoretycznych własności procesów losowych, jak i tworzenie skuteczniejszych modeli statystycznych wykorzystywanych w praktyce.

Rozkład sumy dwóch (lub więcej) zmiennych losowych był na przestrzeni lat przedmiotem licznych badań. Moschopoulos [33] wykazał, że rozkład sumy n niezależnych zmiennych losowych o rozkładzie gamma i różnych parametrach może zostać przedstawiony w postaci pojedynczego szeregu gamma. Beaulieu i in. [3] dokonali porównania czterech różnych przybliżeń dystrybuanty sumy niezależnych zmiennych losowych o rozkładzie logarytmiczno normalnym. Sadooghi-Alvandi i in. [41] wyprowadzili wzór na funkcję gęstości dla sumy n niezależnych zmiennych losowych o rozkładzie jednostajnym z różnymi parametrami. Z kolei Buonocore i in. [4] zaproponowali metodę wyznaczania dystrybuanty oraz funkcji gęstości sumy zmiennych losowych opartą na zasadzie indukcji. Nadarajah i in. [34] przedstawili sposób przybliżonego wyznaczania rozkładu sumy niezależnych zmiennych losowych o rozkładzie beta z różnymi parametrami kształtu, wykazując, że proponowane przybliżenie przewyższa dokładnością klasyczne przybliżenie rozkładem normalnym.

Midhu i in. [32] zaproponowali nową metodę przybliżania rozkładu sumy dwóch niezależnych, nieujemnych zmiennych losowych z wykorzystaniem reprezentacji dystrybuanty opartej na funkcjach kwantylowych.

W pierwszym kroku Midhu i in. [32] wprowadzają wzór aproksymujący dystrybuantę F_Z, gdzie $Z = X + Y$, wykorzystując numeryczną metodę trapezów liczenia całek

$$F_{Z_m}(z) = \frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(\frac{(i+1)z}{m}\right) - F_X\left(\frac{iz}{m}\right)\right)\left(F_Y\left(\frac{z(m-i-1)}{m}\right) + F_Y\left(\frac{z(m-i)}{m}\right)\right),$$

gdzie m oznacza liczbę podziałów dziedziny całkowania. Dodatkowo autorzy wyznaczyli błąd tego przybliżenia w zależności od m. Na podstawie tej aproksymacji skonstruowali nieparametryczny estymator dystrybuanty $\hat{F}_Z$ wykorzystując empiryczne dystrybuanty $\hat{F}_X$ oraz $\hat{F}_Y$. Wzór zaproponowanego estymatora to

$$\hat{F}_{Z_m}(z) = \frac{1}{2}\sum_{i=0}^{m-1}\left(\hat{F}_X\left(\frac{(i+1)z}{m}\right) - \hat{F}_X\left(\frac{iz}{m}\right)\right)\left(\hat{F}_Y\left(\frac{z(m-i-1)}{m}\right) + \hat{F}_Y\left(\frac{z(m-i)}{m}\right)\right).$$

Ponadto, autorzy wyznaczyli dokładny rozkład proponowanego estymatora oraz dowiedli jego silnej zgodności i asymptotycznej normalności.

Analizując pracę Midhu i in. [32] pojawia się naturalne pytanie, jak prezentowałyby się analogiczne wyniki dla przypadku różnicy, iloczynu i ilorazu niezależnych zmiennych losowych. Praca Midhu i in. [32] nie daje odpowiedzi na to pytanie, ale stanowi inspirację do podjęcia dalszych rozważań w tym rozdziale.

4.2. Reprezentacja kwantylowa różnicy, iloczynu i ilorazu zmiennych losowych

Różnice, iloczyny i ilorazy zmiennych losowych mają istotne znaczenie w teorii prawdopodobieństwa i statystyce. W wielu sytuacjach dany problem można sformułować właśnie w kategoriach tego typu zmiennych losowych.

Przykładowo rozważmy masową produkcję wałka, który musi wsuwać się do cylindrycznej tulei. Średnica S produkowanych wałków jest zmienną losową, podobnie jak średnica wewnętrzna C produkowanych tulei cylindrycznych. Montaż odbywa się poprzez losowy wybór wałka i tulei. Wtedy zmienna losowa $C - S$ będzie określać dopasowanie losowo wybranego wałka i tulei. Oczywiście wałek i tuleja będą pasować tylko wtedy, gdy średnica wałka będzie mniejsza od średnicy wewnętrznej tulei.

Również iloczyn zmiennych losowych jest zmienną losową obserwowaną w sytuacjach praktycznych. Na przykład na rynku energii elektrycznej rozliczanym na giełdzie energii każda transakcja charakteryzuje się dwoma zmiennymi losowymi: wolumenem sprzedanej energii (w MWh) oraz ceną (w EUR/MWh). Ostateczna wartość transakcji, będąca całkowitym zyskiem sprzedawcy energii lub całkowitym kosztem nabywcy, jest iloczynem tych dwóch zmiennych.

Ponadto wskaźnik wydajności jest ilorazem dwóch zmiennych mierzących wydajność dwóch różnych metod dla tego samego zadania. Na przykład, rozważmy izolację kabli zasilających prądu przemiennego. Dwoma powszechnie stosowanymi materiałami do izolacji kabli są papier Kraft i papier laminowany polipropylenem (PPLP). Wskaźnik wydajności można wówczas określić jako stosunek zmiennych wytrzymałościowych dla tych dwóch papierów. Jeśli X i Y oznaczają zmienne wytrzymałościowe dla papieru Kraft i PPLP, to zmienna losowa $\frac{X}{Y}$ jest wskaźnikiem wydajności.

Powyższe przykłady pokazują istotną wagę problemu polegającego na umiejętności określania rozkładu zmiennych losowych opisanych jako różnica, iloczyn czy iloraz dwóch zmiennych losowych.

W literaturze można znaleźć prace poświęcone właśnie temu problemowi. Należy jednak zauważyć, że w większości prac rozważany jest przypadek sumy zmiennych losowych. Problem różnicy, iloczynu i ilorazu rozważany jest znacznie rzadziej i w większości przypadków dotyczy bardzo konkretnych rozkładów. Na przykład Puri i Rubin [39] badali wartość bezwzględną różnicy dwóch niezależnych zmiennych losowych o jednakowych rozkładach. Lo [28] zajmował się sumą i różnicą dwóch skorelowanych zmiennych losowych o rozkładzie logarytmiczno normalnym. Joarder i Omar [23] wyprowadzili dokładny rozkład różnicy dwóch niezależnych zmiennych losowych o rozkładzie χ^2. Lomnicki [29] analizował problem wyznaczenia gęstości iloczynu $n \leqslant 7$ zmiennych losowych o jednakowym rozkładzie

normalnym. Ly i in. [30] wyznaczyli rozkład iloczynu zmiennych losowych z wykorzystaniem funkcji typu *copula*. Gaunt i in. [18] wyprowadzili dokładny wzór na gęstość iloczynu n niezależnych zmiennych losowych typu *variance-gamma*. C. Caginalp i G. Caginalp [5] analizowali iloraz normalnych zmiennych losowych. Yuancheng i in. [55] wyznaczyli dokładną postać dystrybuanty ilorazu nieujemnych zmiennych losowych dla pewnych ustalonych rozkładów.

W dalszej części rozdziału wprowadzone zostaną reprezentacje dystrybuanty różnicy, iloczynu i ilorazu niezależnych zmiennych losowych oparte na funkcjach kwantylowych, które następnie będą podstawą do skonstruowania nieparametrycznych estymatorów tych dystrybuant. Proponowane estymatory nie wymagają założenia o konkretnej postaci rozkładu zmiennych losowych X i Y.

Niech X i Y będą dwoma niezależnymi i ciągłymi zmiennymi losowymi o wartościach zawartych, odpowiednio, w przedziałach (A_X, B_X) oraz (A_Y, B_Y), gdzie $A_X, A_Y \in \mathbb{R} \cup \{-\infty\}$ oraz $B_X, B_Y \in \mathbb{R} \cup \{\infty\}$. Niech F_X i F_Y będą dystrybuantami, odpowiednio, zmiennych losowych X oraz Y. Niech F_X będzie ściśle rosnąca na przedziale (A_X, B_X), a F_Y, odpowiednio, na (A_Y, B_Y). Niech Q_X i Q_Y będą funkcjami kwantylowymi, odpowiednio, zmiennych losowych X i Y.

W przypadku różnicy, iloczynu i ilorazu zmiennych losowych konieczne jest wprowadzenie zmodyfikowanej definicji funkcji kwantylowej.

Definicja 4.1. Niech F będzie dystrybuantą pewnej zmiennej losowej. Funkcję kwantylową Q, odpowiadającą funkcji F, definiujemy następująco:

$$Q(u) = \begin{cases} \inf\{x\colon F(x) \geqslant u\}, & \text{dla } u \in (0,1], \\ \lim_{u \to 0^+} Q(u), & \text{dla } u = 0. \end{cases}$$

Korzystając z tej definicji funkcji kwantylowej możemy sformułować poniższe trzy twierdzenia.

Twierdzenie 4.1. *(Dudek i Kuczmaszewska [14]) Niech X i Y będą zmiennymi losowymi o rozkładach zdefiniowanych, odpowiednio, za pomocą dystrybuant F_X i F_Y. Niech Q_X i Q_Y będą, odpowiednio, funkcjami kwantylowymi zmiennych losowych X oraz Y. Dodatkowo zakładamy, że istnieją granice*

$$\lim_{u_1 \to 0^+, u_2 \to 0^+} [Q_X(u_1) - Q_Y(u_2)] \quad \textit{oraz} \quad \lim_{u_1 \to 1^-, u_2 \to 1^-} [Q_X(u_1) - Q_Y(u_2)].$$

Wtedy funkcja opisująca pole nad krzywą $z = Q_X(u_1) - Q_Y(u_2)$ (dla ustalonego z) w kwadracie jednostkowym jest dystrybuantą zmiennej losowej $Z_1 = X - Y$.

Dowód. Niech (u_1, u_2) będzie wartością zmiennej losowej (U_1, U_2) o rozkładzie jednostajnym na kwadracie jednostkowym. Wtedy $Q_X(u_1) - Q_Y(u_2)$ zwraca losową liczbę z rozkładu zmiennej losowej $Z_1 = X - Y$.

Rozważmy krzywą $z = Q_X(u_1) - Q_Y(u_2)$. Wtedy

$$u_2 = F_Y(Q_X(u_1) - z) \tag{4.1}$$

oraz na podstawie definicji dystrybuanty mamy

$$\begin{aligned} F_{Z_1}(z) &= P[Q_X(U_1) - Q_Y(U_2) \leqslant z] \\ &= P[U_2 \geqslant F_Y(Q_X(U_1) - z)] = \int_0^1 \int_{F_Y(Q_X(u_1)-z)}^1 1 \, du_2 du_1 \\ &= 1 - \int_0^1 F_Y(Q_X(u_1) - z) du_1 = 1 - \int_0^1 u_2 \, du_1, \end{aligned} \tag{4.2}$$

co pokazuje, że dystrybuantę F_{Z_1} w punkcie z można wyrazić jako pole nad krzywą $z = Q_X(u_1) - Q_Y(u_2)$ w kwadracie jednostkowym.

W następnym kroku wyprowadzimy wzór na postać tej dystrybuanty.

Zauważmy, że z faktu

$$0 \leqslant F_Y(Q_Y(u_1) - z) \leqslant 1$$

wynika, że

$$0 \leqslant F_X(z + Q_Y(0)) \leqslant u_1 \leqslant F_X(z + Q_Y(1)) \leqslant 1. \tag{4.3}$$

Na podstawie (4.3) mamy

$$\begin{aligned} \int_0^1 u_2 \, du_1 &= \int_0^1 F_Y(Q_X(u_1) - z) \, du_1 = \int_0^{F_X(z+Q_Y(0))} F_Y(Q_X(u_1) - z) \, du_1 \\ &+ \int_{F_X(z+Q_Y(0))}^{F_X(z+Q_Y(1))} F_Y(Q_X(u_1) - z) \, du_1 + \int_{F_X(z+Q_Y(1))}^1 F_Y(Q_X(u_1) - z) \, du_1 \\ &= I_{11} + I_{12} + I_{13}. \end{aligned} \tag{4.4}$$

Rozważamy cztery przypadki.

1° Niech $Q_X(0) - Q_Y(1) \leqslant z < \min(Q_X(0) - Q_Y(0), Q_X(1) - Q_Y(1))$. Wtedy

$$I_{11} = \int_0^{F_X(z+Q_Y(0))} F_Y(Q_X(u_1) - z) \, du_1 = 0 \tag{4.5}$$

oraz

$$I_{12} = \int_0^{F_X(z+Q_Y(1))} F_Y(Q_X(u_1) - z) \, du_1. \tag{4.6}$$

Ponadto, dla $u_1 \geqslant F_X(z + Q_Y(1))$ mamy

$$Q_X(u_1) - z \geqslant Q_Y(1)$$

i

$$1 \geqslant F_Y(Q_X(u_1) - z) \geqslant F_Y(Q_Y(1)) = 1.$$

Stąd $F_Y(Q_X(u_1)-z)=1$ oraz

$$I_{13}=\int_{F_X(z+Q_Y(1))}^{1} F_Y(Q_X(u_1)-z)\,du_1=1-F_X(z+Q_Y(1)). \tag{4.7}$$

Ostatecznie, korzystając z (4.4), (4.5), (4.6) oraz (4.7) mamy

$$\int_0^1 F_Y(Q_X(u_1)-z)\,du_1=\int_0^{F_X(z+Q_Y(1))} F_Y(Q_X(u_1)-z)\,du_1+1-F_X(z+Q_Y(1))$$

i na podstawie (4.2) dostajemy wzór

$$F_{Z_1}(z)=F_X(z+Q_Y(1))-\int_0^{F_X(z+Q_Y(1))} F_Y(Q_X(u_1)-z)\,du_1. \tag{4.8}$$

2° Niech $Q_X(1)-Q_Y(1)\leqslant z<Q_X(0)-Q_Y(0)$. Wtedy otrzymujemy

$$I_{11}=\int_0^0 F_Y(Q_X(u_1)-z)\,du_1=0, \tag{4.9}$$

$$I_{12}=\int_0^1 F_Y(Q_X(u_1)-z)\,du_1 \tag{4.10}$$

oraz

$$I_{13}=\int_1^1 F_Y(Q_X(u_1)-z)\,du_1=0. \tag{4.11}$$

Ostatecznie, na podstawie (4.2), (4.4), (4.9), (4.10) i (4.11) dostajemy

$$F_{Z_1}(z)=1-\int_0^1 F_Y(Q_X(u_1)-z)\,du_1. \tag{4.12}$$

3° Niech $Q_X(0)-Q_Y(0)\leqslant z<Q_X(1)-Q_Y(1)$. Zauważmy, że jeśli $u_1\leqslant F_X(z+Q_Y(0))$, to $Q_X(u_1)-z\leqslant Q_Y(0)$ oraz

$$0\leqslant F_Y(Q_X(u_1)-z)\leqslant F_Y(Q_Y(0))=0.$$

Stąd $F_Y(Q_X(u_1)-z)=0$ i

$$I_{11}=\int_0^{F_X(z+Q_Y(0))} F_Y(Q_X(u_1)-z)\,du_1=0 \tag{4.13}$$

oraz

$$I_{12}=\int_{F_X(z+Q_Y(0))}^{F_X(z+Q_Y(1))} F_Y(Q_X(u_1)-z)\,du_1. \tag{4.14}$$

Analogicznie jak w przypadku 1°, dostajemy

$$I_{13}=\int_{F_X(z+Q_Y(1))}^{1} F_Y(Q_X(u_1)-z)\,du_1=1-F_X(z+Q_Y(1)). \tag{4.15}$$

Na podstawie (4.4), (4.13), (4.14) oraz (4.15)

$$\int_0^1 F_Y(Q_X(u_1)-z)\,du_1=\int_{F_X(z+Q_Y(0))}^{F_X(z+Q_Y(1))} F_Y(Q_X(u_1)-z)\,du_1+1-F_X(z+Q_Y(1)).$$

Korzystając z (4.2) mamy

$$F_{Z_1}(z)=F_X(z+Q_Y(1))-\int_{F_X(z+Q_Y(0))}^{F_X(z+Q_Y(1))} F_Y(Q_X(u_1)-z)\,du_1. \tag{4.16}$$

4° Niech $\max(Q_X(0)-Q_Y(0), Q_X(1)-Q_Y(1)) \leqslant z < Q_X(1)-Q_Y(0)$. Podobnie jak w przypadku 3° otrzymujemy

$$I_{11} = \int_0^{F_X(z+Q_Y(0))} F_Y(Q_X(u_1)-z)\,du_1 = 0, \tag{4.17}$$

$$I_{12} = \int_{F_X(z+Q_Y(0))}^1 F_Y(Q_X(u_1)-z)\,du_1 \tag{4.18}$$

i

$$I_{13} = \int_1^1 F_Y(Q_X(u_1)-z)\,du_1 = 0. \tag{4.19}$$

Ostatecznie, na podstawie (4.4), (4.17), (4.18) i (4.19) dostajemy

$$\int_0^1 F_Y(Q_X(u_1)-z)\,du_1 = \int_{F_X(z+Q_Y(0))}^1 F_Y(Q_X(u_1)-z)\,du_1.$$

Stąd z (4.2) mamy

$$F_{Z_1}(z) = 1 - \int_{F_X(z+Q_Y(0))}^1 F_Y(Q_X(u_1)-z)\,du_1. \tag{4.20}$$

Z (4.8), (4.12), (4.16), (4.20) i podstawienia $u_1 = F_X(v)$ dostajemy ostateczny wzór na podstać dystrybuanty zmiennej losowej Z_1 w punkcie z

$$F_{Z_1}(z) = \begin{cases} 0, & \text{gdy} \quad z < Q_X(0)-Q_Y(1) \\ F_X(z+Q_Y(1)) - \int_{Q_X(0)}^{z+Q_Y(1)} F_Y(v-z) f_X(v)\,dv, & \text{gdy} \quad z \geqslant Q_X(0)-Q_Y(1) \text{ i} \\ & \qquad\quad z < Q_X(0)-Q_Y(0) \text{ i} \\ & \qquad\quad z < Q_X(1)-Q_Y(1) \\ 1 - \int_{Q_X(0)}^{Q_X(1)} F_Y(v-z) f_X(v)\,dv, & \text{gdy} \quad z \geqslant Q_X(1)-Q_Y(1) \text{ i} \\ & \qquad\quad z < Q_X(0)-Q_Y(0) \\ F_X(z+Q_Y(1)) - \int_{z+Q_Y(0)}^{z+Q_Y(1)} F_Y(v-z) f_X(v)\,dv, & \text{gdy} \quad z \geqslant Q_X(0)-Q_Y(0) \text{ i} \\ & \qquad\quad z < Q_X(1)-Q_Y(1) \\ 1 - \int_{z+Q_Y(0)}^{Q_X(1)} F_Y(v-z) f_X(v)\,dv, & \text{gdy} \quad z \geqslant Q_X(0)-Q_Y(0) \text{ i} \\ & \qquad\quad z \geqslant Q_X(1)-Q_Y(1) \text{ i} \\ & \qquad\quad z < Q_X(1)-Q_Y(0) \\ 1, & \text{gdy} \quad z \geqslant Q_X(1)-Q_Y(0) \end{cases}. \tag{4.21}$$

□

Kolejne twierdzenie formułuje analogiczny wynik dla dystrybuanty zmiennej losowej $Z_2 = X \cdot Y$, gdzie X i Y są niezależnymi zmiennymi losowymi o wartościach dodatnich.

Twierdzenie 4.2. *Niech $X > 0$ oraz $Y > 0$ będą dwoma zmiennymi losowymi o rozkładach zdefiniowanych, odpowiednio, za pomocą dystrybuanty F_X oraz F_Y. Niech Q_X oraz Q_Y będą, odpowiednio, funkcjami kwantylowymi zmiennych losowych X oraz Y z założeniem, że $Q_X(u_1) > 0$ dla $u_1 \in [0,1]$ oraz $Q_Y(u_2) > 0$ dla $u_2 \in [0,1]$. Dodatkowo zakładamy, że istnieje granica*

$$\lim_{u_1 \to 0^+, u_2 \to 1^-} [Q_X(u_1) \cdot Q_Y(u_2)].$$

Wtedy funkcja opisująca pole pod krzywą $z = Q_X(u_1) \cdot Q_Y(u_2)$ (dla ustalonego z) w kwadracie jednostkowym jest dystrybuantą zmiennej losowej $Z_2 = X \cdot Y$.

Dowód. Analogicznie jak w przypadku dowodu Twierdzenia 4.1 możemy zauważyć, że $Q_X(u_1) \cdot Q_Y(u_2)$, gdzie $0 < u_1, u_2 < 1$, zwraca losową liczbę z rozkładu zmiennej losowej $Z_2 = X \cdot Y$.

Niech $z = Q_X(u_1) \cdot Q_Y(u_2)$. Wtedy $u_2 = F_Y\left(\frac{z}{Q_X(u_1)}\right)$. Stąd

$$F_{Z_2}(z) = P[Q_X(U_1) \cdot Q_Y(U_2) \leqslant z] = P\left[U_2 \leqslant F_Y\left(\frac{z}{Q_X(U_1)}\right)\right]$$

$$= \int_0^1 \int_0^{F_Y\left(\frac{z}{Q_X(u_1)}\right)} 1 \, du_2 du_1 = \int_0^1 F_Y\left(\frac{z}{Q_X(u_1)}\right) du_1 = \int_0^1 u_2 \, du_1, \tag{4.22}$$

co dowodzi prawdziwości tezy, że funkcja opisująca pole pod krzywą $z = Q_X(u_1) \cdot Q_Y(u_2)$ (dla ustalonego z) w kwadracie jednostkowym jest dystrybuantą zmiennej losowej $Z_2 = X \cdot Y$

Ponadto zauważmy, że jeśli

$$0 \leqslant F_Y\left(\frac{z}{Q_X(u_1)}\right) \leqslant 1,$$

to

$$0 \leqslant F_X\left(\frac{z}{Q_Y(1)}\right) \leqslant u_1 \leqslant F_X\left(\frac{z}{Q_Y(0)}\right) \leqslant 1. \tag{4.23}$$

Na podstawie (4.22) i (4.23) mamy więc

$$F_{Z_2}(z) = \int_0^1 u_2 \, du_1 = \int_0^1 F_Y\left(\frac{z}{Q_X(u_1)}\right) du_1 = \int_0^{F_X\left(\frac{z}{Q_Y(1)}\right)} F_Y\left(\frac{z}{Q_X(u_1)}\right) du_1$$

$$+ \int_{F_X\left(\frac{z}{Q_Y(1)}\right)}^{F_X\left(\frac{z}{Q_Y(0)}\right)} F_Y\left(\frac{z}{Q_X(u_1)}\right) du_1 + \int_{F_X\left(\frac{z}{Q_Y(0)}\right)}^1 F_Y\left(\frac{z}{Q_X(u_1)}\right) du_1 = I_{21} + I_{22} + I_{23}. \tag{4.24}$$

Korzystając z analogicznych metod i rachunków jak w przypadku różnicy zmiennych losowych dostajemy do rozważenia następujące cztery przypadki reprezentacji dystrybuanty iloczynu zmiennych losowych.

1° Dla $Q_X(0)\cdot Q_Y(0) \leqslant z < \min(Q_X(0)\cdot Q_Y(1), Q_X(1)\cdot Q_Y(0))$

$$F_{Z_2}(z) = \int_0^{F_X\left(\frac{z}{Q_Y(0)}\right)} F_Y\left(\frac{z}{Q_X(u_1)}\right) du_1. \tag{4.25}$$

2° Dla $Q_X(1)\cdot Q_Y(0) \leqslant z < Q_X(0)\cdot Q_Y(1)$

$$F_{Z_2}(z) = \int_0^1 F_Y\left(\frac{z}{Q_X(u_1)}\right) du_1. \tag{4.26}$$

3° Dla $Q_X(0)\cdot Q_Y(1) \leqslant z < Q_X(1)\cdot Q_Y(0)$

$$F_{Z_2}(z) = F_X\left(\frac{z}{Q_Y(1)}\right) + \int_{F_X\left(\frac{z}{Q_Y(1)}\right)}^{F_X\left(\frac{z}{Q_Y(0)}\right)} F_Y\left(\frac{z}{Q_X(u_1)}\right) du_1. \tag{4.27}$$

4° Dla $\max(Q_X(0)\cdot Q_Y(1), Q_X(1)\cdot Q_Y(0)) \leqslant z < Q_X(1)\cdot Q_Y(1)$

$$F_{Z_2}(z) = F_X\left(\frac{z}{Q_Y(1)}\right) + \int_{F_X\left(\frac{z}{Q_Y(1)}\right)}^{1} F_Y\left(\frac{z}{Q_X(u_1)}\right) du_1. \tag{4.28}$$

Korzystając z (4.25), (4.26), (4.27), (4.28) i podstawienia $u_1 = F_X(v)$ otrzymujemy postać dystrybuanty zmiennej losowej $Z_2 = X\cdot Y$

$$F_{Z_2}(z) = \begin{cases} 0, & \text{gdy} \quad z < Q_X(0)\cdot Q_Y(0) \\ \int_{Q_X(0)}^{\frac{z}{Q_Y(0)}} F_Y\left(\frac{z}{v}\right) f_X(v)\, dv, & \text{gdy} \quad z \geqslant Q_X(0)\cdot Q_Y(0),\ z < Q_X(0)\cdot Q_Y(1),\ z < Q_X(1)\cdot Q_Y(0) \\ \int_{Q_X(0)}^{Q_X(1)} F_Y\left(\frac{z}{v}\right) f_X(v)\, dv, & \text{gdy} \quad z \geqslant Q_X(1)\cdot Q_Y(0),\ z < Q_X(0)\cdot Q_Y(1) \\ F_X\left(\frac{z}{Q_Y(1)}\right) + \int_{\frac{z}{Q_Y(1)}}^{\frac{z}{Q_Y(0)}} F_Y\left(\frac{z}{v}\right) f_X(v)\, dv, & \text{gdy} \quad z \geqslant Q_X(0)\cdot Q_Y(1),\ z < Q_X(1)\cdot Q_Y(0) \\ F_X\left(\frac{z}{Q_Y(1)}\right) + \int_{\frac{z}{Q_Y(1)}}^{Q_X(1)} F_Y\left(\frac{z}{v}\right) f_X(v)\, dv, & \text{gdy} \quad z \geqslant Q_X(0)\cdot Q_Y(1),\ z \geqslant Q_X(1)\cdot Q_Y(0),\ z < Q_X(1)\cdot Q_Y(1) \\ 1, & \text{gdy} \quad z \geqslant Q_X(1)\cdot Q_Y(1) \end{cases}. \tag{4.29}$$

□

Ostatnim punktem tego podrozdziału jest wyprowadzenie wzoru na dystrybuantę ilorazu niezależnych, dodatnich zmiennych losowych.

Twierdzenie 4.3. *Niech $X > 0$ oraz $Y > 0$ będą zmiennymi losowymi o rozkładach zdefiniowanych, odpowiednio, za pomocą dystrybuanty F_X oraz F_Y. Niech Q_X oraz Q_Y będą, odpowiednio, funkcjami kwantylowymi zmiennych losowych X oraz Y z założeniem, że $Q_X(u_1) > 0$ dla $u_1 \in (0,1]$ oraz $Q_Y(u_2) > 0$ dla $u_2 \in [0,1]$. Dodatkowo zakładamy, że istnieje granica*

$$\lim_{u_1 \to 1^-, u_2 \to 1^-} \left[\frac{Q_X(u_1)}{Q_Y(u_2)} \right].$$

Wtedy funkcja opisująca pole nad krzywą $z = \dfrac{Q_X(u_1)}{Q_Y(u_2)}$ (dla ustalonego z) w kwadracie jednostkowym jest dystrybuantą zmiennej losowej $Z_3 = \dfrac{X}{Y}$.

Dowód. Ponownie, jeśli (u_1, u_2), gdzie $0 < u_1, u_2 < 1$, jest wartością zmiennej losowej (U_1, U_2) o rozkładzie jednostajnym na kwadracie jednostkowym, to $\dfrac{Q_X(u_1)}{Q_Y(u_2)}$ zwraca losową liczbę z rozkładu zmiennej losowej $Z_3 = \dfrac{X}{Y}$.

Niech $z = \dfrac{Q_X(u_1)}{Q_Y(u_2)}$. Wtedy $u_2 = F_Y\left(\dfrac{Q_X(u_1)}{z} \right)$. Stąd

$$F_{Z_3}(z) = P\left[\frac{Q_X(U_1)}{Q_Y(U_2)} \leqslant z \right] = P\left[U_2 \geqslant F_Y\left(\frac{Q_X(U_1)}{z} \right) \right]$$

$$= \int_0^1 \int_{F_Y\left(\frac{Q_X(u_1)}{z}\right)}^1 1 du_2 du_1 = 1 - \int_0^1 F_Y\left(\frac{Q_X(u_1)}{z} \right) du_1, \tag{4.30}$$

co dowodzi, że funkcja opisująca pole nad krzywą $z = \dfrac{Q_X(u_1)}{Q_Y(u_2)}$ (dla ustalonego z) w kwadracie jednostkowym jest dystrybuantą zmiennej losowej $Z_3 = \dfrac{X}{Y}$.

Wzór na podstać dystrybuanty F_{Z_3} w punkcie z wyznaczamy podstępując analogicznie jak w przypadku poprzednich dwóch twierdzeń.

Zauważmy, że z faktu

$$0 \leqslant F_Y\left(\frac{Q_X(u_1)}{z} \right) \leqslant 1$$

wynika, że

$$0 \leqslant F_X(zQ_Y(0)) \leqslant u_1 \leqslant F_X(zQ_Y(1)) \leqslant 1, \tag{4.31}$$

co pozwala stwierdzić, że

$$\int_0^1 u_2 \, du_1 = \int_0^1 F_Y\left(\frac{Q_X(u_1)}{z} \right) du_1 = \int_0^{F_X(zQ_Y(0))} F_Y\left(\frac{Q_X(u_1)}{z} \right) du_1$$

$$+ \int_{F_X(zQ_Y(0))}^{F_X(zQ_Y(1))} F_Y\left(\frac{Q_X(u_1)}{z} \right) du_1 + \int_{F_X(zQ_Y(1))}^1 F_Y\left(\frac{Q_X(u_1)}{z} \right) du_1 = I_{31} + I_{32} + I_{33}. \tag{4.32}$$

Podobnie jak w przypadku różnicy i iloczynu zmiennych losowych musimy rozważyć cztery postaci wzoru dystrybuanty F_{Z_3}.

1° Dla $\dfrac{Q_X(0)}{Q_Y(1)} \leqslant z < \min\left(\dfrac{Q_X(0)}{Q_Y(0)}, \dfrac{Q_X(1)}{Q_Y(1)}\right)$

$$F_{Z_3}(z) = F_X(zQ_Y(1)) - \int_0^{F_X(zQ_Y(1))} F_Y\left(\frac{Q_X(u_1)}{z}\right) du_1. \tag{4.33}$$

2° Dla $\dfrac{Q_X(1)}{Q_Y(1)} \leqslant z < \dfrac{Q_X(0)}{Q_Y(0)}$

$$F_{Z_3}(z) = 1 - \int_0^1 F_Y\left(\frac{Q_X(u_1)}{z}\right) du_1. \tag{4.34}$$

3° Dla $\dfrac{Q_X(0)}{Q_Y(0)} \leqslant z < \dfrac{Q_X(1)}{Q_Y(1)}$

$$F_{Z_3}(z) = F_X(zQ_Y(1)) - \int_{F_X(zQ_Y(0))}^{F_X(zQ_Y(1))} F_Y\left(\frac{Q_X(u_1)}{z}\right) du_1. \tag{4.35}$$

4° Dla $\max\left(\dfrac{Q_X(0)}{Q_Y(0)}, \dfrac{Q_X(1)}{Q_Y(1)}\right) \leqslant z < \dfrac{Q_X(1)}{Q_Y(0)}$

$$F_{Z_3}(z) = 1 - \int_{F_X(zQ_Y(0))}^1 F_Y\left(\frac{Q_X(u_1)}{z}\right) du_1. \tag{4.36}$$

Korzystając z (4.33), (4.34), (4.35), (4.36) i podstawienia $u_1 = F_X(v)$ otrzymujemy postać dystrybuanty F_{Z_3} w punkcie z

$$F_{Z_3}(z) = \begin{cases} 0, & \text{gdy} \quad z < \frac{Q_X(0)}{Q_Y(1)} \\ F_X(zQ_Y(1)) - \int_{Q_X(0)}^{zQ_Y(1))} F_Y\left(\frac{v}{z}\right) f_X(v)\, dv, & \text{gdy} \quad z \geqslant \frac{Q_X(0)}{Q_Y(1)}, \; z < \frac{Q_X(0)}{Q_Y(0)}, \; z < \frac{Q_X(1)}{Q_Y(1)} \\ 1 - \int_{Q_X(0)}^{Q_X(1)} F_Y\left(\frac{v}{z}\right) f_X(v)\, dv, & \text{gdy} \quad z \geqslant \frac{Q_X(1)}{Q_Y(1)}, \; z < \frac{Q_X(0)}{Q_Y(0)} \\ F_X(zQ_Y(1)) - \int_{zQ_Y(0)}^{zQ_Y(1)} F_Y\left(\frac{v}{z}\right) f_X(v)\, dv, & \text{gdy} \quad z \geqslant \frac{Q_X(0)}{Q_Y(0)}, \; z < \frac{Q_X(1)}{Q_Y(1)} \\ 1 - \int_{zQ_Y(0)}^{Q_X(1)} F_Y\left(\frac{v}{z}\right) f_X(v)\, dv, & \text{gdy} \quad z \geqslant \frac{Q_X(0)}{Q_Y(0)}, \; z \geqslant \frac{Q_X(1)}{Q_Y(1)}, \; z < \frac{Q_X(1)}{Q_Y(0)} \\ 1, & \text{gdy} \quad z \geqslant \frac{Q_X(1)}{Q_Y(0)} \end{cases}. \tag{4.37}$$

□

4.3. Aproksymacja i estymacja rozkładu różnicy niezależnych zmiennych losowych

W tym podrozdziale Twierdzenie 4.1 oraz metoda trapezów zostaną wykorzystane do wyznaczenia wzoru aproksymującego dystrybuantę różnicy niezależnych, nieujemnych zmiennych losowych X,Y. Metoda trapezów została wybrana ze względu na równowagę między prostotą obliczeniową, a wystarczającą dokładnością numeryczną, co czyni ją dobrze dopasowaną do konstrukcji proponowanych w dalszej części wzorów aproksymujących ((4.40), (4.42)), na podstawie których zostanie wyznaczony estymator nieparametryczny dystrybuanty różnicy niezależnych, nieujemnych zmiennych losowych.

4.3.1. Aproksymacja rozkładu

W dalszych rozważaniach używane będzie następujące oznaczenie

$$Q_Y^{\delta}(1) := Q_Y(1-\delta) \quad \text{dla} \quad 0 \leqslant \delta < 1.$$

W przypadku, gdy $Q_Y(1) < \infty$, przyjmujemy $\delta = 0$, a w przypadku, gdy $Q_Y(1) = \infty$, przyjmujemy $\delta > 0$. Zauważmy, że w przypadku, gdy $Q_Y(1) = \infty$ mamy

$$\forall_{\varepsilon>0}\exists_{\delta>0} \quad 1-\varepsilon < F_X(Q_Y^{\delta}(1)) < 1. \tag{4.38}$$

Zatem dla dowolnego $\varepsilon > 0$ oraz $z \in \mathbb{R}_+ \cup \{0\}$ otrzymujemy następujące oszacowanie

$$F_X(z+Q_Y(1)) - F_X(z+Q_Y^{\delta}(1)) \leqslant 1 - F_X(z+Q_Y^{\delta}(1)) \leqslant 1 - F_X(Q_Y^{\delta}(1)) < \varepsilon. \tag{4.39}$$

Ustalmy $z \geqslant 0$ i rozważmy przedział

$$[F_X(z), F_X(z+Q_Y^{\delta}(1))]$$

jako dziedzinę całkowania. Dokonujemy podziału tego przedziału następującymi punktami

$$\left\{F_X(z), F_X\left(z+\frac{Q_Y^{\delta}(1)}{m}\right), F_X\left(z+\frac{2Q_Y^{\delta}(1)}{m}\right), ..., F_X(z+Q_Y^{\delta}(1))\right\}.$$

Następnie obliczamy wartości u_2 zgodnie ze wzorem (4.1) dla wymienionych punktów

$$\left\{F_Y(0), F_Y\left(\frac{Q_Y^{\delta}(1)}{m}\right), F_Y\left(\frac{2Q_Y^{\delta}(1)}{m}\right), ..., F_Y(Q_Y^{\delta}(1))\right\}.$$

Korzystając z metody trapezów oraz faktu, że X i Y są niezależnymi zmiennymi losowymi, otrzymujemy przybliżenie $F_{Z_1}(z)$ dane następującym wzorem

$$F_{Z_{1m}}(z) = F_X(z+Q_Y^{\delta}(1)) - \frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(z+\frac{(i+1)\cdot Q_Y^{\delta}(1)}{m}\right) - F_X\left(z+\frac{i\cdot Q_Y^{\delta}(1)}{m}\right)\right)$$

$$\left(F_Y\left(\frac{i\cdot Q_Y^{\delta}(1)}{m}\right)+F_Y\left(\frac{(i+1)\cdot Q_Y^{\delta}(1)}{m}\right)\right), \tag{4.40}$$

gdzie m oznacza liczbę podprzedziałów przedziału $[F_X(z),F_X(z+Q_Y^{\delta}(1))]$.

Rozważmy, teraz przypadek $z<0$. Dla $Q_Y(1)=\infty$ będziemy wykorzystywać poniższy fakt

$$\forall_{\varepsilon>0}\forall_{-Q_Y(1)<z<0}\exists_{0<\delta\leqslant 1-F_Y(Q_X(1-\varepsilon)-z)}\ F_X(z+Q_Y(1))-F_X(z+Q_Y^{\delta}(1))\leqslant\varepsilon. \tag{4.41}$$

Podobnie jak w przypadku $z\geqslant 0$, dzielimy przedział $[F_X(0),F_X(z+Q_Y^{\delta}(1))]$ na m podprzedziałów za pomocą punktów

$$\left\{F_X(0),F_X\left(\frac{z+Q_Y^{\delta}(1)}{m}\right),F_X\left(\frac{2(z+Q_Y^{\delta}(1))}{m}\right),...,F_X(z+Q_Y^{\delta}(1))\right\}$$

i zgodnie ze wzorem (4.1) otrzymujemy wartości w tych punktach

$$\left\{F_Y(-z),F_Y\left(\frac{-z(m-1)Q_Y^{\delta}(1)}{m}\right),F_Y\left(\frac{-z(m-2)+2Q_Y^{\delta}(1)}{m}\right),...,F_Y(Q_Y^{\delta}(1))\right\}.$$

Stąd, dla $z<0$, mamy następujące przybliżenie

$$F_{Z_{1m}}(z)=F_X(z+Q_Y^{\delta}(1))-\frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(\frac{(i+1)\cdot(z+Q_Y^{\delta}(1))}{m}\right)-F_X\left(\frac{i\cdot(z+Q_Y^{\delta}(1))}{m}\right)\right)$$
$$\left(F_Y\left(\frac{-z(m-i)+i\cdot Q_Y^{\delta}(1)}{m}\right)+F_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y^{\delta}(1)}{m}\right)\right). \tag{4.42}$$

Wzory (4.40) i (4.42) aproksymują rozkład różnicy niezależnych, nieujemnych zmiennych losowych.

4.3.2. Błąd aproksymacji

Poniższe twierdzenie przedstawia oszacowanie błędu przybliżenia $F_{Z_1}(z)$ przez $F_{Z_{1m}}(z)$.

Twierdzenie 4.4. *(Dudek i Kuczmaszewska [14]) Niech $\varepsilon>0$. Niech $X\geqslant 0$ będzie zmienną losową o co najmniej raz różniczkowalnej dystrybuancie F_X na przedziale $(Q_X(0),Q_X(1))$. Niech $Y\geqslant 0$ będzie zmienną losową o co najmniej dwukrotnie różniczkowalnej dystrybuancie F_Y na przedziale $(Q_Y(0),Q_Y(1))$. Dodatkowo załóżmy, że F_X' i F_Y'' są ciągłe. Wówczas:*

(i) dla $z\geqslant 0$ istnieje $0\leqslant\delta<1$ taka, że oszacowanie błędu aproksymacji $F_{Z_1}(z)$ przez $F_{Z_{1m}}(z)$ wyraża się wzorem

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)|\leqslant 2\varepsilon+\frac{M_1(z)[M_2(z)]^3(Q_Y^{\delta}(1))^3}{12m^2},$$

gdzie

$$M_1(z)=\max_{F_X(z)\leqslant\eta\leqslant F_X(z+Q_Y^{\delta}(1))}|F_Y''(Q_X(\eta)-z)|,\quad M_2(z)=\max_{z\leqslant\tau\leqslant z+Q_Y^{\delta}(1)}|F_X'(\tau)|$$

oraz m jest liczbą podprzedziałów;

(ii) dla $z<0$ istnieje $0\leqslant\delta\leqslant 1-F_Y(Q_X(1-\varepsilon)-z)$ taka, że oszacowanie błędu aproksymacji $F_{Z_1}(z)$ przez $F_{Z_{1m}}(z)$ wyraża się wzorem

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)|\leqslant 2\varepsilon+\frac{M_1^*(z)[M_2^*(z)]^3(z+Q_Y^\delta(1))^3}{12m^2},$$

gdzie

$$M_1^*(z)=\max_{F_X(0)\leqslant\mu\leqslant F_X(z+Q_Y^\delta(1))}|F_Y''(Q_X(\mu)-z)|\quad oraz\quad M_2^*(z)=\max_{0\leqslant\xi\leqslant z+Q_Y^\delta(1)}|F_X'(\xi)|.$$

Dowód. Zauważmy, że $Q_Y(1)<\infty\implies F_X(z+Q_Y(1))-F_X(z+Q_Y^\delta(1))=0$. Dlatego w dowodzie rozważymy jedynie przypadek, gdy $Q_Y(1)=\infty$. Niech $\varepsilon>0$. Na podstawie (4.38) istnieje $0<\delta<1$ taka, że (4.39) jest prawdziwe dla dowolnego $z\geqslant 0$. Zatem dostajemy

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)|=$$

$$\left|F_X(z+Q_Y^\delta(1))-F_X(z+Q_Y(1))-\frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(z+\frac{(i+1)\cdot Q_Y^\delta(1)}{m}\right)-F_X\left(z+\frac{i\cdot Q_Y^\delta(1)}{m}\right)\right)\right.$$
$$\left.\left(F_Y\left(\frac{i\cdot Q_Y^\delta(1)}{m}\right)+F_Y\left(\frac{(i+1)\cdot Q_Y^\delta(1)}{m}\right)\right)+\int_{F_X(z)}^{F_X(z+Q_Y(1))}F_Y(Q_X(u_1)-z)du_1\right|$$
$$\leqslant|F_X(z+Q_Y^\delta(1))-F_X(z+Q_Y(1))|+\int_{F_X(z+Q_Y^\delta(1))}^{F_X(z+Q_Y(1))}1du_1$$
$$+\left|\int_{F_X(z)}^{F_X(z+Q_Y^\delta(1))}F_Y(Q_X(u_1)-z)du_1-\frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(z+\frac{(i+1)\cdot Q_Y^\delta(1)}{m}\right)-F_X\left(z+\frac{i\cdot Q_Y^\delta(1)}{m}\right)\right)\right.$$
$$\left.\left(F_Y\left(\frac{i\cdot Q_Y^\delta(1)}{m}\right)+F_Y\left(\frac{(i+1)\cdot Q_Y^\delta(1)}{m}\right)\right)\right|.$$

Na podstawie (4.39) uzyskujemy oszacowanie

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)|\leqslant 2\varepsilon+\left|\sum_{i=0}^{m-1}\int_{F_X\left(z+\frac{i\cdot Q_Y^\delta(1)}{m}\right)}^{F_X\left(z+\frac{(i+1)\cdot Q_Y^\delta(1)}{m}\right)}F_Y(Q_X(u_1)-z)du_1\right.$$
$$-\frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(z+\frac{(i+1)\cdot Q_Y^\delta(1)}{m}\right)-F_X\left(z+\frac{i\cdot Q_Y^\delta(1)}{m}\right)\right)$$
$$\left.\left(F_Y\left(\frac{i\cdot Q_Y^\delta(1)}{m}\right)+F_Y\left(\frac{(i+1)\cdot Q_Y^\delta(1)}{m}\right)\right)\right|.$$

Korzystając z równania (5.1.4) w pracy Atkinsona [1] otrzymujemy

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)|\leqslant 2\varepsilon+\sum_{i=0}^{m-1}\frac{\left(F_X\left(z+\frac{(i+1)\cdot Q_Y^\delta(1)}{m}\right)-F_X\left(z+\frac{i\cdot Q_Y^\delta(1)}{m}\right)\right)^3}{12}|F_Y''(Q_X(\eta_i)-z)|,$$

gdzie $\eta_i \in \left[F_X\left(z+\frac{i\cdot Q_Y^\delta(1)}{m}\right), F_X\left(z+\frac{(i+1)\cdot Q_Y^\delta(1)}{m}\right)\right]$ i na mocy twierdzenia o wartości średniej mamy oszacowanie

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)| \leqslant 2\varepsilon + \sum_{i=0}^{m-1} \frac{(F_X'(\tau_i)Q_Y^\delta(1))^3}{12m^3}|F_Y''(Q_X(\eta_i)-z)|,$$

gdzie $\tau_i \in \left[z+\frac{i\cdot Q_Y^\delta(1)}{m}, z+\frac{(i+1)\cdot Q_Y^\delta(1)}{m}\right]$.

Z faktu, że F_Y'' oraz F_X' są ciągłe dostajemy, że

$$|F_Y''(Q_X(\eta_i)-z)| \leqslant \max_{F_X\left(z+\frac{i\cdot Q_Y^\delta(1)}{m}\right)\leqslant\eta_i\leqslant F_X\left(z+\frac{(i+1)\cdot Q_Y^\delta(1)}{m}\right)} |F_Y''(Q_X(\eta_i)-z)|$$

$$\leqslant M_1(z) = \max_{F_X(z)\leqslant\eta\leqslant F_X(z+Q_Y^\delta(1))} |F_Y''(Q_X(\eta)-z)|$$

oraz

$$|F_X'(\tau_i)| \leqslant \max_{z+\frac{i\cdot Q_Y^\delta(1)}{m}\leqslant\tau_i\leqslant z+\frac{(i+1)\cdot Q_Y^\delta(1)}{m}} |F_X'(\tau_i)| \leqslant M_2(z) = \max_{z\leqslant\tau\leqslant z+Q_Y^\delta(1)} |F_X'(\tau)|.$$

Zatem oszacowanie błędu przybliżenia można zapisać w postaci

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)| \leqslant 2\varepsilon + \frac{M_1(z)[M_2(z)]^3(Q_Y^\delta(1))^3}{12m^2}, \tag{4.43}$$

co kończy dowód dla $z \geqslant 0$.

Rozważmy teraz przypadek $z<0$. Analogicznie, jak w przypadku $z\geqslant 0$, dostajemy

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)| = |F_X(z+Q_Y^\delta(1))-F_X(z+Q_Y(1))$$

$$-\frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(\frac{(i+1)\cdot(z+Q_Y^\delta(1))}{m}\right)-F_X\left(\frac{i\cdot(z+Q_Y^\delta(1))}{m}\right)\right)$$

$$\left(F_Y\left(\frac{-z(m-i)+i\cdot Q_Y^\delta(1)}{m}\right)+F_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y^\delta(1)}{m}\right)\right)$$

$$+\int_0^{F_X(z+Q_Y(1))} F_Y(Q_X(u_1)-z)du_1| \leqslant |F_X(z+Q_Y^\delta(1))-F_X(z+Q_Y(1))|$$

$$+\int_{F_X(z+Q_Y^\delta(1))}^{F_X(z+Q_Y(1))} 1du_1 + \left|\int_0^{F_X(z+Q_Y^\delta(1))} F_Y(Q_X(u_1)-z)du_1\right.$$

$$-\frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(\frac{(i+1)\cdot(z+Q_Y^\delta(1))}{m}\right)-F_X\left(\frac{i\cdot(z+Q_Y^\delta(1))}{m}\right)\right)$$

$$\left.\left(F_Y\left(\frac{-z(m-i)+i\cdot Q_Y^\delta(1)}{m}\right)+F_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y^\delta(1)}{m}\right)\right)\right|.$$

Stąd na podstawie (4.41) otrzymujemy

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)| \leqslant 2\varepsilon + \left| \sum_{i=0}^{m-1} \int_{F_X(\frac{i\cdot(z+Q_Y^\delta(1))}{m})}^{F_X(\frac{(i+1)\cdot(z+Q_Y^\delta(1))}{m})} F_Y(Q_X(u_1)-z)du_1 \right.$$

$$-\frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(\frac{(i+1)\cdot(z+Q_Y^\delta(1))}{m}\right)-F_X\left(\frac{i\cdot(z+Q_Y^\delta(1))}{m}\right)\right)$$

$$\left.\left(F_Y\left(\frac{-z(m-i)+i\cdot Q_Y^\delta(1)}{m}\right)+F_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y^\delta(1)}{m}\right)\right)\right|.$$

Ponownie korzystając z równania (5.1.4) z Atkinsona [1] mamy oszacowanie

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)| \leqslant 2\varepsilon + \sum_{i=0}^{m-1} \frac{\left(F_X(\frac{(i+1)\cdot(z+Q_Y^\delta(1))}{m})-F_X(\frac{i\cdot(z+Q_Y^\delta(1))}{m})\right)^3}{12}|F_Y''(Q_X(\mu_i)-z)|,$$

gdzie $\mu_i \in \left[F_X\left(\frac{i\cdot(z+Q_Y^\delta(1))}{m}\right), F_X\left(\frac{(i+1)\cdot(z+Q_Y^\delta(1))}{m}\right)\right]$ i na mocy twierdzenia o wartości średniej

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)| \leqslant 2\varepsilon + \sum_{i=0}^{m-1} \frac{\left(F_X'(\xi_i)(z+Q_Y^\delta(1))\right)^3}{12m^3}|F_Y''(Q_X(\mu_i)-z)|,$$

gdzie $\xi_i \in \left[\frac{i\cdot(z+Q_Y^\delta(1))}{m}, \frac{(i+1)\cdot(z+Q_Y^\delta(1))}{m}\right]$.

Dalej mamy

$$|F_Y''(Q_X(\mu_i)-z)| \leqslant \max_{F_X\left(\frac{i\cdot(z+Q_Y^\delta(1))}{m}\right)\leqslant\mu_i\leqslant F_X\left(\frac{(i+1)\cdot(z+Q_Y^\delta(1))}{m}\right)} |F_Y''(Q_X(\mu_i)-z)|$$

$$\leqslant M_1^*(z) = \max_{F_X(0)\leqslant\mu\leqslant F_X(z+Q_Y^\delta(1))} |F_Y''(Q_X(\mu)-z)|$$

oraz

$$|F_X'(\xi_i)| \leqslant \max_{\frac{i\cdot(z+Q_Y(1))}{m}\leqslant\xi_i\leqslant\frac{(i+1)\cdot(z+Q_Y(1))}{m}} |F_X'(\xi_i)| \leqslant M_2^*(z) = \max_{0\leqslant\xi\leqslant z+Q_Y(1)} |F_X'(\xi)|$$

i ostatecznie oszacowanie błędu przybliżenia dla $z<0$ możemy wyrazić w postaci

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)| \leqslant 2\varepsilon + \frac{M_1^*(z)[M_2^*(z)]^3(z+Q_Y^\delta(1))^3}{12m^2}. \tag{4.44}$$

□

Uwaga 4.1. Zauważmy, że oba oszacowania błędu przybliżenia zbiegają do zera, gdy rośnie liczba punktów podziału ($m\to\infty$) i $\varepsilon\to 0$.

4.3.3. Estymator nieparametryczny

W tym podrozdziale wyprowadzimy nieparametryczny estymator dystrybuanty różnicy niezależnych zmiennych losowych, korzystając ze wzorów aproksymujących (4.40) i (4.42).

Niech $\{X_1,\ldots,X_{n_1}\}$ oraz $\{Y_1,\ldots,Y_{n_2}\}$ będą niezależnymi próbami losowymi o liczebnościach, odpowiednio, n_1 i n_2 pobranymi z populacji o rozkładach zdefiniowanych za pomocą dystrybuant, odpowiednio, F_X i F_Y. Niech $\hat{F}_X$ oraz $\hat{F}_Y$ będą empirycznymi dystrybuantami odpowiadającymi F_X i F_Y oraz niech $Q_Y^*(1) := \max\{Y_i\colon 1 \leqslant i \leqslant n_2\}$.

Nieparametryczny estymator $F_{Z_1}(z)$, gdzie $Z_1 = X - Y$, dany jest wzorem dla

- $z \geqslant 0$

$$\hat{F}_{Z_1}(z) = \hat{F}_X(z+Q_Y^*(1)) - \frac{1}{2}\sum_{i=0}^{m-1}\left(\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right) - \hat{F}_X\left(z+\frac{i\cdot Q_Y^*(1)}{m}\right)\right)$$
$$\left(\hat{F}_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right) + \hat{F}_Y\left(\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)\right) \tag{4.45}$$

- $z < 0$

$$\hat{F}_{Z_1}(z) = \hat{F}_X(z+Q_Y^*(1)) - \frac{1}{2}\sum_{i=0}^{m-1}\left(\hat{F}_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right) - \hat{F}_X\left(\frac{i\cdot(z+Q_Y^*(1))}{m}\right)\right)$$
$$\left(\hat{F}_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right) + \hat{F}_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y^*(1)}{m}\right)\right). \tag{4.46}$$

Wykażemy dalej, że $\hat{F}_{Z_1}$ jest estymatorem silnie zgodnym.

Twierdzenie 4.5. *(Dudek i Kuczmaszewska [14]) Niech $Z_1 = X - Y$, gdzie $X \geqslant 0$ oraz $Y \geqslant 0$ są zmiennymi losowymi o ciągłych dystrybuantach, odpowiednio, F_X oraz F_Y spełniającymi założenia Twierdzenia 4.4. Niech $\hat{F}_{Z_1}(z)$ będzie zdefiniowany wzorem (4.45) dla $z \geqslant 0$ i wzorem (4.46) dla $z < 0$. Wówczas*

$$|\hat{F}_{Z_1}(z) - F_{Z_1}(z)| \to 0 \quad p.p.,$$

gdy $n_1, n_2 \to \infty$ dla każdego $z \in \mathbb{R}$.

Dowód. Zauważmy, że

$$|\hat{F}_{Z_1}(z) - F_{Z_1}(z)| \leqslant |\hat{F}_{Z_1}(z) - F_{Z_{1m}}(z)| + |F_{Z_{1m}}(z) - F_{Z_1}(z)| \quad \text{dla} \quad m \in \mathbb{N}.$$

Rozważmy cztery przypadki

1. Niech $z \geqslant 0$ i $Q_Y(1) = \infty$. Wprowadźmy oznaczenia: $\varepsilon = \varepsilon(n_2) = 1 - F_X(Q_Y^*(1))$ oraz $\delta = \delta(n_2) = 1 - F_Y(Q_Y^*(1))$. Wówczas

$$Q_Y^{\delta}(1) = Q_Y^*(1). \tag{4.47}$$

Łatwo zauważyć, że $\varepsilon \to 0$ oraz $\delta \to 0$, gdy $n_2 \to \infty$.

Stąd na podstawie Twierdzenia 4.6(i) dostajemy, że

$$|F_{Z_{1m}}(z) - F_{Z_1}(z)| \leqslant 2\varepsilon + \frac{M_1(z)[M_2(z)]^3[Q_Y^{\delta}(1)]^3}{12m^2} \leqslant 2\varepsilon + \frac{M_1(z)[M_2(z)]^3[Q_Y^{*}(1)]^3}{12m^2}.$$

Dla ustalonego n_2 i $z \geqslant 0$ mamy

$$\lim_{m\to\infty} \frac{M_1(z)[M_2(z)]^3[Q_Y^{*}(1)]^3}{12m^2} = 0$$

$$\iff \forall_{\varepsilon>0}\exists_{m_0=m_0(z,n_2)}\forall_{m\geqslant m_0} \frac{M_1(z)[M_2(z)]^3[Q_Y^{*}(1)]^3}{12m^2} < \varepsilon.$$

Stąd

$$\forall_{n_2}\forall_{z\geqslant 0}\exists_{\varepsilon}\exists_{m_0}\forall_{m\geqslant m_0}|F_{Z_{1m}}(z) - F_{Z_1}(z)| < 3\varepsilon,$$

co ostatecznie daje

$$|F_{Z_{1m}}(z) - F_{Z_1}(z)| \to 0 \quad \text{gdy} \quad n_2 \to \infty \quad \text{dla} \quad z \geqslant 0. \tag{4.48}$$

Teraz, dla $z \geqslant 0$ wykorzystując (4.47) dostajemy

$$\hat{F}_{Z_1}(z) - F_{Z_{1m}}(z) = \hat{F}_X(z+Q_Y^{*}(1)) - F_X(z+Q_Y^{\delta}(1))$$

$$+\frac{1}{2}\sum_{i=0}^{m-1}\left\{\left(F_X\left(z+\frac{(i+1)\cdot Q_Y^{\delta}(1)}{m}\right)-F_X\left(z+\frac{i\cdot Q_Y^{\delta}(1)}{m}\right)\right)\left(F_Y\left(\frac{i\cdot Q_Y^{\delta}(1)}{m}\right)+F_Y\left(\frac{(i+1)\cdot Q_Y^{\delta}(1)}{m}\right)\right)\right.$$

$$\left.-\left(\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^{*}(1)}{m}\right)-\hat{F}_X\left(z+\frac{i\cdot Q_Y^{*}(1)}{m}\right)\right)\left(\hat{F}_Y\left(\frac{i\cdot Q_Y^{*}(1)}{m}\right)+\hat{F}_Y\left(\frac{(i+1)\cdot Q_Y^{*}(1)}{m}\right)\right)\right.$$

$$= \hat{F}_X(z+Q_Y^{*}(1)) - F_X(z+Q_Y^{*}(1))$$

$$+\frac{1}{2}\sum_{i=0}^{m-1}\left\{\left[F_X\left(z+\frac{(i+1)\cdot Q_Y^{*}(1)}{m}\right)F_Y\left(\frac{i\cdot Q_Y^{*}(1)}{m}\right)-\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^{*}(1)}{m}\right)\hat{F}_Y\left(\frac{i\cdot Q_Y^{*}(1)}{m}\right)\right]\right.$$

$$\left.+\left[\hat{F}_X\left(z+\frac{i\cdot Q_Y^{*}(1)}{m}\right)\hat{F}_Y\left(\frac{i\cdot Q_Y^{*}(1)}{m}\right)-F_X\left(z+\frac{i\cdot Q_Y^{*}(1)}{m}\right)F_Y\left(\frac{i\cdot Q_Y^{*}(1)}{m}\right)\right]\right.$$

$$+\left[F_X\left(z+\frac{(i+1)\cdot Q_Y^{*}(1)}{m}\right)F_Y\left(\frac{(i+1)\cdot Q_Y^{*}(1)}{m}\right)-\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^{*}(1)}{m}\right)\hat{F}_Y\left(\frac{(i+1)\cdot Q_Y^{*}(1)}{m}\right)\right]$$

$$\left.+\left[\hat{F}_X\left(z+\frac{i\cdot Q_Y^{*}(1)}{m}\right)\hat{F}_Y\left(\frac{(i+1)\cdot Q_Y^{*}(1)}{m}\right)-F_X\left(z+\frac{i\cdot Q_Y^{*}(1)}{m}\right)F_Y\left(\frac{(i+1)\cdot Q_Y^{*}(1)}{m}\right)\right]\right\}$$

$$= \hat{F}_X(z+Q_Y^{*}(1)) - F_X(z+Q_Y^{*}(1)) + \frac{1}{2}\sum_{i=0}^{m-1}\{A_i(z)+B_i(z)+C_i(z)+D_i(z)\}.$$

Zauważmy, że $|\hat{F}_X(t) - F_X(t)| \to 0$ *p.p.*, gdy $n_1 \to \infty$ oraz $|\hat{F}_Y(t) - F_Y(t)| \to 0$ *p.p.*, gdy $n_2 \to \infty$.

Zatem

$$|\hat{F}_X(z+Q_Y^*(1))-F_X(z+Q_Y^*(1))|=0 \quad p.p., \quad \text{gdy} \quad n_1\to\infty.$$

Powyższe fakty wykorzystamy do oszacowania wyrażenia $A_i(z)$. Zauważmy ponadto, że

$$\begin{aligned}
&|A_i(z)|\\
&=\left|F_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)F_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)-\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)\hat{F}_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)\right|\\
&=\left|F_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)F_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)-\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)F_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)\right.\\
&\left.+\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)F_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)-\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)\hat{F}_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)\right|\\
&=\left|F_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)\left(F_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)-\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)\right)\right.\\
&\left.+\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)\left(F_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)-\hat{F}_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)\right)\right|\\
&\leqslant\left|F_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)\left(F_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)-\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)\right)\right|\\
&+\left|\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)\left(F_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)-\hat{F}_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)\right)\right|.
\end{aligned}$$

Zatem

$$|A_i(z)|\to 0 \ p.p., \quad \text{gdy} \quad n_1,n_2\to\infty. \tag{4.49}$$

Analogicznie, łatwo pokazać, że $|B_i(z)|\to 0$, $|C_i(z)|\to 0$ i $|D_i(z)|\to 0$ $p.p.$, gdy $n_1,n_2\to\infty$. Stąd dla $z\geqslant 0$

$$|\hat{F}_{Z_1}(z)-F_{Z_{1m}}(z)|\to 0 \ p.p., \quad \text{gdy} \quad n_1,n_2\to\infty. \tag{4.50}$$

Ostatecznie na podstawie (4.48) oraz (4.50) dostajemy, że

$$|\hat{F}_{Z_1}(z)-F_{Z_1}(z)|\to 0 \ p.p., \quad \text{gdy} \quad n_1,n_2\to\infty.$$

2. Niech $z\geqslant 0$ oraz $Q_Y(1)<\infty$. Przyjmujemy $\delta=0$ i na podstawie (4.43) mamy

$$|F_{Z_1}(z)-F_{Z_{1m}}(z)|\leqslant\frac{M_1(z)[M_2(z)]^3(Q_Y(1))^3}{12m^2},$$

co dowodzi, że dla dowolnego n_2 i ustalonego $z\geqslant 0$ zachodzi

$$\forall_{\varepsilon>0}\exists_{m_0=m_0(z)}\forall_{m\geqslant m_0}\frac{M_1(z)[M_2(z)]^3(Q_Y(1))^3}{12m^2}<\varepsilon.$$

Ponadto zauważmy, że

$$\hat{F}_{Z_1}(z)-F_{Z_{1m}}(z)=\hat{F}_X(z+Q_Y^*(1))-F_X(z+Q_Y(1))$$

$$+\frac{1}{2}\sum_{i=0}^{m-1}\left\{\left(F_X\left(z+\frac{(i+1)\cdot Q_Y(1)}{m}\right)-F_X\left(z+\frac{i\cdot Q_Y(1)}{m}\right)\right)\left(F_Y\left(\frac{i\cdot Q_Y(1)}{m}\right)+F_Y\left(\frac{(i+1)\cdot Q_Y(1)}{m}\right)\right)\right.$$
$$\left.-\left(\hat{F}_X\left(z+\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)-\hat{F}_X\left(z+\frac{i\cdot Q_Y^*(1)}{m}\right)\right)\left(\hat{F}_Y\left(\frac{i\cdot Q_Y^*(1)}{m}\right)+\hat{F}_Y\left(\frac{(i+1)\cdot Q_Y^*(1)}{m}\right)\right)\right\}.$$

Stosując analogiczne narzędzia jak w przypadku 1. tzn. wykorzystując ciągłość dystrybuant F_X i F_Y, zbieżność dystrybuanty empirycznej do rzeczywistej oraz silną zgodność estymatora kwantyla otrzymujemy, że

$$|\hat{F}_{Z_1}(z)-F_{Z_{1m}}(z)|\to 0,\quad \text{gdy}\quad n_1,n_2\to\infty.$$

3. Niech $-Q_Y^*(1)<z<0$ i $Q_Y(1)=\infty$. Oznaczmy

$$\varepsilon=\varepsilon(n_2,z)=1-F_X(z+Q_Y^*(1))\quad\text{oraz}\quad\delta=\delta(n_2)=1-F_Y(Q_Y^*(1)).$$

Tak jak w przypadku 1. zachodzi $Q_Y^\delta(1)=Q_Y^*(1)$ oraz $\varepsilon\to 0$, $\delta\to 0$, gdy $n_2\to\infty$.
Na podstawie Twierdzenia 4.6(ii) mamy, że

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)|\leqslant 2\varepsilon+\frac{M_1^*(z)[M_2^*(z)]^3(z+Q_Y^\delta(1))^3}{12m^2}\leqslant 2\varepsilon+\frac{M_1^*(z)[M_2^*(z)]^3[Q_Y^*(1)]^3}{12m^2}.$$

Zatem

$$\forall_{n_2}\forall_{-Q_Y^*(1)<z<0}\exists\varepsilon\exists_{m_0=m_0(n_2,z)}\forall_{m\geqslant m_0}|F_{Z_{1m}}(z)-F_{Z_1}(z)|<3\varepsilon.$$

Stąd

$$|F_{Z_{1m}}(z)-F_{Z_1}(z)|\to 0,\quad\text{gdy}\quad n_2\to\infty\quad\text{dla}\quad -Q_Y^*(1)<z<0. \tag{4.51}$$

Analogicznie jak dla przypadku $z\geqslant 0$, na podstawie (4.47) dostajemy równość

$$\hat{F}_{Z_1}(z)-F_{Z_{1m}}(z)=\hat{F}_X(z+Q_Y^*(1))-F_X(z+Q_Y^\delta(1))$$
$$+\frac{1}{2}\sum_{i=0}^{m-1}\left\{\left(F_X\left(\frac{(i+1)\cdot(z+Q_Y^\delta(1))}{m}\right)-F_X\left(\frac{i\cdot(z+Q_Y^\delta(1))}{m}\right)\right)\right.$$
$$\left(F_Y\left(\frac{-z(m-i)+(i+1)\cdot Q_Y^\delta(1)}{m}\right)+F_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y^\delta(1)}{m}\right)\right)$$
$$-\left(\hat{F}_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)-\hat{F}_X\left(\frac{i\cdot(z+Q_Y^*(1))}{m}\right)\right)$$
$$\left.\left(\hat{F}_Y\left(\frac{-z(m-i)+(i+1)\cdot Q_Y^*(1)}{m}\right)+\hat{F}_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y^*(1)}{m}\right)\right)\right\}$$
$$=\hat{F}_X(z+Q_Y^*(1))-F_X(z+Q_Y^*(1))$$
$$+\frac{1}{2}\sum_{i=0}^{m-1}\left\{\left[F_X(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m})F_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)\right.\right.$$
$$\left.\left.-\hat{F}_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)\hat{F}_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)\right]\right.$$

$$
\begin{aligned}
&+\left[\hat{F}_X\left(\frac{i\cdot(z+Q_Y^*(1))}{m}\right)\hat{F}_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)\right.\\
&\left.-F_X\left(\frac{i\cdot(z+Q_Y^*(1))}{m}\right)F_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)\right]\\
&+\left[F_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)F_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y^*(1)}{m}\right)\right.\\
&\left.-\hat{F}_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)\hat{F}_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y^*(1)}{m}\right)\right]\\
&+\left[\hat{F}_X\left(\frac{i\cdot(z+Q_Y^*(1))}{m}\right)\hat{F}_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y^*(1)}{m}\right)\right.\\
&\left.\left.-F_X\left(\frac{i\cdot(z+Q_Y^*(1))}{m}\right)F_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y^*(1)}{m}\right)\right]\right\}\\
&=\hat{F}_X(z+Q_Y^*(1))-F_X(z+Q_Y^*(1))+\frac{1}{2}\sum_{i=0}^{m-1}\{E_i(z)+F_i(z)+G_i(z)+H_i(z)\}.
\end{aligned}
$$

Następnie mamy

$$
\begin{aligned}
|E_i(z)| &= \left|F_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)F_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)\right.\\
&\left.-\hat{F}_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)\hat{F}_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)\right|\\
&=\left|F_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)F_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)\right.\\
&-\hat{F}_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)F_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)\\
&+\hat{F}_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)F_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)\\
&\left.-\hat{F}_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)\hat{F}_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)\right|\\
\leqslant &\left|F_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)\left(F_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)-\hat{F}_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)\right)\right|\\
&+\left|\hat{F}_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right)\left(F_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)-\hat{F}_Y\left(\frac{-z(m-i)+i\cdot Q_Y^*(1)}{m}\right)\right)\right|.
\end{aligned}
$$

Stąd dostajemy, że $|E_i(z)|\to 0$ $p.p.$, gdy $n_1,n_2\to\infty$.

Podobnie dowodzimy, że $|F_i(z)|\to 0$, $|G_i(z)|\to 0$ i $|H_i(z)|\to 0$ $p.p.$, gdy $n_1,n_2\to\infty$.

Na podstawie powyższych rozważań dostajemy, że

$$
|\hat{F}_{Z_1}(z)-F_{Z_{1m}}(z)|\to 0 \ p.p., \quad \text{gdy} \quad n_1,n_2\to\infty. \tag{4.52}
$$

Ostatecznie, na podstawie (4.51) oraz (4.52) mamy

$$
|\hat{F}_{Z_1}(z)-F_{Z_1}(z)|\to 0, \ p.p., \quad \text{gdy} \quad n_1,n_2\to\infty.
$$

4. Niech $-Q_Y^*(1) < z < 0$ i $Q_Y(1) < \infty$. Przyjmujemy $\delta = 0$ i na podstawie (4.44) mamy

$$|F_{Z_{1m}}(z) - F_{Z_1}(z)| \leqslant \frac{M_1^*(z)[M_2^*(z)]^3(z+Q_Y(1))^3}{12m^2},$$

co dowodzi, że dla dowolnego n_2 i ustalonego $z \geqslant 0$ zachodzi

$$\forall_{\varepsilon>0}\exists_{m_0=m_0(z)}\forall_{m\geqslant m_0}\frac{M_1(z)[M_2(z)]^3(Q_Y(1))^3}{12m^2} < \varepsilon.$$

Ponadto zauważmy, że

$$\hat{F}_{Z_1}(z) - F_{Z_{1m}}(z) = \hat{F}_X(z+Q_Y^*(1)) - F_X(z+Q_Y(1))$$
$$+\frac{1}{2}\sum_{i=0}^{m-1}\left\{\left(F_X\left(\frac{(i+1)\cdot(z+Q_Y(1))}{m}\right) - F_X\left(\frac{i\cdot(z+Q_Y(1))}{m}\right)\right)\right.$$
$$\left(F_Y\left(\frac{-z(m-i)+(i+1)\cdot Q_Y(1)}{m}\right) + F_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y(1)}{m}\right)\right)$$
$$-\left(\hat{F}_X\left(\frac{(i+1)\cdot(z+Q_Y^*(1))}{m}\right) - \hat{F}_X\left(\frac{i\cdot(z+Q_Y^*(1))}{m}\right)\right)$$
$$\left.\left(\hat{F}_Y\left(\frac{-z(m-i)+(i+1)\cdot Q_Y^*(1)}{m}\right) + \hat{F}_Y\left(\frac{-z(m-i-1)+(i+1)\cdot Q_Y^*(1)}{m}\right)\right)\right.\!.$$

Ponownie stosując analogiczne narzędzia jak w przypadku pozostałych przypadkach tzn. wykorzystując ciągłość dystrybuant F_X i F_Y, zbieżność dystrybuanty empirycznej do rzeczywistej oraz silną zgodność estymatora kwantyla otrzymujemy, że

$$|\hat{F}_{Z_1}(z) - F_{Z_{1m}}(z)| \to 0, \quad \text{gdy} \quad n_1, n_2 \to \infty.$$

□

Uwaga 4.2. Dla $z \geqslant 0$, korzystając z Lematu 1 z pracy Midhu i in. [32] oraz z faktu, że $\hat{F}(z+Q^*(1)) = 1$, łatwo pokazać, że zmienna losowa $2n_1n_2\widehat{F}_{Z_1}(z)$ ma następującą funkcję masy prawdopodobieństwa:

$$P(2n_1n_2\widehat{F}_{Z_1}(z) = a) = P(2n_1n_2 - (C_1+2C_2) = a) = P(C_1+2C_2 = 2n_1n_2 - a)$$
$$= P(C_1+2C_2 = k) = \begin{cases} \sum\limits_{j=0}^{\frac{1}{4}(2k+(-1)^k-1)} \frac{(n_1n_2)!}{j!(k-2j)!(n_1n_2-k+j)!} q_1^j q_2^{k-2j} q_3^{n_1n_2-k+j}, & \text{gdy } k \leqslant n_1n_2, \\ \sum\limits_{j=k-n_1n_2}^{\frac{1}{4}(2k+(-1)^k-1)} \frac{(n_1n_2)!}{j!(k-2j)!(n_1n_2-k+j)!} q_1^j q_2^{k-2j} q_3^{n_1n_2-k+j}, & \text{gdy} k > n_1n_2, \end{cases}$$

gdzie $k = 2n_1n_2 - a$, $a = 0, 1, ..., 2n_1n_2$,

$$q_1 = \sum_{i=0}^{m-1}\left[\left(F_X\left(z+\frac{(i+1)Q_Y^*(1)}{m}\right) - F_X\left(z+\frac{iQ_Y^*(1)}{m}\right)\right)F_Y\left(\frac{iQ_Y^*(1)}{m}\right)\right],$$

$$q_2 = \sum_{i=0}^{m-1}\left[\left(F_X\left(z+\frac{(i+1)Q_Y^*(1)}{m}\right) - F_X\left(z+\frac{iQ_Y^*(1)}{m}\right)\right)\left(F_Y\left(\frac{(i+1)Q_Y^*(1)}{m}\right) - F_Y\left(\frac{iQ_Y^*(1)}{m}\right)\right)\right]$$

oraz

$$q_3 = 1 - q_1 - q_2.$$

4.3.4. Symulacje

W tej części przedstawione zostaną symulacje, których celem jest zilustrowanie rozważań teoretycznych zawartych w podrozdziale 4.3. W pierwszym kroku z wykorzystaniem metody Monte Carlo i programu *R* studio zostanie zbadany błąd aproksymacji dystrybuanty F_{Z_1} przez $F_{Z_{1m}}$ (wzór (4.40) lub (4.42)). W drugim etapie, ponownie z wykorzystaniem *R* studio, badany będzie estymator dany wzorem (4.45) lub (4.46). Rozważony zostanie przypadek, gdy $n_1 = n_2 = n$ i proponowany estymator zostanie porównany z klasyczną dystrybuantą empiryczną (ECDF) dla $X_i - Y_i$, gdzie $i = 1,2,...,n$. Dodatkowo zaprezentowane zostaną wyniki dla przypadku gdy $n_1 \neq n_2$.

Błąd aproksymacji

Błąd przybliżenia został wyznaczony dla czterech różnych przypadków rozkładów zmiennych losowych X i Y:

1. X ma rozkład wykładniczy z parametrem 1, a Y ma rozkład wykładniczy z parametrem 2,
2. X i Y mają jednakowy rozkład jednostajny na przedziale $[0,1]$,
3. X i Y mają jednakowy rozkład Weibulla z parametrami $(2,1)$,
4. X ma rozkład wykładniczy z parametrem 1, a Y ma rozkład Weibulla z parametrami $(2,1)$.

W każdym przypadku wyznaczono wartości z z zależności $F_{Z_1}(z) = u$, gdzie $u \in \{0,2;0,6;0,8\}$, a $Z_1 = X - Y$. Wyniki symulacji zostały przedstawione w Tabeli 4.1.

Tabela 4.1: Błąd aproksymacji $F_{Z_{1m}}$. **Źródło:** [14]

$X \sim \exp(1), Y \sim \exp(2)$				$X,Y \sim \text{Unif}[0,1]$			
F_{Z_1}	$m=5$	$m=10$	$m=15$	F_{Z_1}	$m=5$	$m=10$	$m=15$
0.2	0.0094	0.0024	0.0010	0.2	0.0000	0.0000	0.0000
0.6	0.0100	0.0026	0.0011	0.6	0.0049	0.0005	0.0002
0.8	0.0054	0.0014	0.0006	0.8	0.0037	0.0012	0.0004
$X,Y \sim \text{Weibull}(2,1)$				$X \sim \exp(1), Y \sim \text{Weibull}(2,1)$			
F_{Z_1}	$m=5$	$m=10$	$m=15$	F_{Z_1}	$m=5$	$m=10$	$m=15$
0.2	0.0008	0.0006	0.0004	0.2	0.0010	0.0009	0.0008
0.6	0.0066	0.0017	0.0008	0.6	0.0107	0.0030	0.0015
0.8	0.0143	0.0031	0.0010	0.8	0.0056	0.0015	0.0008

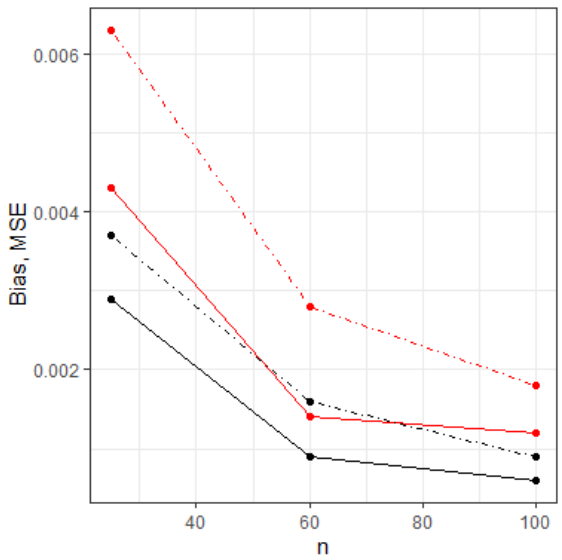
Z wyników przedstawionych w Tabeli 4.1 wynika wspólny wniosek, że dla ustalonej wartości z w każdym przypadku błąd aproksymacji maleje wraz ze wzrostem podprzedziałów m.

Ocena estymatora nieparametrycznego ($n_1 = n_2 = n$)

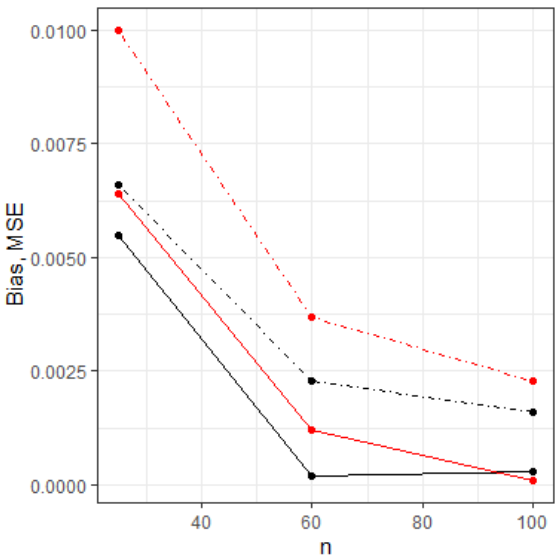
W drugim kroku symulacji przeanalizowano estymator dany wzorami (4.45) i (4.46). W tym celu zostały wygenerowane 1000 razy próby z rozkładów: jednostajnego, wykładniczego oraz Weibulla. Rozmiary prób przyjęto $n \in \{25, 60, 100\}$. W każdym przypadku wyznaczono wartości z z zależności $F_{Z_1}(z) = u$, gdzie $u \in \{0,2; 0,4; 0,6; 0,8\}$, a $Z_1 = X - Y$. Dla każdej próby obliczono estymator dany wzorem (4.45) lub (4.46) dla wyznaczonych z oraz $m \in \{5, 10, 15\}$. Następnie obliczono obciążenie (bias) i błąd średniokwadratowy (MSE) zaproponowanego estymatora. W każdym przypadku obliczono również bias i MSE dla klasycznej empirycznej dystrybuanty (ECDF) dla $X_i - Y_i$, gdzie $i = 1, 2, ..., n$. Wyniki symulacji zostały zebrane w Tabelach 4.2 - 4.5. W Tabeli 4.2 przedstawiono wyniki dla różnicy dwóch zmiennych losowych o rozkładzie jednostajnym na przedziale $[0, 1]$. W Tabeli 4.3 zaprezentowano wyniki dla dwóch zmiennych losowych o rozkładzie wykładniczym z różnymi parametrami. W Tabeli 4.4 zebrano wyniki dla dwóch zmiennych losowych o rozkładzie Weibulla z tymi samymi parametrami. W Tabeli 4.5 przedstawiono wyniki dla próbek o różnych rozkładach – wykładniczym i Weibulla. Dodatkowo, część wyników przedstawiono na Rysunkach 4.1 - 4.16. Dla każdego z rozważanych rozkładów i wszystkich punktów kwantylowych przedstawiono wykresy obciążenia i MSE estymatora (4.45) lub (4.46) (dla $m = 15$) oraz ECDF. Ciągłe czarne odpowiadają obciążeniu dla estymatora (4.45) lub (4.46) natomiast czerwone odpowiadają obciążeniu ECDF. Linia przerywana czarna odpowiada MSE estymatora (4.45) lub (4.46), a czerwona reprezentuje MSE dla ECDF.

Tabela 4.2: Obciążenie i MSE estymatora dla $X, Y \sim \text{Unif}[0,1]$. **Źródło:** [14]

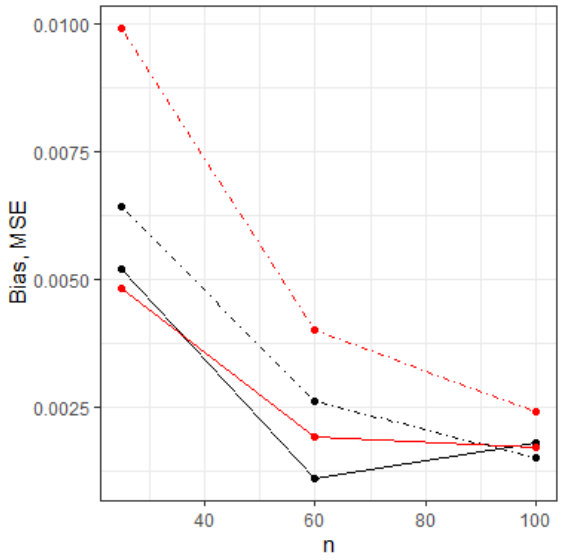
$n = 25$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0031	0.0036	0.0030	0.0037	0.0029	0.0037	0.0043	0.0063
0.4	0.0061	0.0064	0.0057	0.0066	0.0055	0.0066	0.0064	0.0100
0.6	0.0087	0.0063	0.0060	0.0064	0.0052	0.0064	0.0048	0.0099
0.8	0.0108	0.0040	0.0077	0.0041	0.0072	0.0040	0.0121	0.0069
$n = 60$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0012	0.0016	0.0010	0.0016	0.0009	0.0016	0.0014	0.0028
0.4	0.0008	0.0023	0.0005	0.0023	0.0002	0.0023	0.0012	0.0037
0.6	0.0053	0.0025	0.0012	0.0026	0.0011	0.0026	0.0019	0.0040
0.8	0.0062	0.0015	0.0034	0.0015	0.0026	0.0015	0.0029	0.0027
$n = 100$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0007	0.0009	0.0007	0.0009	0.0006	0.0009	0.0012	0.0018
0.4	0.0006	0.0015	0.0005	0.0015	0.0003	0.0016	0.0001	0.0023
0.6	0.0026	0.0015	0.0019	0.0015	0.0018	0.0015	0.0017	0.0024
0.8	0.0040	0.0010	0.0011	0.0010	0.0002	0.0010	0.0014	0.0018



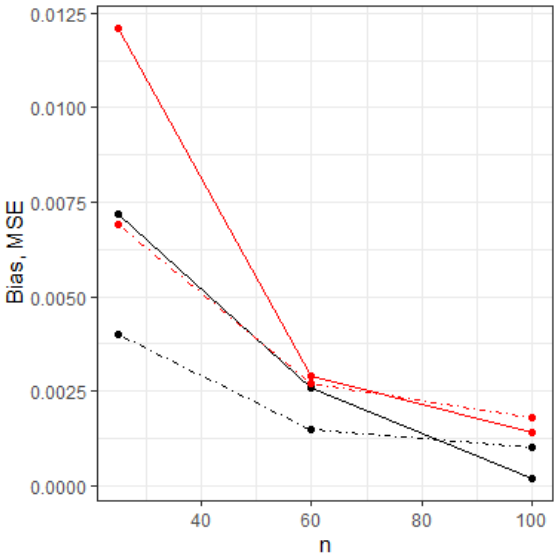
Rysunek 4.1: Obciążenie i MSE dla $F_{Z_1}(z) = 0.2$, gdzie $X, Y \sim$Unif$[0,1]$.
Źródło: [14]



Rysunek 4.2: Obciążenie i MSE dla $F_{Z_1}(z) = 0.4$, gdzie $X, Y \sim$Unif$[0,1]$.
Źródło: [14]



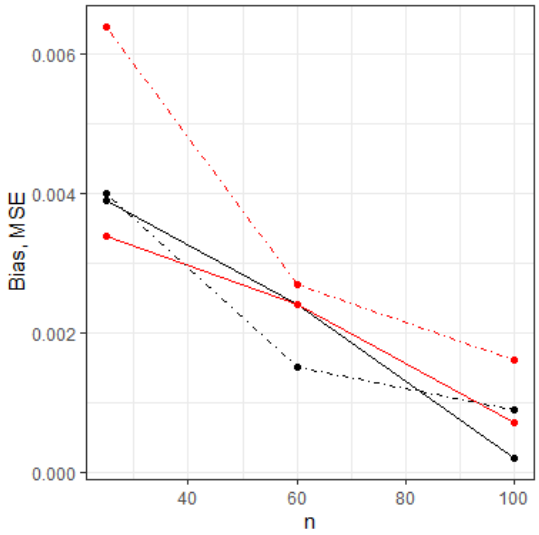
Rysunek 4.3: Obciążenie i MSE dla $F_{Z_1}(z) = 0.6$, gdzie $X, Y \sim$Unif$[0,1]$.
Źródło: [14]



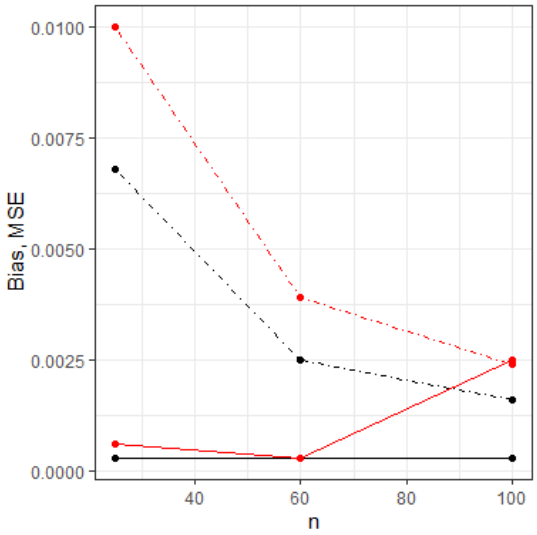
Rysunek 4.4: Obciążenie i MSE dla $F_{Z_1}(z) = 0.8$, gdzie $X, Y \sim$Unif$[0,1]$.
Źródło: [14]

Tabela 4.3: Obciążenie i MSE estymatora dla $X \sim \exp(1)$, $Y \sim \exp(2)$. **Źródło:** [14]

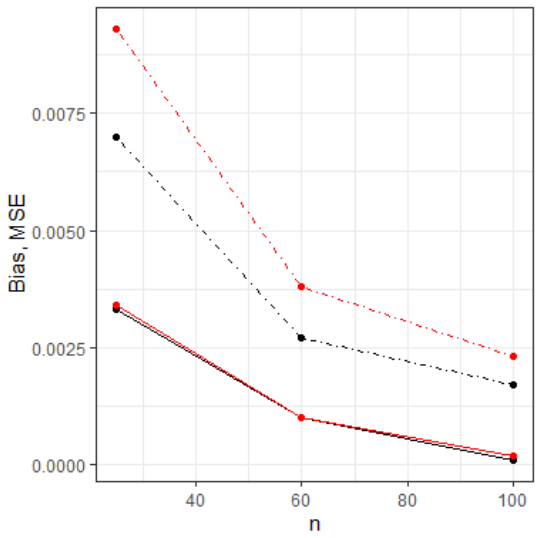
$n = 25$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0086	0.0040	0.0050	0.0040	0.0039	0.0040	0.0034	0.0064
0.4	0.0072	0.0066	0.0021	0.0068	0.0003	0.0068	0.0006	0.0100
0.6	0.0077	0.0069	0.0038	0.0070	0.0033	0.0070	0.0034	0.0093
0.8	0.0064	0.0049	0.0042	0.0050	0.0040	0.0050	0.0040	0.0068
$n = 60$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0075	0.0016	0.0028	0.0015	0.0024	0.0015	0.0024	0.0027
0.4	0.0108	0.0027	0.0021	0.0026	0.0003	0.0025	0.0003	0.0039
0.6	0.0071	0.0028	0.0018	0.0027	0.0010	0.0027	0.0010	0.0038
0.8	0.0049	0.0019	0.0024	0.0019	0.0019	0.0020	0.0026	0.0025
$n = 100$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0066	0.0010	0.0009	0.0009	0.0002	0.0009	0.0007	0.0016
0.4	0.0114	0.0016	0.0018	0.0015	0.0003	0.0016	0.0025	0.0024
0.6	0.0069	0.0017	0.0013	0.0017	0.0001	0.0017	0.0002	0.0023
0.8	0.0033	0.0012	0.0004	0.0013	0.0002	0.0013	0.0013	0.0017



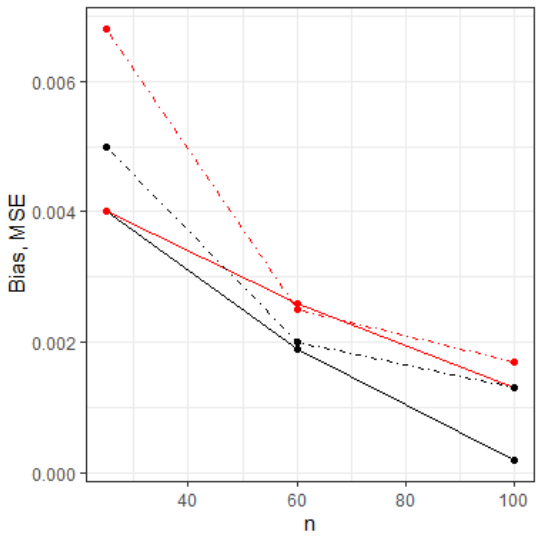
Rysunek 4.5: Bias i MSE dla $F_{Z_1}(z) = 0.2$, gdzie $X \sim$exp(1), $Y \sim$exp(2). **Źródło:** [14]



Rysunek 4.6: Bias i MSE dla $F_{Z_1}(z) = 0.4$, gdzie $X \sim$exp(1), $Y \sim$exp(2). **Źródło:** [14]



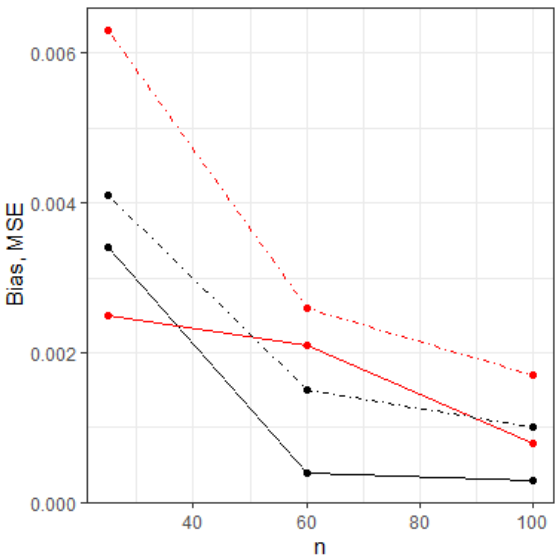
Rysunek 4.7: Bias i MSE dla $F_{Z_1}(z) = 0.6$, gdzie $X \sim$exp(1), $Y \sim$exp(2). **Źródło:** [14]



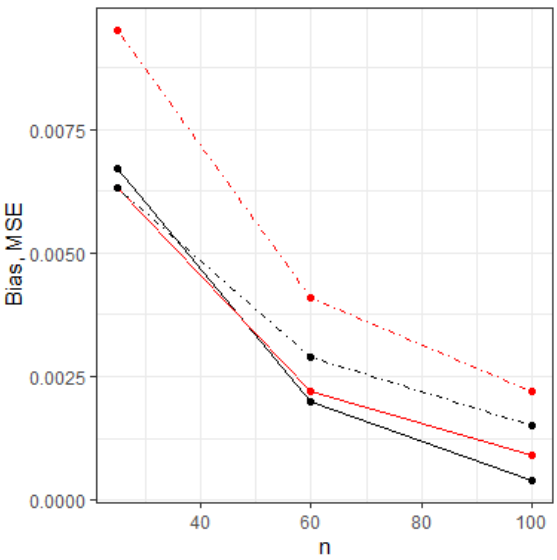
Rysunek 4.8: Bias i MSE dla $F_{Z_1}(z) = 0.8$, gdzie $X \sim$exp(1), $Y \sim$exp(2). **Źródło:** [14]

Tabela 4.4: Obciążenie i MSE estymatora dla $X, Y \sim$ Weibull $(2,1)$. **Źródło:** [14]

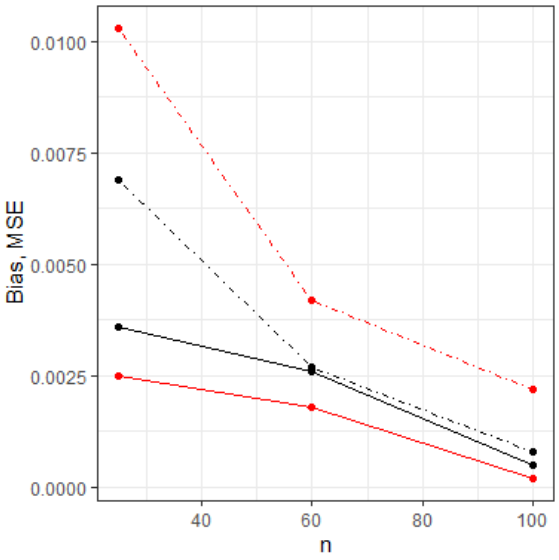
$n = 25$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0153	0.0043	0.0044	0.0042	0.0034	0.0041	0.0025	0.0063
0.4	0.0101	0.0062	0.0067	0.0064	0.0067	0.0063	0.0063	0.0095
0.6	0.0100	0.0068	0.0044	0.0069	0.0036	0.0069	0.0025	0.0103
0.8	0.0188	0.0037	0.0079	0.0036	0.0064	0.0036	0.0044	0.0058
$n = 60$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0150	0.0018	0.0028	0.0016	0.0004	0.0015	0.0021	0.0026
0.4	0.0087	0.0028	0.0034	0.0028	0.0020	0.0029	0.0022	0.0041
0.6	0.0098	0.0027	0.0038	0.0027	0.0026	0.0027	0.0018	0.0042
0.8	0.0178	0.0016	0.0042	0.0014	0.0015	0.0013	0.0013	0.0024
$n = 100$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0097	0.0015	0.0019	0.0010	0.0003	0.0010	0.0008	0.0017
0.4	0.0074	0.0015	0.0013	0.0015	0.0004	0.0015	0.0009	0.0022
0.6	0.0073	0.0015	0.0019	0.0015	0.0005	0.0015	0.0002	0.0022
0.8	0.0161	0.0011	0.0037	0.0008	0.0009	0.0008	0.0008	0.0015



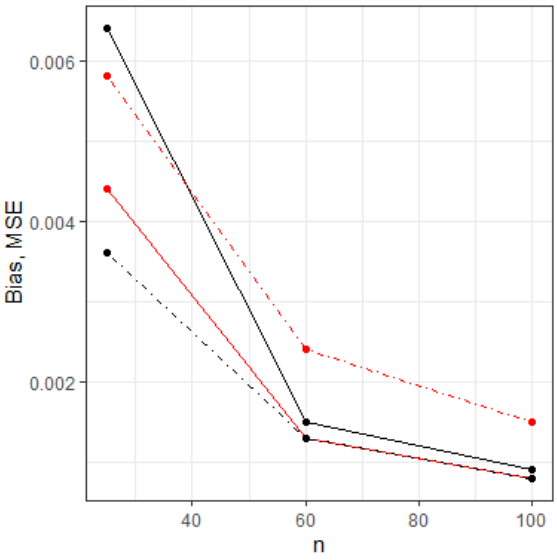
Rysunek 4.9: Obciążenie i MSE dla $F_{Z_1}(z) = 0.2$, gdzie $X, Y \sim$Weibull$(2,1)$.
Źródło: [14]



Rysunek 4.10: Obciążenie i MSE dla $F_{Z_1}(z) = 0.4$, gdzie $X, Y \sim$Weibull$(2,1)$.
Źródło: [14]



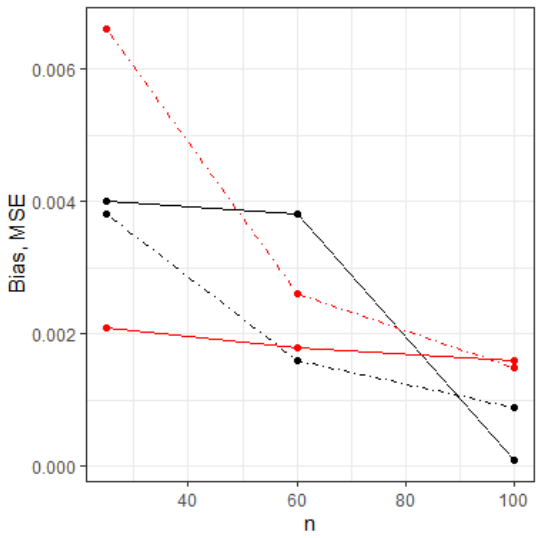
Rysunek 4.11: Obciążenie i MSE dla $F_{Z_1}(z) = 0.6$, gdzie $X, Y \sim$Weibull$(2,1)$.
Źródło: [14]



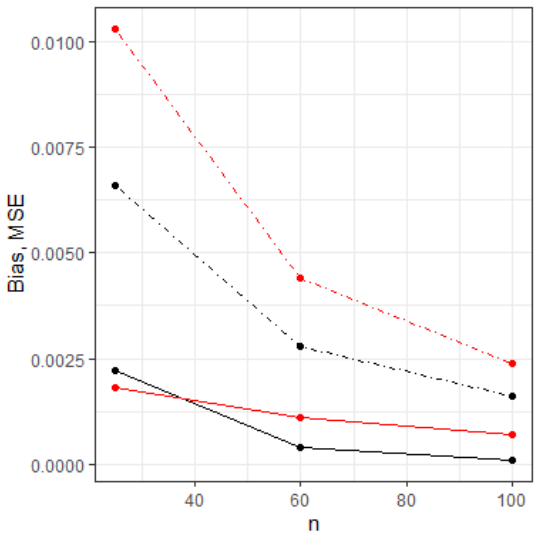
Rysunek 4.12: Obciążenie i MSE dla $F_{Z_1}(z) = 0.8$, gdzie $X, Y \sim$Weibull$(2,1)$.
Źródło: [14]

Tabela 4.5: Obciążenie i MSE estymatora dla $X \sim \exp(1), Y \sim$ Weibull $(2,1)$. **Źródło:** [14]

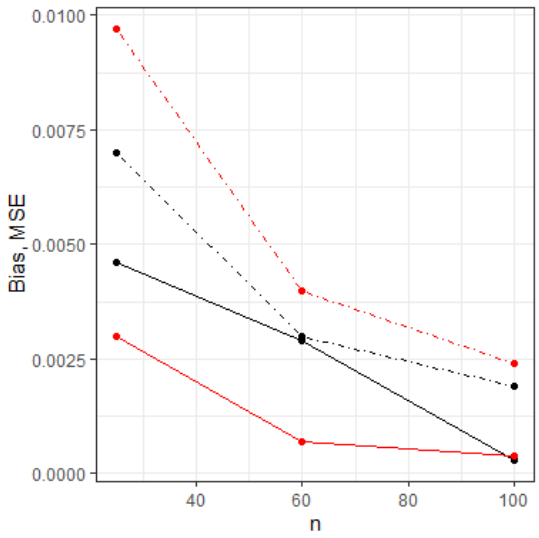
$n = 25$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0173	0.0040	0.0055	0.0038	0.0040	0.0038	0.0021	0.0066
0.4	0.0066	0.0063	0.0035	0.0066	0.0022	0.0066	0.0018	0.0103
0.6	0.0136	0.0069	0.0062	0.0070	0.0046	0.0070	0.0030	0.0097
0.8	0.0073	0.0051	0.0032	0.0051	0.0024	0.0051	0.0029	0.0064
$n = 60$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0163	0.0021	0.0054	0.0017	0.0038	0.0016	0.0018	0.0026
0.4	0.0057	0.0027	0.0014	0.0028	0.0004	0.0028	0.0011	0.0044
0.6	0.0133	0.0030	0.0044	0.0029	0.0029	0.0030	0.0007	0.0040
0.8	0.0065	0.0020	0.0018	0.0020	0.0009	0.0020	0.0024	0.0025
$n = 100$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0126	0.0012	0.0017	0.0009	0.0001	0.0009	0.0016	0.0015
0.4	0.0054	0.0015	0.0006	0.0016	0.0001	0.0016	0.0007	0.0024
0.6	0.0123	0.0019	0.0024	0.0019	0.0003	0.0019	0.0004	0.0024
0.8	0.0058	0.0012	0.0012	0.0012	0.0002	0.0012	0.0012	0.0015



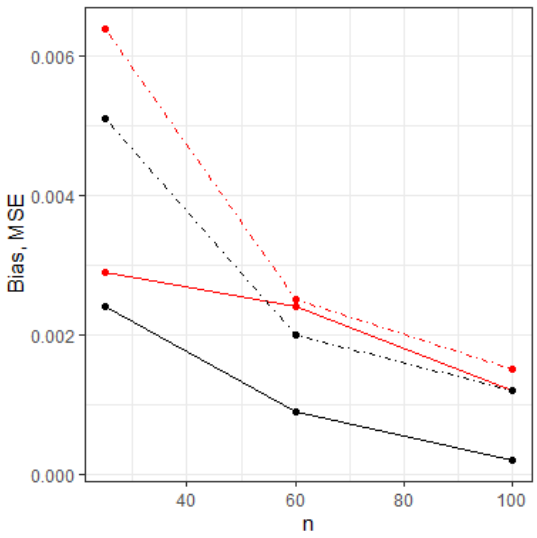
Rysunek 4.13: Obciążenie i MSE dla $F_{Z_1}(z) = 0.2$, gdzie $X \sim$exp(1) oraz $Y \sim$Weibull$(2,1)$. **Źródło:** [14]



Rysunek 4.14: Obciążenie i MSE dla $F_{Z_1}(z) = 0.4$, gdzie $X \sim$exp(1) oraz $Y \sim$Weibull$(2,1)$. **Źródło:** [14]



Rysunek 4.15: Obciążenie i MSE dla $F_{Z_1}(z) = 0.6$, gdzie $X \sim$exp(1) oraz $Y \sim$Weibull$(2,1)$. **Źródło:** [14]



Rysunek 4.16: Obciążenie i MSE dla $F_{Z_1}(z) = 0.8$, gdzie $X \sim$exp(1) oraz $Y \sim$Weibull$(2,1)$. **Źródło:** [14]

Na podstawie wyników przedstawionych w Tabelach 4.2 - 4.5 oraz na Rysunkach 4.1 - 4.16 można sformułować wspólne wnioski dla każdego badanego przypadku. Łatwo zauważyć, że wraz ze wzrostem liczności próby zarówno obciążenie, jak i MSE maleją. Wraz ze wzrostem liczby podziałów m wartość MSE w większości przypadków pozostaje na zbliżonym poziomie, natomiast obciążenie maleje. Należy podkreślić, że dla każdego m i każdego

rozmiaru próby estymator (4.45) lub (4.46) charakteryzuje się mniejszym MSE niż tradycyjna empiryczna dystrybuanta ECDF.

Ocena estymatora nieparametrycznego ($n_1 \neq n_2$)

W ostatnim kroku zbadano estymator dany wzorem (4.45) lub (4.46) dla przypadku różnych rozmiarów próby. Dla rozkładów wykładniczych i Weibulla przeprowadzono 1000 niezależnych powtórzeń symulacji. W każdym powtórzeniu generowano dwie próby losowe o liczebnościach (n_1, n_2) równych kolejno $(30, 25)$, $(60, 65)$ oraz $(105, 100)$. Następnie dla każdej pary prób wyznaczano wartości proponowanego estymatora dla $m \in \{5, 10, 15\}$, a także wartość estymatora porównawczego opartego na wszystkich różnicach par obserwacji (ozn. jako *wszystkie pary*). Na podstawie uzyskanych wartości każdego estymatora obliczono jego obciążenie (bias) oraz błąd średniokwadratowy (MSE) dla tych samych z, które były rozważane w przypadku $n_1 = n_2 = n$. Wyniki zebrano w Tabelach 4.6 oraz 4.7.

Tabela 4.6: Obciążenie i MSE dla $X \sim \exp(1)$, $Y \sim \exp(2)$.

$n_1 = 30$, $n_2 = 25$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0068	0.0034	0.0029	0.0034	0.0021	0.0033	0.0020	0.0034
0.4	0.0115	0.0053	0.0059	0.0055	0.0047	0.0055	0.0045	0.0056
0.6	0.0111	0.0060	0.0057	0.0060	0.0049	0.0060	0.0046	0.0060
0.8	0.0062	0.0038	0.0039	0.0040	0.0034	0.0039	0.0031	0.0040
$n_1 = 60$, $n_2 = 65$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0062	0.0015	0.0026	0.0015	0.0015	0.0014	0.0009	0.0015
0.4	0.0113	0.0016	0.0039	0.0014	0.0011	0.0014	0.0008	0.0014
0.6	0.0099	0.0030	0.0052	0.0030	0.0037	0.0030	0.0029	0.0031
0.8	0.0048	0.0020	0.0021	0.0020	0.0018	0.0020	0.0014	0.0021
$n_1 = 105$, $n_2 = 100$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0060	0.0009	0.0014	0.0009	0.0002	0.0009	0.0006	0.0009
0.4	0.0072	0.0015	0.0013	0.0016	0.0001	0.0016	0.0007	0.0016
0.6	0.0068	0.0017	0.0008	0.0017	0.0005	0.0017	0.0019	0.0017
0.8	0.0039	0.0011	0.0005	0.0011	0.0001	0.0011	0.0006	0.0011

Tabela 4.7: Obciążenie i MSE dla $X, Y \sim \text{Weibull}(2;1)$.

$n_1 = 30$, $n_2 = 25$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0120	0.0038	0.0046	0.0037	0.0035	0.0037	0.0024	0.0037
0.4	0.0083	0.0062	0.0048	0.0064	0.0034	0.0064	0.0033	0.0065
0.6	0.0097	0.0054	0.0055	0.0056	0.0041	0.0056	0.0037	0.0057
0.8	0.0175	0.0033	0.0052	0.0031	0.0032	0.0032	0.0016	0.0032
$n_1 = 60$, $n_2 = 65$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0099	0.0017	0.0022	0.0016	0.0010	0.0016	0.0022	0.0016
0.4	0.0080	0.0025	0.0037	0.0025	0.0026	0.0025	0.0019	0.0026
0.6	0.0092	0.0024	0.0030	0.0024	0.0021	0.0024	0.0013	0.0025
0.8	0.0167	0.0016	0.0039	0.0014	0.0012	0.0014	0.0008	0.0014
$n_1 = 105$, $n_2 = 100$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0082	0.0011	0.0014	0.0010	0.0001	0.0010	0.0012	0.0010
0.4	0.0066	0.0016	0.0008	0.0016	0.0003	0.0016	0.0012	0.0016
0.6	0.0072	0.0015	0.0012	0.0016	0.0001	0.0016	0.0010	0.0016
0.8	0.0155	0.0010	0.0035	0.0007	0.0007	0.0007	0.0006	0.0007

Tabela 4.8: Obciążenie i MSE estymatora dla $X \sim \exp(1),\ Y \sim \text{Weibull}\ (2,1)$.

$n_1 = 30, n_2 = 25$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0021	0.0035	0.0020	0.0035	0.0021	0.0035	0.0022	0.0036
0.4	0.0079	0.0054	0.0041	0.0057	0.0032	0.0057	0.0027	0.0057
0.6	0.0141	0.0058	0.0072	0.0057	0.0055	0.0058	0.0045	0.0058
0.8	0.0074	0.0044	0.0038	0.0044	0.0031	0.0044	0.0023	0.0044
$n_1 = 60, n_2 = 65$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0017	0.0015	0.0012	0.0015	0.0009	0.0016	0.0010	0.0016
0.4	0.0070	0.0022	0.0026	0.0023	0.0019	0.0023	0.0011	0.0024
0.6	0.0145	0.0029	0.0054	0.0028	0.0040	0.0029	0.0026	0.0029
0.8	0.0062	0.0022	0.0014	0.0022	0.0007	0.0014	0.0014	0.0014
$n_1 = 105, n_2 = 100$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_1}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0005	0.0009	0.0001	0.0009	0.0000	0.0009	0.0001	0.0009
0.4	0.0061	0.0015	0.0008	0.0015	0.0001	0.0016	0.0008	0.0016
0.6	0.0123	0.0016	0.0022	0.0015	0.0008	0.0015	0.0008	0.0015
0.8	0.0058	0.0011	0.0005	0.0011	0.0004	0.0011	0.0012	0.0011

Przeprowadzone badania symulacyjne pokazują, że zaproponowany, nowy estymator dobrze przybliża dystrybuantę zmiennej Z_1 dla rozważanych przypadków. Jego dokładność wyraźnie zależy od liczby podprzedziałów m oraz rozmiarów próbek. Dla $m = 5$ obciążenie jest na ogół większe, natomiast zwiększenie wartości m prowadzi do jego istotnego zmniejszenia. Dla największych rozważanych prób oraz $m = 15$ obciążenie proponowanego estymatora jest w większości przypadków bardzo bliskie zeru, a niekiedy mniejsze niż dla estymatora porównawczego.

Wartości błędu średniokwadratowego wszystkich analizowanych estymatorów są bardzo zbliżone. Dla $m = 10$ i $m = 15$ różnice pomiędzy proponowanym estymatorem a estymatorem opartym na wszystkich różnicach par obserwacji są niewielkie i występują zazwyczaj dopiero na dalszych miejscach po przecinku. W niektórych przypadkach proponowany estymator osiąga nawet nieznacznie mniejszy MSE niż estymator oparty na wszystkich różnicach par. Ponadto wraz ze wzrostem liczebności prób wartości MSE maleją, co potwierdza poprawę jakości oszacowania przy większej liczbie obserwacji.

Uzyskane wyniki nie wskazują na jednoznaczną przewagę jednej metody w każdym rozważanym przypadku. Pokazują jednak, że przy odpowiednim wyborze parametru m proponowany estymator osiąga dokładność porównywalną z estymatorem wykorzystującym wszystkie różnice $X_i - Y_j$, a jednocześnie może charakteryzować się mniejszym obciążeniem. Stanowi zatem wartościową alternatywną metodę estymacji dystrybuanty różnicy dwóch zmiennych losowych.

4.4. Aproksymacja i estymacja rozkładu iloczynu zmiennych losowych

W tym podrozdziale reprezentacja kwantylowa dystrybuanty iloczynu niezależnych, dodatnich zmiennych losowych dana wzorem (4.29) i metoda trapezów zostaną wykorzystane do wyznaczenia wzoru aproksymującego dystrybuantę iloczynu dodatnich zmiennych losowych X i Y. Podobnie jak w przypadku różnicy zmiennych losowych na podstawie wzoru aproksymującego zostanie wyznaczony nieparametryczny estymator dystrybuanty iloczynu niezależnych, nieujemnych zmiennych losowych.

4.4.1. Aproksymacja rozkładu

Rozważmy następujący przedział $\left[F_X(Q_X(0)), F_X\left(\frac{z}{Q_Y(0)}\right)\right]$ jako dziedzinę całkowania. Dzielimy ten przedział na m podprzedziałów za pomocą punktów

$$\left\{F_X(Q_X(0)), F_X\left(\frac{\frac{z}{Q_Y(0)}+(m-1)Q_X(0)}{m}\right), F_X\left(\frac{\frac{2z}{Q_Y(0)}+(m-2)Q_X(0)}{m}\right), ..., F_X\left(\frac{z}{Q_Y(0)}\right)\right\}.$$

Następnie obliczamy wartości $u_2 = F_Y\left(\frac{z}{Q_X(u_1)}\right)$ w wyżej wymienionych punktach:

$$\left\{F_Y\left(\frac{z}{Q_X(0)}\right), F_Y\left(\frac{zm}{\frac{z}{Q_Y(0)}+(m-1)Q_X(0)}\right), F_Y\left(\frac{zm}{\frac{2z}{Q_Y(0)}+(m-2)Q_X(0)}\right), ..., F_Y(Q_Y(0))\right\}.$$

Korzystając z metody trapezów oraz faktu, że X i Y są niezależnymi zmiennymi losowymi otrzymujemy wzór aproksymujący dystrybuantę iloczynu tych zmiennych losowych w następującej postaci

$$F_{Z_{2m}(z)} = \frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)}+(m-i-1)Q_X(0)}{m}\right) - F_X\left(\frac{\frac{iz}{Q_Y(0)}+(m-i)Q_X(0)}{m}\right)\right)$$
$$\cdot\left(F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)}+(m-i)Q_X(0)}\right) + F_Y\left(\frac{zm}{\frac{(i+1)z}{Q_Y(0)}+(m-i-1)Q_X(0)}\right)\right). \tag{4.53}$$

W następnym paragrafie zajmiemy się wyznaczeniem błędu tej aproksymacji.

4.4.2. Błąd aproksymacji

W celu oszacowania błędu aproksymacji dystrybuanty $F_{Z_2}(z)$, gdzie $Z_2 = X \cdot Y$, za pomocą wzoru (4.53) dowiedziemy następującego twierdzenia.

Twierdzenie 4.6. *Niech $X > 0$ będzie zmienną losową o co najmniej raz różniczkowalnej dystrybuancie F_X na przedziale $(Q_X(0), Q_X(1))$, przy czym $Q_X(0) > 0$. Niech Y będzie zmienną losową o co najmniej dwukrotnie różniczkowalnej dystrybuancie F_Y na przedziale $(Q_Y(0), Q_Y(1))$, przy czym $Q_Y(0) > 0$. Dodatkowo załóżmy, że F'_X i F''_Y są ciągłe. Wówczas oszacowanie błędu aproksymacji $F_{Z_2}(z)$ za pomocą $F_{Z_{2m}}(z)$ dane jest wzorem*

$$|F_{Z_2}(z) - F_{Z_{2m}}(z)| \leqslant \frac{M_1(z)[M_2(z)]^3\left(\frac{z}{Q_Y(0)} - Q_X(0)\right)^3}{12m^2}, \tag{4.54}$$

gdzie

$$M_1(z) = \max_{F_X\left(Q_X(0)\right) \leqslant \eta \leqslant F_X\left(\frac{z}{Q_Y(0)}\right)} \left|F''_Y\left(\frac{z}{Q_X(\eta)}\right)\right|, \quad M_2(z) = \max_{Q_X(0) \leqslant \tau \leqslant \frac{z}{Q_Y(0)}} |F'_X(\tau)|$$

oraz m jest liczbą podprzedziałów.

Dowód. Na podstawie (4.29) i (4.53) mamy

$$|F_{Z_2}(z) - F_{Z_{2m}}(z)| = \left| \int_{F_X(Q_X(0))}^{F_X\left(\frac{z}{Q_Y(0)}\right)} F_Y\left(\frac{z}{Q_X(u_1)}\right) du_1 \right.$$

$$-\frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}{m}\right) - F_X\left(\frac{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}{m}\right)\right)$$

$$\left. \cdot \left(F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}\right) + F_Y\left(\frac{zm}{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}\right)\right)\right|$$

$$= \left| \sum_{i=0}^{m-1} \int_{F_X\left(\frac{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}{m}\right)}^{F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}{m}\right)} F_Y\left(\frac{z}{Q_X(u_1)}\right) du_1 \right.$$

$$-\frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}{m}\right) - F_X\left(\frac{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}{m}\right)\right)$$

$$\left. \left(F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}\right) + F_Y\left(\frac{zm}{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}\right)\right)\right|.$$

Z równania (5.1.4) w pracy Atkinsona [1] otrzymujemy, że

$$|F_{Z_2}(z) - F_{Z_{2m}}(z)| \leqslant \sum_{i=0}^{m-1} \frac{\left(F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}{m}\right) - F_X\left(\frac{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}{m}\right)\right)^3}{12} \left|F''_Y\left(\frac{z}{Q_X(\eta_i)}\right)\right|,$$

gdzie $\eta_i \in \left[F_X\left(\frac{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}{m} \right), F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}{m} \right) \right]$ i na podstawie twierdzenia o wartości średniej mamy

$$|F_{Z_2}(z) - F_{Z_{2m}}(z)| \leqslant \sum_{i=0}^{m-1} \frac{\left(F_X'(\tau_i)\left(\frac{z}{Q_Y(0)} - Q_X(0)\right)\right)^3}{12m^3} \left| F_Y''\left(\frac{z}{Q_X(\eta_i)} \right) \right|,$$

gdzie $\tau_i \in \left[\frac{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}{m}, \frac{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}{m} \right]$.

Z faktu, że F_Y'' oraz F_X' są ciągłe dostajemy, że

$$\left| F_Y''\left(\frac{z}{Q_X(\eta_i)} \right) \right| \leqslant \max_{F_X\left(\frac{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}{m} \right) \leqslant \eta_i \leqslant F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}{m} \right)} \left| F_Y''\left(\frac{z}{Q_X(\eta_i)} \right) \right|$$

$$\leqslant M_1(z) = \max_{F_X\left(Q_X(0)\right) \leqslant \eta \leqslant F_X\left(\frac{z}{Q_Y(0)}\right)} \left| F_Y''\left(\frac{z}{Q_X(\eta)} \right) \right|$$

oraz

$$|F_X'(\tau_i)| \leqslant \max_{\frac{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}{m} \leqslant \tau_i \leqslant \frac{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}{m}} |F_X'(\tau_i)| \leqslant M_2(z) = \max_{Q_X(0) \leqslant \tau \leqslant \frac{z}{Q_Y(0)}} |F_X'(\tau)|,$$

co dowodzi tezy. □

Uwaga 4.3. Łatwo zauważyć, że dla ustalonego z wyrażenie $|F_{Z_2}(z) - F_{Z_{2m}}(z)| \to 0$, gdy $m \to \infty$.

4.4.3. Estymator nieparametryczny

Niech $\{X_1, \ldots, X_{n_1}\}$ oraz $\{Y_1, \ldots, Y_{n_2}\}$ będą niezależnymi próbami losowymi o rozmiarach, odpowiednio, n_1 i n_2 pobranymi z populacji o ciągłych dystrybuantach F_X i F_Y. Niech $\hat{F}_X$ i $\hat{F}_Y$ będą empirycznymi dystrybuantami, odpowiednio, F_X oraz F_Y. Wprowadźmy oznaczenia

$$Q_Y^*(0) := \min\{Y_i,\ 1 \leqslant i \leqslant n_2\} \quad \text{oraz} \quad Q_X^*(0) := \min\{X_i,\ 1 \leqslant i \leqslant n_2\}.$$

Dla $z > 0$ nieparametryczny estymator $F_{Z_2}(z)$ jest wyrażony następującym wzorem

$$\hat{F}_{Z_2}(z) = \frac{1}{2} \sum_{i=0}^{m-1} \left(\hat{F}_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X^*(0)}{m} \right) - \hat{F}_X\left(\frac{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)}{m} \right) \right)$$

$$\cdot \left(\hat{F}_Y\left(\frac{zm}{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)} \right) + \hat{F}_Y\left(\frac{zm}{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X^*(0)} \right) \right). \tag{4.55}$$

Poniższe twierdzenie dowodzi silnej zgodności estymatora $\hat{F}_{Z_2}(z)$.

Twierdzenie 4.7. *Niech $Z_2 = X \cdot Y$ będzie zmienną losową, gdzie $X > 0$ oraz $Y > 0$ są zmiennymi losowymi o ciągłych dystrybuantach, odpowiednio, F_X oraz F_Y spełniającymi założenia Twierdzenia 4.6. Niech $\hat{F}_{Z_2}$ będzie zdefiniowany wzorem (4.55). Wówczas*

$$|\hat{F}_{Z_2}(z) - F_{Z_2}(z)| = O\left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}}\right) + O(m^{-2}).$$

Dowód. Zauważmy, że

$$\begin{aligned}
\hat{F}_{Z_2}(z) - F_{Z_{2m}}(z) = \frac{1}{2}\sum_{i=0}^{m-1} &\left(\hat{F}_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X^*(0)}{m}\right) - \hat{F}_X\left(\frac{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)}{m}\right)\right) \\
&\left(\hat{F}_Y\left(\frac{zm}{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)}\right) + \hat{F}_Y\left(\frac{zm}{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X^*(0)}\right)\right) \\
-\frac{1}{2}\sum_{i=0}^{m-1} &\left(F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}{m}\right) - F_X\left(\frac{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}{m}\right)\right) \\
&\left(F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}\right) + F_Y\left(\frac{zm}{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}\right)\right) \\
= \frac{1}{2}\sum_{i=0}^{m-1} &\left\{\left[\hat{F}_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X^*(0)}{m}\right)\hat{F}_Y\left(\frac{zm}{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)}\right)\right.\right. \\
&\left.\left. - F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}{m}\right)F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}\right)\right] \right. \\
&+ \left[\hat{F}_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X^*(0)}{m}\right)\hat{F}_Y\left(\frac{zm}{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X^*(0)}\right)\right. \\
&\left. - F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}{m}\right)F_Y\left(\frac{zm}{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}\right)\right] \\
&+ \left[F_X\left(\frac{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}{m}\right)F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}\right)\right. \\
&\left. - \hat{F}_X\left(\frac{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)}{m}\right)\hat{F}_Y\left(\frac{zm}{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)}\right)\right] \\
&+ \left[F_X\left(\frac{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}{m}\right)F_Y\left(\frac{zm}{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}\right)\right. \\
&\left.\left. - \hat{F}_X\left(\frac{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)}{m}\right)\hat{F}_Y\left(\frac{zm}{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X^*(0)}\right)\right]\right\}
\end{aligned}$$

$$= \frac{1}{2}\sum_{i=0}^{m-1}\{A_i(z)+B_i(z)+C_i(z)+D_i(z)\}.$$

Rozważając składnik $A_i(z)$ dostajemy następujące oszacowanie

$$|A_i(z)| \leqslant \left|\hat{F}_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)}+(m-i-1)Q_X^*(0)}{m}\right)-F_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)}+(m-i-1)Q_X^*(0)}{m}\right)\right|$$

$$\cdot\hat{F}_Y\left(\frac{zm}{\frac{iz}{Q_Y^*(0)}+(m-i)Q_X^*(0)}\right)+\left|F_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)}+(m-i-1)Q_X^*(0)}{m}\right)\right.$$

$$\left.-F_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)}+(m-i-1)Q_X(0)}{m}\right)\right|\cdot\hat{F}_Y\left(\frac{zm}{\frac{iz}{Q_Y^*(0)}+(m-i)Q_X^*(0)}\right)$$

$$+F_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)}+(m-i-1)Q_X(0)}{m}\right)\cdot\left|\hat{F}_Y\left(\frac{zm}{\frac{iz}{Q_Y^*(0)}+(m-i)Q_X^*(0)}\right)-F_Y\left(\frac{zm}{\frac{iz}{Q_Y^*(0)}+(m-i)Q_X^*(0)}\right)\right|$$

$$+F_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)}+(m-i-1)Q_X(0)}{m}\right)\cdot\left|F_Y\left(\frac{zm}{\frac{iz}{Q_Y^*(0)}+(m-i)Q_X^*(0)}\right)-F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)}+(m-i)Q_X^*(0)}\right)\right|$$

$$+\left|F_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)}+(m-i-1)Q_X(0)}{m}\right)\cdot F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)}+(m-i)Q_X^*(0)}\right)\right.$$

$$\left.-F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)}+(m-i-1)Q_X(0)}{m}\right)\cdot F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)}+(m-i)Q_X(0)}\right)\right|$$

$$\leqslant J_1+J_2+J_3+J_4+\left|F_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)}+(m-i-1)Q_X(0)}{m}\right)-F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)}+(m-i-1)Q_X(0)}{m}\right)\right|$$

$$\cdot F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)}+(m-i)Q_X^*(0)}\right)+F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)}+(m-i-1)Q_X(0)}{m}\right)$$

$$\cdot\left|F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)}+(m-i)Q_X^*(0)}\right)-F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)}+(m-i)Q_X(0)}\right)\right|$$

$$= J_1+J_2+J_3+J_4+J_5+J_6.$$

Wykorzystując fakt, że $|\hat{F}_X(x)-F_X(x)| = O\left(n_1^{-\frac{1}{2}}\right)$ uzyskujemy

$$J_1 \leqslant \left|\hat{F}_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)}+(m-i-1)Q_X^*(0)}{m}\right)-F_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)}+(m-i-1)Q_X^*(0)}{m}\right)\right| = O\left(n_1^{-\frac{1}{2}}\right).$$

Na podstawie pracy Bahadura [2] mamy

$$Q^*(p) = Q(p) - \frac{F_n(p) - F(p)}{f(Q(p))} + O\Big(n^{-\frac{3}{4}} \log n\Big) \leqslant Q(p) + O\Big(n_1^{-\frac{1}{2}}\Big) + O\Big(n^{-\frac{3}{4}} \log n\Big).$$

Stąd

$$Q^*(p) - Q(p) = O\Big(n_1^{-\frac{1}{2}}\Big). \tag{4.56}$$

Z twierdzenia o wartości średniej oraz (4.56) otrzymujemy

$$\begin{aligned} J_2 &\leqslant \Big| F_X\Big(\frac{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X^*(0)}{m}\Big) - F_X\Big(\frac{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X(0)}{m}\Big)\Big| \\ &\leqslant C\left[\frac{(m-i-1)|Q_X^*(0) - Q_X(0)|}{m}\right] = O\Big(n_1^{-\frac{1}{2}}\Big). \end{aligned}$$

Analogicznie $|\hat{F}_Y(y) - F_Y(y)| = O\Big(n_1^{-\frac{1}{2}}\Big)$. Zatem

$$J_3 \leqslant \left| \hat{F}_Y\left(\frac{zm}{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)}\right) - F_Y\left(\frac{zm}{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)}\right)\right| = O\Big(n_1^{-\frac{1}{2}}\Big).$$

Korzystając z twierdzenia o wartości średniej oraz (4.56) i z faktu, że $Q_Y(0) \leqslant Q_Y^*(0)$, otrzymujemy

$$\begin{aligned} J_4 &\leqslant \left| F_Y\left(\frac{zm}{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)}\right) - F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)} + (m-i)Q_X^*(0)}\right)\right| \\ &\leqslant C\left[\frac{\left|\frac{iz}{Q_Y^*(0)} - \frac{iz}{Q_Y(0)}\right|}{[\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)][\frac{iz}{Q_Y(0)} + (m-i)Q_X^*(0)]}\right] \\ &\leqslant C\left[\frac{\frac{|Q_Y(0) - Q_Y^*(0)|}{Q_Y(0)\cdot Q_Y^*(0)}}{\frac{iz}{Q_Y^*(0)}\frac{iz}{Q_Y(0)}}\right] \leqslant C|Q_Y(0) - Q_Y^*(0)| = O\Big(n_1^{-\frac{1}{2}}\Big), \end{aligned}$$

$$J_5 \leqslant \left| F_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X(0)}{m}\right) - F_X\left(\frac{\frac{(i+1)z}{Q_Y(0)} + (m-i-1)Q_X(0)}{m}\right)\right|$$

$$\leqslant C\left[\frac{\frac{(i+1)z}{Q_Y^*(0)} - \frac{(i+1)z}{Q_Y(0)}}{m}\right] \leqslant C\left[\frac{|Q_Y(0) - Q_Y^*(0)|}{Q_Y(0)\cdot Q_Y^*(0)}\right] \leqslant C\left[\frac{|Q_Y(0) - Q_Y^*(0)|}{Q_Y(0)\cdot Q_Y(0)}\right] = O\Big(n_1^{-\frac{1}{2}}\Big)$$

oraz

$$J_6 \leqslant \left| F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)} + (m-i)Q_X^*(0)}\right) - F_Y\left(\frac{zm}{\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)}\right)\right|$$

$$\leqslant C\left[\frac{|Q_X(0) - Q_X^*(0)|}{[\frac{iz}{Q_Y(0)} + (m-i)Q_X^*(0)][\frac{iz}{Q_Y(0)} + (m-i)Q_X(0)]}\right] \leqslant C\left[\frac{|Q_X(0) - Q_X^*(0)|}{[\frac{iz}{Q_Y(0)}][\frac{iz}{Q_Y(0)}]}\right] = O\Big(n_1^{-\frac{1}{2}}\Big).$$

Stąd

$$|A_i(z)| = O\left(n_1^{-\frac{1}{2}} + n_2^{-\frac{1}{2}}\right).$$

Podobnie można pokazać, że $|B_i(z)| = O\left(n_1^{-\frac{1}{2}} + n_2^{-\frac{1}{2}}\right)$, $|C_i(z)| = O\left(n_1^{-\frac{1}{2}} + n_2^{-\frac{1}{2}}\right)$ oraz $|D_i(z)| = O\left(n_1^{-\frac{1}{2}} + n_2^{-\frac{1}{2}}\right)$. Zatem

$$|\hat{F}_{Z_2}(z) - F_{Z_{2m}}(z)| \leqslant |A_i(z)| + |B_i(z)| + |C_i(z)| + |D_i(z)| = O\left(n_1^{-\frac{1}{2}} + n_2^{-\frac{1}{2}}\right). \tag{4.57}$$

Ostatecznie na podstawie (4.57) oraz Twierdzenia 4.6 dostajemy

$$|\hat{F}_{Z_2}(z) - F_{Z_2}(z)| \leqslant |\hat{F}_{Z_2}(z) - F_{Z_{2m}}(z)| + |F_{Z_{2m}}(z) - F_{Z_2}(z)| = O\left(n_1^{-\frac{1}{2}} + n_2^{-\frac{1}{2}}\right) + O\left(m^{-2}\right).$$

□

Uwaga 4.4. Deterministyczny błąd aproksymacji wynikający z zastosowania reguły trapezów maleje w tempie $O\left(m^{-2}\right)$. Składnik stochastyczny estymatora jest rzędu $O\left(n_1^{-\frac{1}{2}} + n_2^{-\frac{1}{2}}\right)$. Wybierając m tak, aby $m^{-2} \leqslant n_1^{-\frac{1}{2}} + n_2^{-\frac{1}{2}}$, na przykład $m \geqslant n^{\frac{1}{4}}$ przy $n_1 = n_2$, błąd kwadratury można uczynić znikomym w porównaniu do zmienności próbkowania, przy zachowaniu umiarkowanych kosztów obliczeniowych.

Uwaga 4.5. Korzystając z Lematu 1 z pracy Midhu i in. [32] łatwo pokazać, że zmienna losowa $2n_1n_2\widehat{F}_{Z_2}(z)$ ma następującą funkcję masy prawdopodobieństwa:

$$P(2n_1n_2\widehat{F}_{Z_2}(z) = k) = \begin{cases} \sum\limits_{j=0}^{\frac{1}{4}(2k+(-1)^k-1)} \dfrac{(n_1n_2)!}{j!(k-2j)!(n_1n_2-k+j)!} q_1^j q_2^{k-2j} q_3^{n_1n_2-k+j}, & \text{gdy } k \leqslant n_1n_2, \\ \sum\limits_{j=k-n_1n_2}^{\frac{1}{4}(2k+(-1)^k-1)} \dfrac{(n_1n_2)!}{j!(k-2j)!(n_1n_2-k+j)!} q_1^j q_2^{k-2j} q_3^{n_1n_2-k+j}, & \text{gdy } k > n_1n_2, \end{cases}$$

gdzie $k = 0, 1, ..., 2n_1n_2$,

$$q_1 = \sum_{i=0}^{m-1} \left[\left(F_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X^*(0)}{m} \right) - F_X\left(\frac{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)}{m} \right) \right) \cdot F_Y\left(\frac{zm}{\frac{(i+1)Q_Y^*(0)}{m} + (m-i-1)Q_X^*(0)} \right) \right],$$

$$q_2 = \sum_{i=0}^{m-1} \left[\left(F_X\left(\frac{\frac{(i+1)z}{Q_Y^*(0)} + (m-i-1)Q_X^*(0)}{m} \right) - F_X\left(\frac{\frac{iz}{Q_Y^*(0)} + (m-i)Q_X^*(0)}{m} \right) \right) \cdot \left(F_Y\left(\frac{zm}{\frac{iQ_Y^*(0)}{m} + (m-i)Q_X^*(0)} \right) - F_Y\left(\frac{zm}{\frac{(i+1)Q_Y^*(0)}{m} + (m-i-1)Q_X^*(0)} \right) \right) \right]$$

oraz

$$q_3 = 1 - q_1 - q_2.$$

4.4.4. Symulacje

Analogicznie jak w przypadku różnicy zmiennych losowych, w tej części zostaną przedstawione symulacje mające na celu analizę błędu aproksymacji za pomocą $F_{Z_{2m}}$ oraz weryfikację jakości estymatora $\hat{F}_{Z_2}(z)$. Symulacje ponownie przeprowadzono w programie *R* studio.

Błąd aproksymacji

W pierwszym kroku zbadano błąd aproksymacji $F_{Z_{2m}}$ dla trzech różnych przypadków:

1. X, Y – przesunięty rozkład wykładniczy z parametrami (1;0,5),
2. X, Y – przesunięty rozkład Weibulla z parametrami (2;1;0,5),
3. X, Y – rozkład Pareto z parametrami (1,3).

Wyniki symulacji zostały przedstawione w Tabeli 4.9.

Tabela 4.9: Błąd aproksymacji $F_{Z_{2m}}$.

$X,Y \sim \exp(1;0,5)$				$X,Y \sim \text{Weibull}(2;1;0,5)$				$X,Y \sim \text{Pareto}(1,3)$			
F_{Z_2}	$m=5$	$m=10$	$m=15$	F_{Z_2}	$m=5$	$m=10$	$m=15$	F_{Z_2}	$m=5$	$m=10$	$m=15$
0.2	0.0003	0.0001	0.0000	0.2	0.0185	0.0050	0.0022	0.2	0.0017	0.0004	0.0002
0.4	0.0024	0.0006	0.0002	0.4	0.0073	0.0024	0.0010	0.4	0.0047	0.0012	0.0005
0.6	0.0105	0.0029	0.0013	0.6	0.0154	0.0041	0.0019	0.6	0.0082	0.0020	0.0009
0.8	0.0323	0.0086	0.0038	0.8	0.0501	0.0162	0.0074	0.8	0.0116	0.0029	0.0012

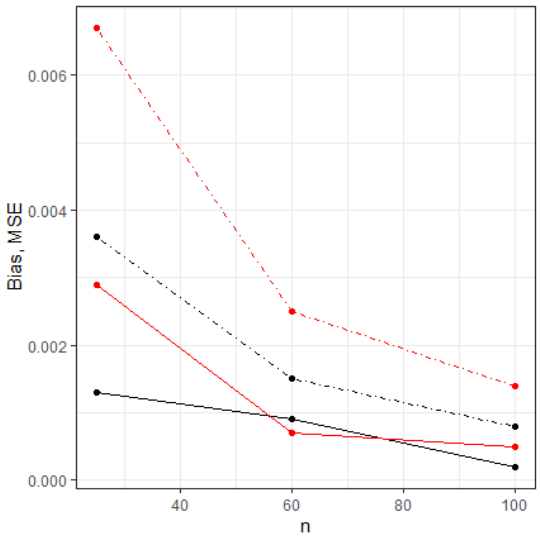
Wyniki symulacji wyraźnie pokazują, że wraz ze wzrostem liczby podziałów m, błąd przybliżenia maleje. Jest to wspólny wniosek dla każdego z badanych rozkładów.

Ocena estymatora nieparametrycznego ($n_1 = n_2 = n$)

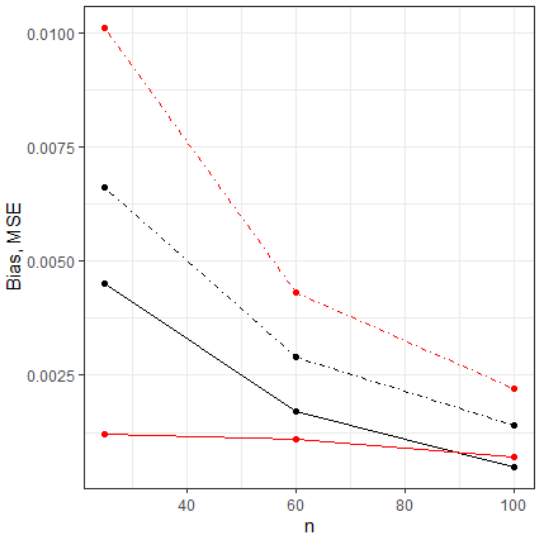
W celu oceny estymatora danego wzorem (4.55) wygenerowano 1000 razy próbki z rozkładów: wykładniczego, Weibulla oraz Pareto. Rozmiary prób przyjęto $n \in \{25, 60, 100\}$. W każdym przypadku wyznaczono wartości z z zależności $F_{Z_2}(z) = u$, gdzie $u \in \{0.2, 0.4, 0.6, 0.8\}$, a $Z_2 = X \cdot Y$. Następnie, dla każdej próby, obliczono estymator podany wzorem (4.55) dla wyznaczonych z oraz $m \in \{5, 10, 15\}$. W kolejnym kroku wyznaczono obciążenie (bias) oraz błąd średniokwadratowy (MSE) estymatora (4.55). Dodatkowo, dla porównania, policzono obciążenie oraz MSE klasycznej empirycznej funkcji dystrybuanty (ECDF) dla $X_i \cdot Y_i$, gdzie $i = 1, 2, ..., n$. W Tabelach 4.10-4.12 przedstawiono wyniki symulacji. Część wyników zebrano na Rysunkach 4.19 - 4.28. Dla każdego z rozważanych rozkładów przedstawiono wykresy obciążenia oraz MSE proponowanego estymatora dla $m = 15$ oraz estymatora ECDF. Ciągłe czarne i czerwone linie odpowiadają obciążeniu dla, odpowiednio, estymatora (4.55) oraz ECDF. Linia przerywana czarna i czerwona odpowiada MSE, odpowiednio, estymatora (4.55) oraz ECDF.

Tabela 4.10: Obciążenie i MSE dla $X, Y \sim \exp(1; 0.5)$.

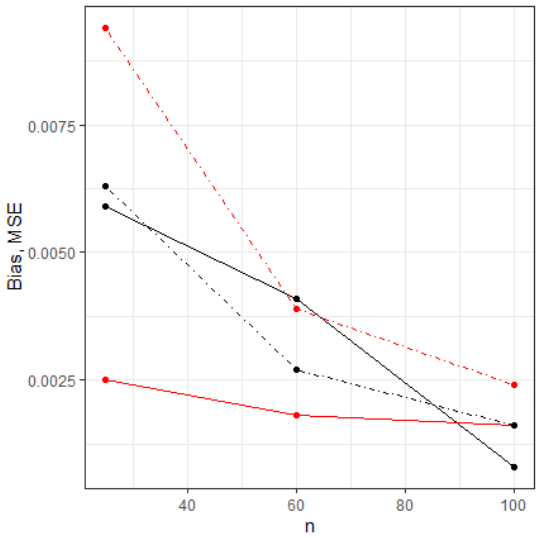
$n = 25$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0041	0.0037	0.0021	0.0036	0.0013	0.0036	0.0029	0.0067
0.4	0.0104	0.0065	0.0057	0.0066	0.0045	0.0066	0.0012	0.0101
0.6	0.0197	0.0064	0.0084	0.0063	0.0059	0.0063	0.0025	0.0094
0.8	0.0392	0.0056	0.0128	0.0043	0.0074	0.0040	0.0031	0.0067
$n = 60$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0022	0.0015	0.0012	0.0015	0.0009	0.0015	0.0007	0.0025
0.4	0.0053	0.0028	0.0025	0.0029	0.0017	0.0029	0.0011	0.0043
0.6	0.0174	0.0029	0.0066	0.0027	0.0041	0.0027	0.0018	0.0039
0.8	0.0366	0.0017	0.0092	0.0017	0.0039	0.0016	0.0019	0.0026
$n = 100$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0009	0.0008	0.0004	0.0008	0.0002	0.0008	0.0005	0.0014
0.6	0.0044	0.0014	0.0012	0.0015	0.0005	0.0014	0.0007	0.0022
0.6	0.0137	0.0017	0.0029	0.0016	0.0008	0.0016	0.0016	0.0024
0.8	0.0370	0.0022	0.0090	0.0010	0.0032	0.0009	0.0015	0.0016



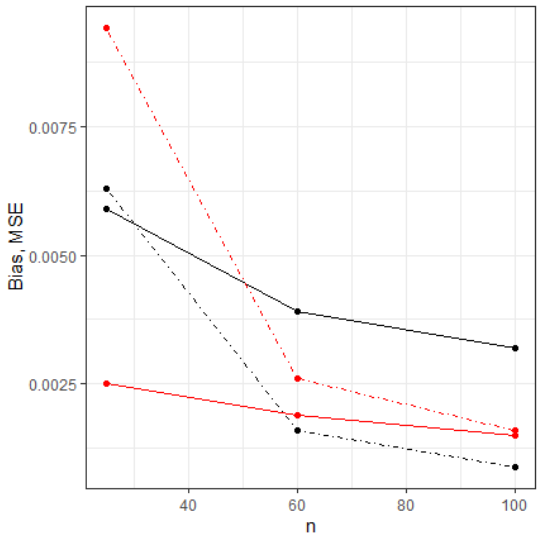
Rysunek 4.17: Obciążenie i MSE dla $F_{Z_2}(z) = 0.2$, gdy $X, Y \sim \exp(1; 0.5)$



Rysunek 4.18: Obciążenie i MSE dla $F_{Z_2}(z) = 0.4$, gdy $X, Y \sim \exp(1; 0.5)$



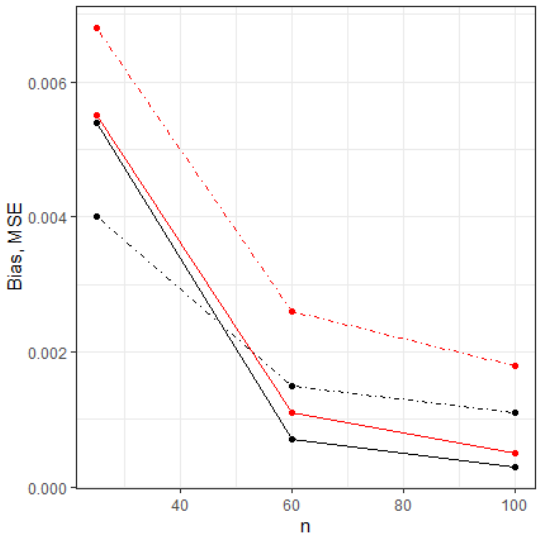
Rysunek 4.19: Obciążenie i MSE dla $F_{Z_2}(z) = 0.6$, gdy $X, Y \sim \exp(1; 0.5)$



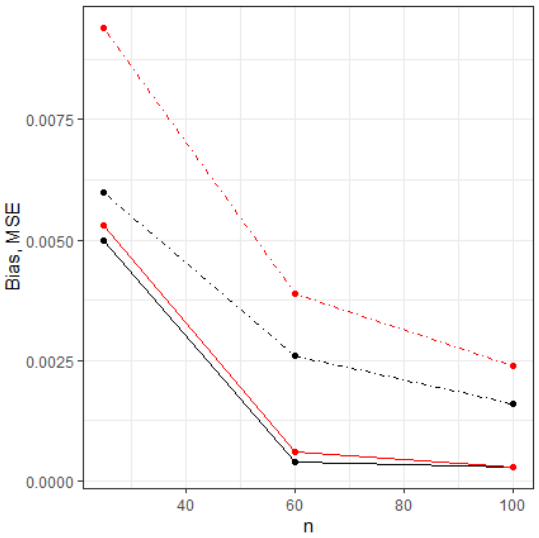
Rysunek 4.20: Obciążenie i MSE dla $F_{Z_2}(z) = 0.8$, gdy $X, Y \sim \exp(1; 0.5)$

Tabela 4.11: Obciążenie i MSE dla $X, Y \sim \text{Weibull}(2; 1; 0.5)$.

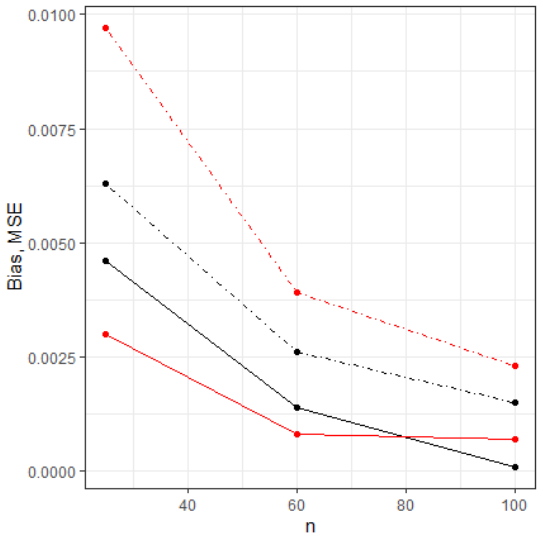
$n = 25$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0061	0.0041	0.0059	0.0040	0.0054	0.0040	0.0055	0.0068
0.4	0.0070	0.0057	0.0059	0.0060	0.0050	0.0060	0.0053	0.0094
0.6	0.0132	0.0062	0.0050	0.0062	0.0046	0.0063	0.0030	0.0097
0.8	0.0340	0.0047	0.0087	0.0037	0.0048	0.0037	0.0051	0.0061
$n = 60$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0047	0.0016	0.0013	0.0015	0.0007	0.0015	0.0011	0.0026
0.4	0.0043	0.0025	0.0015	0.0026	0.0004	0.0026	0.0006	0.0039
0.6	0.0104	0.0024	0.0023	0.0026	0.0014	0.0026	0.0008	0.0039
0.8	0.0331	0.0026	0.0080	0.0016	0.0032	0.0015	0.0016	0.0026
$n = 100$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0037	0.0012	0.0012	0.0011	0.0003	0.0011	0.0005	0.0018
0.4	0.0041	0.0015	0.0008	0.0016	0.0003	0.0016	0.0003	0.0024
0.6	0.0091	0.0014	0.0010	0.0015	0.0001	0.0015	0.0007	0.0023
0.8	0.0307	0.0020	0.0079	0.0010	0.0029	0.0010	0.0015	0.0016



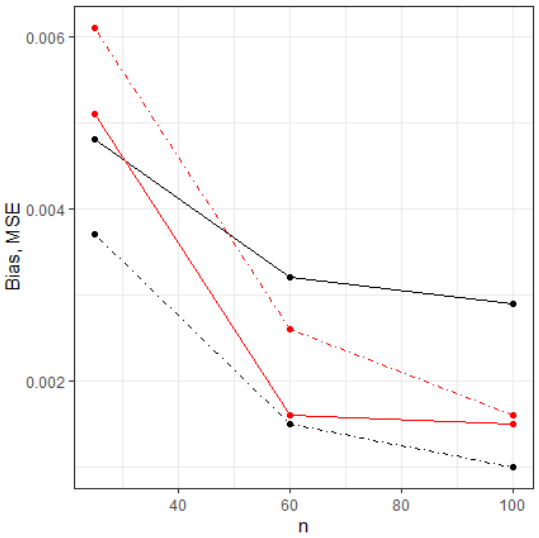
Rysunek 4.21: Obciążenie i MSE dla $F_{Z_2}(z) = 0.2$, gdy $X, Y \sim \text{Weibull}(2;1;0.5)$



Rysunek 4.22: Obciążenie i MSE dla $F_{Z_2}(z) = 0.4$, gdy $X, Y \sim \text{Weibull}(2;1;0.5)$



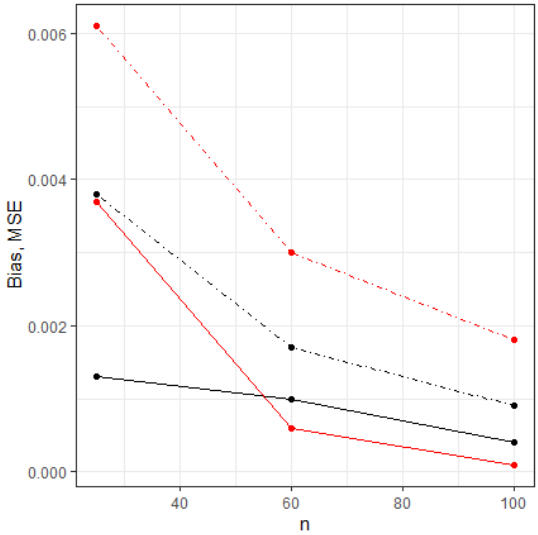
Rysunek 4.23: Obciążenie i MSE dla $F_{Z_2}(z) = 0.6$, gdy $X, Y \sim \text{Weibull}(2;1;0.5)$



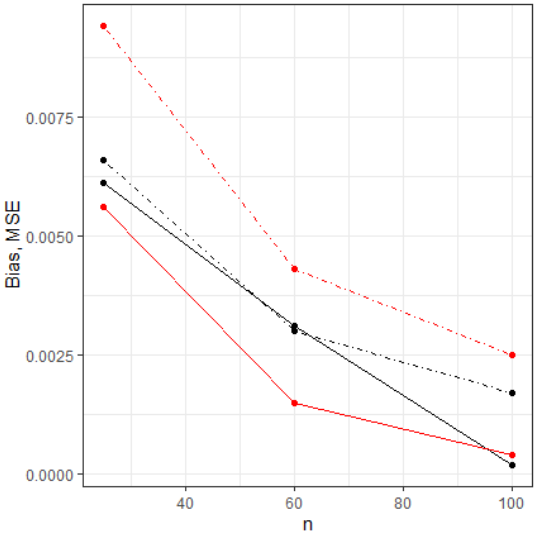
Rysunek 4.24: Obciążenie i MSE dla $F_{Z_2}(z) = 0.8$, gdy $X, Y \sim \text{Weibull}(2;1;0.5)$

Tabela 4.12: Obciążenie i MSE dla $X, Y \sim \text{Pareto}(1;3)$.

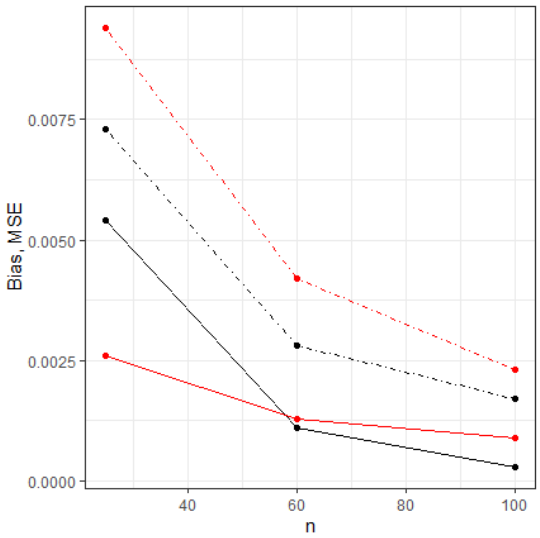
$n = 25$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0042	0.0037	0.0021	0.0037	0.0013	0.0038	0.0037	0.0061
0.4	0.0116	0.0066	0.0072	0.0066	0.0061	0.0066	0.0056	0.0094
0.6	0.0131	0.0075	0.0069	0.0073	0.0054	0.0073	0.0026	0.0094
0.8	0.0139	0.0052	0.0047	0.0049	0.0033	0.0049	0.0021	0.0063
$n = 60$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0033	0.0016	0.0013	0.0017	0.0010	0.0017	0.0006	0.0030
0.4	0.0081	0.0030	0.0039	0.0029	0.0031	0.0030	0.0015	0.0043
0.6	0.0090	0.0029	0.0024	0.0028	0.0011	0.0028	0.0013	0.0042
0.8	0.0118	0.0020	0.0035	0.0018	0.0018	0.0018	0.0016	0.0025
$n = 100$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0023	0.0009	0.0008	0.0009	0.0004	0.0009	0.0001	0.0018
0.4	0.0047	0.0017	0.0010	0.0017	0.0002	0.0017	0.0004	0.0025
0.6	0.0076	0.0017	0.0016	0.0017	0.0003	0.0017	0.0009	0.0023
0.8	0.0113	0.0013	0.0022	0.0011	0.0006	0.0011	0.0004	0.0015



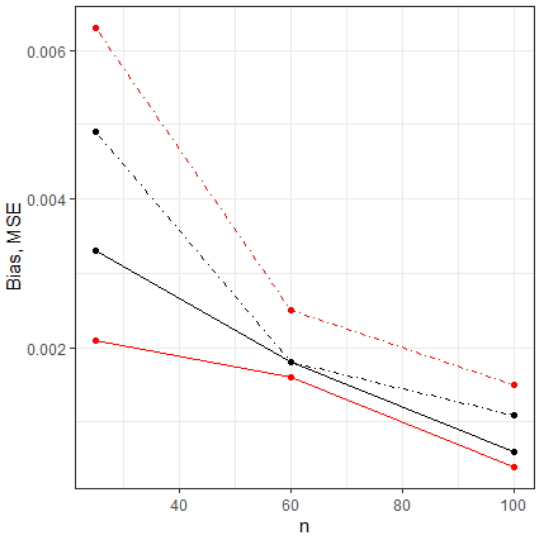
Rysunek 4.25: Obciążenie i MSE dla $F_{Z_2}(z) = 0.2$, gdy $X, Y \sim \text{Pareto}(1;3)$



Rysunek 4.26: Obciążenie i MSE dla $F_{Z_2}(z) = 0.4$, gdy $X, Y \sim \text{Pareto}(1;3)$



Rysunek 4.27: Obciążenie i MSE dla $F_{Z_2}(z) = 0.6$, gdy $X, Y \sim \text{Pareto}(1;3)$



Rysunek 4.28: Obciążenie i MSE dla $F_{Z_2}(z) = 0.8$, gdy $X, Y \sim \text{Pareto}(1;3)$

Na podstawie powyższych tabel i wykresów można sformułować wspólne wnioski dla wszystkich badanych rozkładów. W każdym przypadku, wraz ze wzrostem liczności próby obciążenie i MSE estymatora (4.55) maleją. W przypadku obciążenia nie można wskazać jednoznacznej przewagi żadnego z estymatorów. W zależności od rozważanego przypadku estymator (4.55) osiąga wartości zarówno większe, jak i mniejsze od odpowiadających im wartości uzyskanych dla ECDF. Bardzo ważnym wnioskiem, który świadczy o efektywności estymatora (4.55), jest fakt, że MSE estymatora (4.55) jest w każdym przypadku mniejsze niż MSE dla ECDF.

Ocena estymatora nieparametrycznego ($n_1 \neq n_2$)

Badania symulacyjne dla estymatora (4.55) przeprowadzono analogicznie jak w przypadku różnicy dwóch zmiennych losowych. Zastosowano tę samą procedurę generowania prób, liczbę powtórzeń symulacji, rozmiary prób oraz wartości parametru m. Próbki generowano z rozkładu wykładniczego, Weibulla lub Pareto. Dla każdego rozważanego przypadku obliczono obciążenie (bias) oraz błąd średniokwadratowy (MSE) estymatora (4.55). Dodatkowo, dla porównania, wyznaczono obciążenia i MSE dla estymatora opartego na wszystkich iloczynach par obserwacji (ozn. jako *wszystkie pary*). Uzyskane wyniki przedstawiono w Tabelach 4.13 oraz 4.14.

Tabela 4.13: Obciążenie i MSE dla $X, Y \sim \exp(1; 0.5)$.

$n_1 = 30$, $n_2 = 25$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0072	0.0034	0.0051	0.0034	0.0045	0.0034	0.0027	0.0035
0.4	0.0098	0.0060	0.0049	0.0060	0.0036	0.0061	0.0014	0.0061
0.6	0.0151	0.0055	0.0030	0.0056	0.0005	0.0056	0.0029	0.0056
0.8	0.0345	0.0048	0.0084	0.0036	0.0033	0.0035	0.0014	0.0034
$n_1 = 60$, $n_2 = 65$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0038	0.0013	0.0031	0.0013	0.0028	0.0013	0.0022	0.0013
0.4	0.0066	0.0026	0.0036	0.0027	0.0024	0.0027	0.0012	0.0027
0.6	0.0131	0.0025	0.0025	0.0024	0.0007	0.0024	0.0021	0.0024
0.8	0.0330	0.0029	0.0081	0.0017	0.0027	0.0016	0.0013	0.0016
$n_1 = 105$, $n_2 = 100$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0030	0.0009	0.0025	0.0009	0.0024	0.0009	0.0010	0.0009
0.4	0.0032	0.0015	0.0004	0.0016	0.0002	0.0016	0.0010	0.0016
0.6	0.0121	0.0017	0.0021	0.0016	0.0006	0.0016	0.0015	0.0016
0.8	0.0321	0.0023	0.0079	0.0010	0.0021	0.0009	0.0024	0.0009

Tabela 4.14: Obciążenie i MSE dla $X, Y \sim \text{Weibull}(2;1;0.5)$.

$n_1 = 30$, $n_2 = 25$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0081	0.0036	0.0071	0.0035	0.0060	0.0035	0.0045	0.0034
0.4	0.0058	0.0062	0.0056	0.0062	0.0031	0.0063	0.0028	0.0063
0.6	0.0138	0.0056	0.0062	0.0057	0.0056	0.0057	0.0031	0.0057
0.8	0.0311	0.0046	0.0105	0.0037	0.0074	0.0035	0.0026	0.0035
$n_1 = 60$, $n_2 = 65$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0077	0.0018	0.0028	0.0017	0.0017	0.0017	0.0023	0.0017
0.4	0.0053	0.0025	0.0024	0.0026	0.0013	0.0027	0.0019	0.0027
0.6	0.0131	0.0025	0.0045	0.0025	0.0038	0.0025	0.0018	0.0025
0.8	0.0309	0.0027	0.0089	0.0015	0.0035	0.0015	0.0012	0.0015
$n_1 = 105$, $n_2 = 100$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0070	0.0010	0.0011	0.0009	0.0003	0.0009	0.0005	0.0009
0.4	0.0050	0.0015	0.0015	0.0015	0.0003	0.0015	0.0001	0.0015
0.6	0.0106	0.0015	0.0016	0.0015	0.0007	0.0015	0.0008	0.0016
0.8	0.0301	0.0021	0.0080	0.0009	0.0033	0.0009	0.0005	0.0008

Tabela 4.15: Obciążenie i MSE dla $X, Y \sim \text{Pareto}(1;3)$.

$n_1 = 30$, $n_2 = 25$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0054	0.0033	0.0029	0.0034	0.0022	0.0034	0.0011	0.0034
0.4	0.0077	0.0065	0.0034	0.0065	0.0024	0.0065	0.0011	0.0065
0.6	0.0125	0.0067	0.0058	0.0067	0.0041	0.0067	0.0029	0.0067
0.8	0.0154	0.0044	0.059	0.0042	0.0041	0.0043	0.0025	0.0042
$n_1 = 60$, $n_2 = 65$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0037	0.0014	0.0019	0.0015	0.0015	0.0015	0.0010	0.0015
0.4	0.0068	0.0026	0.0026	0.0026	0.0017	0.0026	0.0007	0.0026
0.6	0.0104	0.0027	0.0041	0.0027	0.0029	0.0027	0.0020	0.0027
0.8	0.0129	0.0019	0.0040	0.0018	0.0023	0.0017	0.0009	0.0017
$n_1 = 105$, $n_2 = 100$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_2}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0029	0.0009	0.0012	0.0009	0.0009	0.0009	0.0005	0.0009
0.4	0.0045	0.0016	0.0007	0.0016	0.0001	0.0016	0.0006	0.0016
0.6	0.0074	0.0017	0.0009	0.0017	0.0004	0.0017	0.0014	0.0017
0.8	0.0111	0.0013	0.0021	0.0011	0.0005	0.0011	0.0009	0.0011

Powyższe wyniki są analogiczne z obserwacjami przedstawionymi dla przypadku różnicy. Wraz ze wzrostem wartości parametru m następuje wyraźne zmniejszenie obciążenia proponowanego estymatora, przy czym dla $m = 15$ osiąga ono wartości porównywalne z estymatorem opartym na wszystkich parach obserwacji, a w wybranych przypadkach jest nawet od niego mniejsze.

Jednocześnie wartości błędu średniokwadratowego wszystkich analizowanych estymatorów pozostają bardzo zbliżone. Podobnie jak w przypadku estymacji dystrybuanty różnicy, wzrost liczebności prób prowadzi do zmniejszenia wartości MSE, co potwierdza poprawę jakości oszacowania wraz ze wzrostem liczby obserwacji.

4.5. Aproksymacja i estymacja rozkładu ilorazu zmiennych losowych

W tej części ponownie wykorzystano metodę trapezów, która została zastosowana w przypadku różnicy i iloczynu zmiennych losowych, do aproksymacji dystrybuanty ilorazu dodatnich zmiennych losowych X i Y.

4.5.1. Aproksymacja rozkładu

Ponownie, jak w przypadku różnicy zmiennych losowych, niezbędne będzie wprowadzenie oznaczenia

$$Q_Y^{\delta}(1) := Q_Y(1-\delta) \quad \text{dla dowolnej} \quad 0 \leqslant \delta < 1.$$

Najpierw rozpatrzmy przypadek $0 < z < 1$. Gdy $Q_Y(1) = \infty$, prawdziwe jest następujące oszacowanie

$$\forall_{0<z<1} \forall_{0<\delta\leqslant 1-F_Y\left(\frac{Q_X(1-\varepsilon)}{z}\right)} \quad F_X(zQ_Y(1)) - F_X(zQ_Y^{\delta}(1)) \leqslant \varepsilon. \tag{4.58}$$

Rozważmy następujący przedział $\left[F_X\left(Q_X(0)\right), F_X\left(zQ_Y^{\delta}(1)\right)\right]$ jako dziedzinę całkowania. Dzielimy ten przedział na m podprzedziałów za pomocą punktów

$$\left\{F_X\left(Q_X(0)\right), F_X\left(\frac{zQ_Y^{\delta}(1)+(m-1)Q_X(0)}{m}\right), F_X\left(\frac{2zQ_Y^{\delta}(1)+(m-2)Q_X(0)}{m}\right), ..., F_X\left(zQ_Y^{\delta}(1)\right)\right\}.$$

Następnie obliczamy odpowiadające im wartości zgodnie ze wzorem $u_2 = F_Y\left(\frac{Q_X(u_1)}{z}\right)$:

$$\left\{F_Y\left(\frac{Q_X(0)}{z}\right), F_Y\left(\frac{zQ_Y^{\delta}(1)+(m-1)Q_X(0)}{zm}\right), F_Y\left(\frac{2zQ_Y^{\delta}(1)+(m-2)Q_X(0)}{zm}\right), ..., F_Y(Q_Y^{\delta}(1))\right\}.$$

Korzystając z reguły trapezów oraz faktu, że X i Y są niezależnymi zmiennymi losowymi otrzymujemy przybliżenie $F_{Z_3}(z)$ dane wzorem

$$F_{Z_{3m}}(z) = F_X\left(zQ_Y^{\delta}(1)\right) - \frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(\frac{(i+1)zQ_Y^{\delta}(1)+(m-i-1)Q_X(0)}{m}\right)\right.$$

$$\left.-F_X\left(\frac{izQ_Y^{\delta}(1)+(m-i)Q_X(0)}{m}\right)\right)\cdot\left(F_Y\left(\frac{(i+1)zQ_Y^{\delta}(1)+(m-i-1)Q_X(0)}{zm}\right)\right.$$

$$\left.+F_Y\left(\frac{izQ_Y^{\delta}(1)+(m-i)Q_X(0)}{zm}\right)\right), \tag{4.59}$$

gdzie m oznacza liczbę podprzedziałów.

Rozważmy teraz $z \geqslant 1$. W przypadku $Q_X(1) = \infty$ mamy

$$\forall_{z\geqslant 1}\forall_{\varepsilon>0}\exists_{0<\delta<1} \quad F_X(Q_X(1)) - F_X(Q_X^{\delta}(1)) = 1 - F_X(Q_X^{\delta}(1)) < \varepsilon. \tag{4.60}$$

Przyjmujemy przedział $[F_X(zQ_Y(0)), F_X(Q_X^{\delta}(1))]$ jako dziedzinę całkowania, dzielimy go na m podprzedziałów za pomocą punktów

$$\left\{F_X(zQ_Y(0)), F_X\left(\frac{zQ_Y(0)(m-1)+Q_X^{\delta}(1)}{m}\right), F_X\left(\frac{zQ_Y(0)(m-2)+2Q_X^{\delta}(1)}{m}\right), ..., F_X(Q_X^{\delta}(1))\right\}.$$

Następnie obliczamy odpowiadające im wartości u_2:

$$\left\{F_Y(Q_Y(0)), F_Y\left(\frac{zQ_Y(0)(m-1)+Q_X^{\delta}(1)}{zm}\right), F_Y\left(\frac{zQ_Y(0)(m-2)+2Q_X^{\delta}(1)}{zm}\right), ..., F_Y\left(\frac{Q_Y^{\delta}(1)}{z}\right)\right\}.$$

Ponownie korzystając z metody trapezów oraz faktu, że X i Y są niezależnymi zmiennymi losowymi, otrzymujemy przybliżenie

$$F_{Z_{3m}}(z) = F_X\left(Q_X^{\delta}(1)\right) - \frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{m}\right)\right.$$

$$\left.-F_X\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{m}\right)\right)\cdot\left(F_Y\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{zm}\right)\right.$$

$$\left.+F_Y\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{zm}\right)\right). \tag{4.61}$$

W kolejnym paragrafie zostanie oszacowany błąd tej aproksymacji.

4.5.2. Błąd aproksymacji

Poniższe twierdzenie pozwala na oszacowanie błędu aproksymacji danej wzorami (4.59), (4.61).

Twierdzenie 4.8. *Niech $\varepsilon > 0$. Niech $X > 0$ będzie zmienną losową o co najmniej raz różniczkowalnej dystrybuancie F_X na przedziale $(Q_X(0), Q_X(1))$. Niech $Y > 0$ będzie zmienną losową o co najmniej dwukrotnie różniczkowalnej dystrybuancie F_Y na przedziale $(Q_Y(0), Q_Y(1))$, przy czym $Q_Y(0) > 0$. Dodatkowo zakładamy, że F_X' i F_Y'' są ciągłe. Wówczas:*

(i) dla $0 < z < 1$ istnieje $0 \leqslant \delta \leqslant 1 - F_Y\left(\frac{Q_X(1-\varepsilon)}{z}\right)$, taka że oszacowanie błędu aproksymacji $F_{Z_3}(z)$ przez $F_{Z_{3m}}(z)$ wyraża się wzorem

$$|F_{Z_3}(z) - F_{Z_{3m}}(z)| \leqslant 2\varepsilon + \frac{M_1(z)[M_2(z)]^3\left(zQ_Y^{\delta}(1) - Q_X(0)\right)^3}{12m^2}, \tag{4.62}$$

gdzie

$$M_1(z) = \max_{F_X(Q_X(0))\leqslant\eta\leqslant F_X(zQ_Y^{\delta}(1))}\left|F_Y''\left(\frac{Q_X(\eta)}{z}\right)\right|, \quad M_2(z) = \max_{Q_X(0)\leqslant\tau\leqslant zQ_Y^{\delta}(1)}|F_X'(\tau)|$$

oraz m jest liczbą podprzedziałów;

(ii) dla $z \geqslant 1$ *istnieje* $0 \leqslant \delta < 1$, *taka że oszacowanie błędu aproksymacji* $F_{Z_3}(z)$ *przez* $F_{Z_{3m}}(z)$ *wynosi*

$$|F_{Z_3}(z) - F_{Z_{3m}}(z)| \leqslant 2\varepsilon + \frac{M_1(z)[M_2(z)]^3\left(Q_X^{\delta}(1) - zQ_Y(0)\right)^3}{12m^2}, \tag{4.63}$$

gdzie

$$M_1^*(z) = \max_{F_X(zQ_Y(0)) \leqslant \eta \leqslant F_X(Q_X^{\delta}(1))} \left|F_Y''\left(\frac{Q_X(\eta)}{z}\right)\right| \quad \textit{and} \quad M_2^*(z) = \max_{zQ_Y(0) \leqslant \tau \leqslant Q_X^{\delta}(1)} |F_X'(\tau)|.$$

Dowód. Najpierw rozważamy przypadek $0 < z < 1$. Zauważmy, że na podstawie (4.37) oraz (4.59) mamy

$$|F_{Z_3}(z) - F_{Z_{3m}}(z)| \leqslant |F_X(zQ_Y(1)) - F_X(zQ_Y^{\delta}(1)))| + \left|\int_{F_X(zQ_Y^{\delta}(1))}^{F_X(zQ_Y(1))} 1du_1\right|$$

$$+\left|\int_{F_X(Q_X(0))}^{F_X(zQ_Y^{\delta}(1))} F_Y\left(\frac{Q_X(u_1)}{z}\right)du_1 - \frac{1}{2}\sum_{i=0}^{m-1}\left(F_X\left(\frac{(i+1)zQ_Y^{\delta}(1)+(m-i-1)Q_X(0)}{m}\right)\right)\right.$$

$$-F_X\left(\frac{izQ_Y^{\delta}(1)+(m-i)Q_X(0)}{m}\right)\Bigg)\cdot\left(F_Y\left(\frac{(i+1)zQ_Y^{\delta}(1)+(m-i-1)Q_X(0)}{zm}\right)\right.$$

$$\left.\left.+F_Y\left(\frac{izQ_Y^{\delta}(1)+(m-i)Q_X(0)}{zm}\right)\right)\right|.$$

Z (4.58) oraz równania (5.1.4) w pracy Atkinsona [1] otrzymujemy

$$|F_{Z_3}(z) - F_{Z_{3m}}(z)|$$

$$\leqslant 2\varepsilon + \sum_{i=0}^{m-1}\frac{\left(F_X\left(\frac{(i+1)zQ_Y^{\delta}(1)+(m-i-1)Q_X(0)}{m}\right) - F_X\left(\frac{izQ_Y^{\delta}(1)+(m-i)Q_X(0)}{m}\right)\right)^3}{12}\left|F_Y''\left(\frac{Q_X(\eta_i)}{z}\right)\right|,$$

gdzie

$$\eta_i \in \left[F_X\left(\frac{izQ_Y^{\delta}(1)+(m-i)Q_X(0)}{m}\right), F_X\left(\frac{(i+1)zQ_Y^{\delta}(1)+(m-i-1)Q_X(0)}{m}\right)\right].$$

Na podstawie twierdzenia o wartości średniej dostajemy

$$|F_{Z_3}(z) - F_{Z_{3m}}(z)| \leqslant 2\varepsilon + \sum_{i=0}^{m-1}\frac{\left(F_X'(\tau_i)\left(zQ_Y^{\delta}(1) - Q_X(0)\right)\right)^3}{12m^3}\left|F_Y''\left(\frac{Q_X(\eta_i)}{z}\right)\right|,$$

gdzie $\tau_i \in \left[\frac{izQ_Y^{\delta}(1)+(m-i)Q_X(0)}{m}, \frac{(i+1)zQ_Y^{\delta}(1)+(m-i-1)Q_X(0)}{m}\right]$.

Z ciągłości F_Y'' i F_X' mamy

$$\left|F_Y''\left(\frac{Q_X(\eta_i)}{z}\right)\right| \leqslant \max_{F_X\left(\frac{izQ_Y^{\delta}(1)+(m-i)Q_X(0)}{m}\right) \leqslant \eta_i \leqslant F_X\left(\frac{(i+1)zQ_Y^{\delta}(1)+(m-i-1)Q_X(0)}{m}\right)}\left|F_Y''\left(\frac{Q_X(\eta_i)}{z}\right)\right|$$

$$\leqslant M_1(z) = \max_{F_X(Q_X(0))\leqslant\eta\leqslant F_X(zQ_Y^{\delta}(1))} \left| F_Y'' \left(\frac{Q_X(\eta)}{z} \right) \right|$$

oraz

$$|F_X'(\tau_i)| \leqslant \max_{\frac{izQ_Y^{\delta}(1)+(m-i)Q_X(0)}{m}\leqslant\tau_i\leqslant\frac{(i+1)zQ_Y^{\delta}(1)+(m-i-1)Q_X(0)}{m}} |F_X'(\tau_i)|$$
$$\leqslant M_2(z) = \max_{Q_X(0)\leqslant\tau\leqslant zQ_Y^{\delta}(1)} |F_X'(\tau)|.$$

Stąd oszacowanie błędu aproksymacji jest postaci

$$|F_{Z_3}(z) - F_{Z_{3m}}(z)| \leqslant 2\varepsilon + \frac{M_1(z)[M_2(z)]^3 \left(zQ_Y^{\delta}(1) - Q_X(0)\right)^3}{12m^2}, \tag{4.64}$$

co kończy dowód dla $0 < z < 1$.

Dla $z \geqslant 1$, na podstawie (4.37) oraz (4.61) dostajemy

$$|F_{Z_3}(z) - F_{Z_{3m}}(z)| \leqslant |F_X(Q_X(1) - F_X(Q_X^{\delta}(1)))| + \left| \int_{F_X(Q_X^{\delta}(1))}^{F_X(Q_X(1))} 1 du_1 \right|$$

$$+ \left| \int_{F_X(zQ_Y(0))}^{F_X(Q_X^{\delta}(1))} F_Y\left(\frac{Q_X(u_1)}{z} \right) du_1 - \frac{1}{2} \sum_{i=0}^{m-1} \left(F_X\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{m} \right) \right) \right.$$
$$- F_X\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{m} \right) \Bigg) \cdot \left(F_Y\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{zm} \right) \right.$$
$$\left. \left. + F_Y\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{zm} \right) \right) \right|.$$

Na podstawie (4.60) oraz równania (5.1.4) w pracy Atkinsona [1] otrzymujemy

$$|F_{Z_3}(z) - F_{Z_{3m}}(z)|$$
$$\leqslant 2\varepsilon + \sum_{i=0}^{m-1} \frac{\left(F_X\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{m} \right) - F_X\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{m} \right) \right)^3}{12} \left| F_Y''\left(\frac{Q_X(\eta_i)}{z} \right) \right|,$$

gdzie

$$\eta_i \in \left[F_X\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{m} \right), F_X\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{m} \right) \right].$$

Stosując twierdzenie o wartości średniej otrzymujemy

$$|F_{Z_3}(z) - F_{Z_{3m}}(z)| \leqslant 2\varepsilon + \sum_{i=0}^{m-1} \frac{\left(F_X'(\tau_i)\left(Q_X^{\delta}(1) - zQ_Y(0) \right) \right)^3}{12m^3} \left| F_Y''\left(\frac{Q_X(\eta_i)}{z} \right) \right|,$$

gdzie $\tau_i \in \left[\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{m}, \frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{m} \right]$.

Na podstawie ciągłości F_Y'' oraz F_X' dostajemy, że

$$\left|F_Y''\left(\frac{Q_X(\eta_i)}{z}\right)\right| \leqslant \max_{F_X\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{m}\right)\leqslant\eta_i\leqslant F_X\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{m}\right)} \left|F_Y''\left(\frac{Q_X(\eta_i)}{z}\right)\right|$$

$$\leqslant M_1^*(z) = \max_{F_X(zQ_Y(0))\leqslant\eta\leqslant F_X(Q_X^{\delta}(1))} \left|F_Y''\left(\frac{Q_X(\eta)}{z}\right)\right|$$

oraz

$$|F_X'(\tau_i)| \leqslant \max_{\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{m}\leqslant\tau_i\leqslant\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{m}} |F_X'(\tau_i)|$$

$$\leqslant M_2^*(z) = \max_{zQ_Y(0)\leqslant\tau\leqslant Q_X^{\delta}(1)} |F_X'(\tau)|.$$

Stąd oszacowanie błędu aproksymacji jest postaci

$$|F_{Z_3}(z) - F_{Z_{3m}}(z)| \leqslant 2\varepsilon + \frac{M_1(z)[M_2(z)]^3\left(Q_X^{\delta}(1) - zQ_Y(0)\right)^3}{12m^2},$$

co kończy dowód dla $z \geqslant 1$. □

Łatwo zauważyć, że podobnie jak w przypadku różnicy i iloczynu zmiennych losowych, wraz ze wzrostem liczby punktów podziału błąd aproksymacji maleje.

4.5.3. Estymator nieparametryczny

Niech $\{X_1,\ldots,X_{n_1}\}$ oraz $\{Y_1,\ldots,Y_{n_2}\}$ będą niezależnymi próbami losowymi pobranymi z populacji o dystrybuantach, odpowiednio, F_X i F_Y. Niech $\hat{F}_X$ i $\hat{F}_Y$ będą empirycznymi dystrybuantami odpowiadającymi, odpowiednio, F_X oraz F_Y.

Oznaczmy

$$Q_X^*(0) := \min\{X_i,\ \ 1\leqslant i\leqslant n_1\},\quad Q_Y^*(0) := \min\{Y_i,\ \ 1\leqslant i\leqslant n_2\},$$

$$Q_X^*(1) := \max\{X_i,\ \ 1\leqslant i\leqslant n_1\},\quad Q_Y^*(1) := \max\{Y_i,\ \ 1\leqslant i\leqslant n_2\}.$$

Nieparametryczny estymator $F_{Z_3}(z)$, gdzie $Z_3 = \dfrac{X}{Y}$, dany jest wzorem dla

- $0 < z < 1$

$$\hat{F}_{Z_3}(z) = \hat{F}_X\left(zQ_Y^*(1)\right) - \frac{1}{2}\sum_{i=0}^{m-1}\left(\hat{F}_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\right)\right.$$

$$\left.-\hat{F}_X\left(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{m}\right)\right)\cdot\left(\hat{F}_Y\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{zm}\right)\right.$$

$$\left.+\hat{F}_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{zm}\right)\right), \tag{4.65}$$

- $z \geqslant 1$

$$\hat{F}_{Z_3}(z) = \hat{F}_X\big(Q_X^*(1)\big) - \frac{1}{2}\sum_{i=0}^{m-1}\left(\hat{F}_X\left(\frac{zQ_Y^*(0)(m-i-1)+(i+1)Q_X^*(1)}{m}\right)\right.$$
$$\left.-\hat{F}_X\left(\frac{zQ_Y^*(0)(m-i)+iQ_X^*(1)}{m}\right)\right)\cdot\left(\hat{F}_Y\left(\frac{zQ_Y^*(0)(m-i-1)+(i+1)Q_X^*(1)}{zm}\right)\right.$$
$$\left.+\hat{F}_Y\left(\frac{zQ_Y^*(0)(m-i)+iQ_X^*(1)}{zm}\right)\right). \tag{4.66}$$

Poniższe twierdzenie dowodzi silnej zgodność estymatora $\hat{F}_{Z_3}(z)$.

Twierdzenie 4.9. *Niech $Z_3 = \dfrac{X}{Y}$, gdzie $X > 0$ oraz $Y > 0$ są zmiennymi losowymi o ciągłych dystrybuantach, odpowiednio, F_X oraz F_Y, spełniającymi założenia Twierdzenia 4.8. Niech $\hat{F}_{Z_3}(z)$ będzie zdefiniowany wzorem (4.65) dla $0 < z < 1$ oraz wzorem (4.66) dla $z \geqslant 1$. Wówczas*

$$|\hat{F}_{Z_3}(z) - F_{Z_3}(z)| \to 0,$$

gdy $n_1, n_2 \to \infty$.

Dowód. Zauważmy najpierw, że

$$|\hat{F}_{Z_3}(z) - F_{Z_3}(z)| \leqslant |\hat{F}_{Z_3}(z) - F_{Z_{3m}}(z)| + |F_{Z_{3m}}(z) - F_{Z_3}(z)|. \tag{4.67}$$

Rozważać będziemy trzy przypadki.

1. Niech $0 < z < 1$ i $Q_Y(1) = \infty$. Wprowadźmy następujące oznaczenia: $\varepsilon = \varepsilon(n_2, z) = 1 - F_X(zQ_Y^*(1))$ oraz $\delta = \delta(n_2) = 1 - F_Y(Q_Y^*(1))$. Stąd mamy

$$Q_Y^{\delta}(1) = Q_Y^*(1). \tag{4.68}$$

Łatwo zauważyć, że $\varepsilon \to 0$ oraz $\delta \to 0$ gdy $n_2 \to \infty$.

Zauważmy, że dla ustalonego n_2 i ustalonego $0 < z < 1$ mamy

$$\lim_{m\to\infty} \frac{M_1(z)[M_2(z)]^3\big(zQ_Y^{\delta}(1) - Q_X(0)\big)^3}{12m^2} = 0$$

$$\iff \forall_{\varepsilon>0}\exists_{m_0=m_0(z,n_2)}\forall_{m\geqslant m_0} \frac{M_1(z)[M_2(z)]^3\big(zQ_Y^{\delta}(1) - Q_X(0)\big)^3}{12m^2} < \varepsilon.$$

Stąd na podstawie Twierdzenia 4.8(i) otrzymujemy

$$\forall_{n_2}\forall_{0<z<1}\exists_{\varepsilon}\exists_{m_0}\forall_{m\geqslant m_0}|F_{Z_{3m}}(z) - F_{Z_3}(z)| < 3\varepsilon,$$

co pokazuje, że

$$|F_{Z_{3m}}(z) - F_{Z_3}(z)| \to 0, \quad \text{gdy} \quad n_2 \to \infty \quad \text{dla} \quad 0 < z < 1. \tag{4.69}$$

Ponadto, dla $0 < z < 1$ mamy

$$\hat{F}_{Z_3}(z) - F_{Z_{3m}}(z) = \hat{F}_X(zQ_Y^*(1)) - F_X(zQ_Y^\delta(1))$$

$$+\frac{1}{2}\sum_{i=0}^{m-1}\Bigg\{\Bigg(F_X\Bigg(\frac{(i+1)zQ_Y^\delta(1)+(m-i-1)Q_X(0)}{m}\Bigg)$$

$$-F_X\Bigg(\frac{izQ_Y^\delta(1)+(m-i)Q_X(0)}{m}\Bigg)\Bigg)\cdot\Bigg(F_Y\Bigg(\frac{(i+1)zQ_Y^\delta(1)+(m-i-1)Q_X(0)}{zm}\Bigg)$$

$$+F_Y\Bigg(\frac{izQ_Y^\delta(1)+(m-i)Q_X(0)}{zm}\Bigg)\Bigg)-\Bigg(\hat{F}_X\Bigg(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\Bigg)$$

$$-\hat{F}_X\Bigg(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{m}\Bigg)\Bigg)\cdot\Bigg(\hat{F}_Y\Bigg(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{zm}\Bigg)$$

$$+\hat{F}_Y\Bigg(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{zm}\Bigg)\Bigg)\Bigg\} = \hat{F}_X(zQ_Y^*(1)) - F_X(zQ_Y^\delta(1))$$

$$+\frac{1}{2}\sum_{i=0}^{m-1}\Bigg\{\Bigg[F_X\Bigg(\frac{(i+1)zQ_Y^\delta(1)+(m-i-1)Q_X(0)}{m}\Bigg)F_Y\Bigg(\frac{izQ_Y^\delta(1)+(m-i)Q_X(0)}{zm}\Bigg)$$

$$-\hat{F}_X\Bigg(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\Bigg)\hat{F}_Y\Bigg(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{zm}\Bigg)\Bigg]$$

$$+\Bigg[\hat{F}_X\Bigg(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{m}\Bigg)\hat{F}_Y\Bigg(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{zm}\Bigg)$$

$$-F_X\Bigg(\frac{izQ_Y^\delta(1)+(m-i)Q_X(0)}{m}\Bigg)F_Y\Bigg(\frac{izQ_Y^\delta(1)+(m-i)Q_X(0)}{zm}\Bigg)\Bigg]$$

$$+\Bigg[F_X\Bigg(\frac{(i+1)zQ_Y^\delta(1)+(m-i-1)Q_X(0)}{m}\Bigg)F_Y\Bigg(\frac{(i+1)zQ_Y^\delta(1)+(m-i-1)Q_X(0)}{zm}\Bigg)$$

$$-\hat{F}_X\Bigg(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\Bigg)\hat{F}_Y\Bigg(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{zm}\Bigg)\Bigg]$$

$$+\Bigg[\hat{F}_X\Bigg(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{m}\Bigg)\hat{F}_Y\Bigg(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{zm}\Bigg)$$

$$-F_X\Bigg(\frac{izQ_Y^\delta(1)+(m-i)Q_X(0)}{m}\Bigg)F_Y\Bigg(\frac{(i+1)zQ_Y^\delta(1)+(m-i-1)Q_X(0)}{zm}\Bigg)\Bigg]\Bigg\}$$

$$= \hat{F}_X(zQ_Y^*(1)) - F_X(zQ_Y^\delta(1)) + \frac{1}{2}\sum_{i=0}^{m-1}\{E_i(z)+F_i(z)+G_i(z)+H_i(z)\}.$$

Na podstawie (4.68) oraz faktu, że $|\hat{F}_X(t) - F_X(t)| \to 0$ $p.p.$, gdy $n_1 \to \infty$ dostajemy

$$|\hat{F}_X(zQ_Y^*(1)) - F_X(zQ_Y^\delta(1))| \to 0 \quad p.p., \quad \text{gdy} \quad n_1 \to \infty.$$

Następnie korzystając z (4.68) otrzymujemy

$$
\begin{aligned}
|E_i(z)| = \Bigg| & F_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X(0)}{m}\right)F_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X(0)}{zm}\right) \\
& -\hat{F}_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\right)\hat{F}_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{zm}\right)\Bigg| \\
\leqslant & \left|F_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X(0)}{m}\right)-F_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\right)\right| \\
& \cdot F_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X(0)}{zm}\right) \\
& +\left|F_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\right)-\hat{F}_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\right)\right| \\
& \cdot F_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X(0)}{zm}\right) \\
& +\left|F_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X(0)}{zm}\right)-F_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{zm}\right)\right| \\
& \cdot \hat{F}_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\right) \\
& +\left|F_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{zm}\right)-\hat{F}_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{zm}\right)\right| \\
& \cdot \hat{F}_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\right) = K_1+K_2+K_3+K_4.
\end{aligned}
$$

Na podstawie założenia o ciągłości F_X dostajemy

$$K_1 \leqslant \left|F_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X(0)}{m}\right)-F_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\right)\right| \to 0,$$

gdy $n_1 \to \infty$.

Ponadto, z faktu, że $|\hat{F}_X(x)-F_X(x)| \to 0$ $p.p.$ otrzymujemy

$$K_2 \leqslant \left|F_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\right)-\hat{F}_X\left(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\right)\right| \to 0,$$

$p.p.$ gdy $n_1 \to \infty$.

Ponownie, na podstawie założenia o ciągłości F_Y mamy

$$K_3 \leqslant \left|F_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X(0)}{zm}\right)-F_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{zm}\right)\right| \to 0, \text{ gdy } n_2 \to \infty$$

oraz z faktu, że $|\hat{F}_Y(y)-F_Y(y)| \to 0$ $p.p.$ uzyskujemy

$$K_4 \leqslant \left|F_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{zm}\right)-\hat{F}_Y\left(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{zm}\right)\right| \to 0$$

$p.p.$, gdy $n_2 \to \infty$.

Zatem

$$|E_i(z)| \to 0 \quad p.p., \quad \text{gdy} \quad n_1, n_2 \to \infty. \tag{4.70}$$

Łatwo pokazać, że $|F_i(z)| \to 0$, $|G_i(z)| \to 0$, $|H_i(z)| \to 0$ $p.p.$, gdy $n_1, n_2 \to \infty$. Stąd dla $0 < z < 1$

$$|\hat{F}_{Z_3}(z) - F_{Z_{3m}}(z)| \to 0 \quad p.p., \tag{4.71}$$

gdy $n_1, n_2 \to \infty$.

Zatem na podstawie (4.67), (4.69) oraz (4.71), otrzymujemy

$$|\hat{F}_{Z_3}(z) - F_{Z_3}(z)| \to 0 \;\; p.p., \;\; \text{gdy} \;\; n_1, n_2 \to \infty.$$

2. Niech $0 < z < 1$ oraz $Q_Y(1) < \infty$. Gdy $Q_Y(1) < \infty$ przyjmujemy $\delta = 0$ i na podstawie (4.64) mamy

$$|F_{Z_3}(z) - F_{Z_{3m}}(z)| \leqslant \frac{M_1(z)[M_2(z)]^3\big(zQ_Y(1) - Q_X(0)\big)^3}{12m^2},$$

co dowodzi, że dla dowolnego n_2 i ustalonego $0 < z < 1$ zachodzi

$$\forall_{\varepsilon>0}\exists_{m_0=m_0(z)}\forall_{m\geqslant m_0} \frac{M_1(z)[M_2(z)]^3\big(zQ_Y(1) - Q_X(0)\big)^3}{12m^2} < \varepsilon.$$

Ponadto zauważmy, że

$$\begin{aligned}
\hat{F}_{Z_3}(z) - F_{Z_{3m}}(z) &= \hat{F}_X(zQ_Y^*(1)) - F_X(zQ_Y(1)) \\
&+\frac{1}{2}\sum_{i=0}^{m-1}\Bigg\{\Bigg(F_X\Bigg(\frac{(i+1)zQ_Y(1)+(m-i-1)Q_X(0)}{m}\Bigg) \\
&-F_X\Bigg(\frac{izQ_Y(1)+(m-i)Q_X(0)}{m}\Bigg)\Bigg)\cdot\Bigg(F_Y\Bigg(\frac{(i+1)zQ_Y(1)+(m-i-1)Q_X(0)}{zm}\Bigg) \\
&+F_Y\Bigg(\frac{izQ_Y(1)+(m-i)Q_X(0)}{zm}\Bigg)\Bigg) - \Bigg(\hat{F}_X\Bigg(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{m}\Bigg) \\
&-\hat{F}_X\Bigg(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{m}\Bigg)\Bigg)\cdot\Bigg(\hat{F}_Y\Bigg(\frac{(i+1)zQ_Y^*(1)+(m-i-1)Q_X^*(0)}{zm}\Bigg) \\
&+\hat{F}_Y\Bigg(\frac{izQ_Y^*(1)+(m-i)Q_X^*(0)}{zm}\Bigg)\Bigg)\Bigg\}.
\end{aligned}$$

Stosując analogiczne narzędzia jak w przypadku 1. tzn. wykorzystując ciągłość dystrybuant F_X i F_Y, zbieżność dystrybuanty empirycznej do rzeczywistej oraz silną zgodność estymatora kwantyla otrzymujemy, że

$$|\hat{F}_{Z_3}(z) - F_{Z_{3m}}(z)| \to 0, \quad \text{gdy} \quad n_1, n_2 \to \infty.$$

3. Niech $z \geqslant 1$. Przyjmujemy $\varepsilon = \varepsilon(n_1) = 1 - F_X(Q_X^*(1))$ oraz $\delta = \delta(n_1) = 1 - F_X(Q_X^*(1))$. Stąd

$$Q_X^{\delta}(1) = Q_X^*(1). \tag{4.72}$$

Ponownie łatwo zauważyć, że $\varepsilon \to 0$ i $\delta \to 0$ gdy $n_1 \to \infty$.

Następnie, dla ustalonego n_1 i ustalonego $z \geqslant 1$ mamy

$$\lim_{m\to\infty} \frac{M_1(z)[M_2(z)]^3\left(Q_X^{\delta}(1) - zQ_Y(0)\right)^3}{12m^2} = 0$$

$$\iff \forall_{\varepsilon>0}\exists_{m_0=m_0(z,n_2)}\forall_{m\geqslant m_0} \frac{M_1(z)[M_2(z)]^3\left(Q_X^{\delta}(1) - zQ_Y(0)\right)^3}{12m^2} < \varepsilon.$$

Stąd na podstawie Twierdzenia 4.8(ii) dostajemy

$$\forall_{n_2}\forall_{z\geqslant 0}\exists_{\varepsilon}\exists_{m_0}\forall_{m\geqslant m_0}|F_{Z_{3m}}(z) - F_{Z_3}(z)| < 3\varepsilon$$

i w efekcie

$$|F_{Z_{3m}}(z) - F_{Z_3}(z)| \to 0, \quad \text{gdy} \quad n_1 \to \infty \quad \text{dla} \quad z \geqslant 1. \tag{4.73}$$

Ponadto, dla $z \geqslant 1$ uzyskujemy

$$\hat{F}_{Z_3}(z) - F_{Z_{3m}}(z) = \hat{F}_X(Q_X^*(1)) - F_X(Q_X^{\delta}(1))$$

$$+\frac{1}{2}\sum_{i=0}^{m-1}\left\{\left(F_X\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{m}\right)\right.\right.$$

$$-F_X\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{m}\right)\Bigg)\cdot\left(F_Y\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{zm}\right)\right.$$

$$+F_Y\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{zm}\right)\Bigg) - \left(\hat{F}_X\left(\frac{zQ_Y^*(0)(m-i-1)+(i+1)Q_X^*(1)}{m}\right)\right.$$

$$-\hat{F}_X\left(\frac{zQ_Y^*(0)(m-i)+iQ_X^*(1)}{m}\right)\Bigg)\cdot\left(\hat{F}_Y\left(\frac{zQ_Y^*(0)(m-i-1)+(i+1)Q_X^*(1)}{zm}\right)\right.$$

$$\left.\left.+\hat{F}_Y\left(\frac{zQ_Y^*(0)(m-i)+iQ_X^*(1)}{zm}\right)\Bigg)\right\} = \hat{F}_X(Q_X^*(1)) - F_X(Q_X^{\delta}(1))$$

$$+\frac{1}{2}\sum_{i=0}^{m-1}\left\{\left[F_X\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{m}\right)F_Y\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{zm}\right)\right.\right.$$

$$\left.-\hat{F}_X\left(\frac{zQ_Y^*(0)(m-i-1)+(i+1)Q_X^*(1)}{m}\right)\hat{F}_Y\left(\frac{zQ_Y^*(0)(m-i)+iQ_X^*(1)}{zm}\right)\right]$$

$$+\left[\hat{F}_X\left(\frac{zQ_Y^*(0)(m-i)+iQ_X^*(1)}{m}\right)\hat{F}_Y\left(\frac{zQ_Y^*(0)(m-i)+iQ_X^*(1)}{zm}\right)\right.$$

$$\left.\left.-F_X\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{m}\right)F_Y\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{zm}\right)\right]\right.$$

$$+\left[F_X\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{m}\right)F_Y\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{zm}\right)\right.$$
$$\left.-\hat{F}_X\left(\frac{zQ_Y^*(0)(m-i-1)+(i+1)Q_X^*(1)}{m}\right)\hat{F}_Y\left(\frac{zQ_Y^*(0)(m-i-1)+(i+1)Q_X^*(1)}{zm}\right)\right]$$
$$+\left[\hat{F}_X\left(\frac{zQ_Y^*(0)(m-i)+iQ_X^*(1)}{m}\right)\hat{F}_Y\left(\frac{zQ_Y^*(0)(m-i-1)+(i+1)Q_X^*(1)}{zm}\right)\right.$$
$$\left.\left.-F_X\left(\frac{zQ_Y(0)(m-i)+iQ_X^{\delta}(1)}{m}\right)F_Y\left(\frac{zQ_Y(0)(m-i-1)+(i+1)Q_X^{\delta}(1)}{zm}\right)\right]\right\}$$
$$=\hat{F}_X(Q_X^*(1))-F_X(Q_X^{\delta}(1))+\frac{1}{2}\sum_{i=0}^{m-1}\{E_i^{(2)}(z)+F_i^{(2)}(z)+G_i^{(2)}(z)+H_i^{(2)}(z)\}.$$

Na podstawie (4.72) oraz faktu, że $|\hat{F}_X(t)-F_X(t)|\to 0$ *p.p.*, gdy $n_1\to\infty$, otrzymujemy

$$|\hat{F}_X(Q_X^*(1))-F_X(Q_X^{\delta}(1))|\to 0 \quad p.p., \quad \text{gdy} \quad n_1\to\infty.$$

Następnie, wykorzystując (4.72), założenie o ciągłości F_X oraz F_Y, fakty, że $|\hat{F}_X(t)-F_X(t)|\to 0$ *p.p.*, gdy $n_1\to\infty$ i $|\hat{F}_Y(t)-F_Y(t)|\to 0$ *p.p.*, gdy $n_2\to\infty$ oraz stosując analogiczne rozważania jak w przypadku $0<z<1$ dostajemy

$$|E_i^{(2)}(z)|\to 0, \qquad |F_i^{(2)}(z)|\to 0,$$
$$|G_i^{(2)}(z)|\to 0, \qquad |H_i^{(2)}(z)|\to 0 \quad p.p., \quad \text{gdy} \quad n_1,n_2\to\infty.$$

Zatem dla $z\geqslant 1$

$$|\hat{F}_{Z_3}(z)-F_{Z_{3m}}(z)|\to 0 \quad p.p., \quad \text{gdy} \quad n_1,n_2\to\infty. \tag{4.74}$$

Stąd, na podstawie (4.73) oraz (4.74), otrzymujemy

$$|\hat{F}_{Z_3}(z)-F_{Z_3}(z)|\to 0, \quad \text{gdy} \quad n_1,n_2\to\infty.$$

□

Uwaga 4.6. Dla $z\geqslant 1$ wykorzystując Lemat 1 z pracy Midhu i in. [32] oraz fakt, że $\hat{F}_X(Q_X^*(1))=1$ łatwo pokazać, że zmienna losowa $2n_1n_2\widehat{F}_{Z_3}(z)$ ma funkcję masy prawdopodobieństwa daną wzorem

$$P(2n_1n_2\widehat{F}_{Z_3}(z)=a)=P(2n_1n_2-(C_1+2C_2)=a)=P(C_1+2C_2=2n_1n_2-a)$$
$$=P(C_1+2C_2=k)=\begin{cases}\displaystyle\sum_{j=0}^{\frac{1}{4}(2k+(-1)^k-1)}\frac{(n_1n_2)!}{j!(k-2j)!(n_1n_2-k+j)!}q_1^jq_2^{k-2j}q_3^{n_1n_2-k+j}, & \text{if } k\leqslant n_1n_2,\\ \displaystyle\sum_{j=k-n_1n_2}^{\frac{1}{4}(2k+(-1)^k-1)}\frac{(n_1n_2)!}{j!(k-2j)!(n_1n_2-k+j)!}q_1^jq_2^{k-2j}q_3^{n_1n_2-k+j}, & \text{if } k>n_1n_2,\end{cases}$$

gdzie $k = 2n_1n_2 - a$ oraz $a = 0, 1, ..., 2n_1n_2$, gdzie

$$q_1 = \sum_{i=0}^{m-1} \left[\left(F_X\left(\frac{zQ_Y^*(0)(m-i-1)+(i+1)Q_X^*(1)}{m}\right) - F_X\left(\frac{zQ_Y^*(0)(m-i)+iQ_X^*(1)}{m}\right)\right) \cdot F_Y\left(\frac{zQ_Y^*(0)(m-i)+iQ_X^*(1)}{zm}\right)\right],$$

$$q_2 = \sum_{i=0}^{m-1} \left[\left(F_X\left(\frac{zQ_Y^*(0)(m-i-1)+(i+1)Q_X^*(1)}{m}\right) - F_X\left(\frac{zQ_Y^*(0)(m-i)+iQ_X^*(1)}{m}\right)\right) \cdot \left(F_Y\left(\frac{zQ_Y^*(0)(m-i-1)+(i+1)Q_X^*(1)}{zm}\right) - F_Y\left(\frac{zQ_Y^*(0)(m-i)+iQ_X^*(1)}{zm}\right)\right)\right]$$

i

$$q_3 = 1 - q_1 - q_2.$$

4.5.4. Symulacje

Analogicznie jak w przypadku różnicy i iloczynu zmiennych losowych w tej części przedstawione zostały symulacje mające na celu zbadanie błędu aproksymacji za pomocą wzoru $F_{Z_{3m}}$ oraz ocenę estymatora $\hat{F}_{Z_3}(z)$. W symulacjach ponownie wykorzystano *R* studio.

Błąd aproksymacji

Błąd przybliżenia został obliczony dla trzech różnych przypadków:

1. X, Y – przesunięty rozkład wykładniczy z parametrami (1;0,5),
2. X, Y – przesunięty rozkład Weibulla z parametrami (2;1;0,5),
3. X, Y – rozkład Pareto z parametrami (1,3).

Wyniki symulacji zostały przedstawione w Tabeli 4.16.

Tabela 4.16: Błąd aproksymacji $F_{Z_{3m}}$.

$X,Y \sim \exp(1;0,5)$				$X,Y \sim \text{Weibull}(2;1;0,5)$				$X,Y \sim \text{Pareto}(1,3)$			
F_{Z_3}	$m=5$	$m=10$	$m=15$	F_{Z_3}	$m=5$	$m=10$	$m=15$	F_{Z_3}	$m=5$	$m=10$	$m=15$
0.2	0.0046	0.0008	0.0001	0.2	0.0125	0.0031	0.0013	0.2	0.00007	0.00007	0.00007
0.4	0.0042	0.0010	0.0003	0.4	0.0072	0.0018	0.0008	0.4	0.00004	0.00004	0.00004
0.6	0.0048	0.0014	0.0007	0.6	0.0080	0.0022	0.0010	0.6	0.00004	0.00004	0.00004
0.8	0.0061	0.0023	0.0016	0.8	0.0106	0.0014	0.0003	0.8	0.00008	0.00008	0.00008

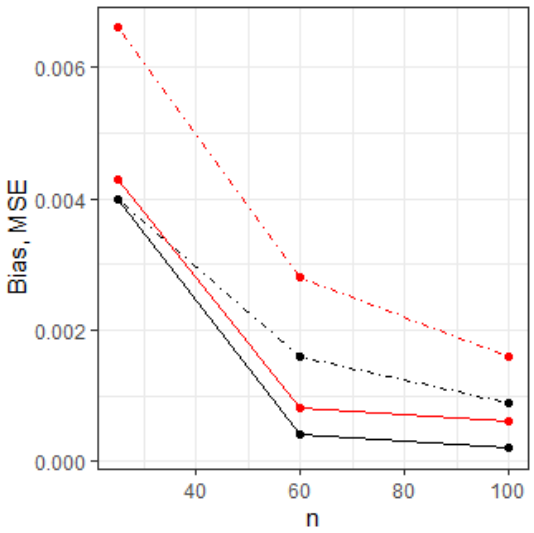
Wyniki symulacji wyraźnie pokazują, że w przypadku rozkładu wykładniczego i Weibulla wraz ze wzrostem liczby podziałów m błąd przybliżenia maleje. Ciekawą obserwacją, w przypadku rozkładu Pareto, jest stały błąd aproksymacji dla ustalonej wartość F_{Z_3}, który nie maleje wraz ze wzrostem liczby podziałów m. Takie rezultaty wynikają z liniowej postaci funkcji $u_2 = F_Y\left(\frac{Q_X(u_1)}{z}\right)$ dla tego rozkładu i zastosowania metody trapezów.

Ocena estymatora nieparametrycznego ($n_1 = n_2 = n$)

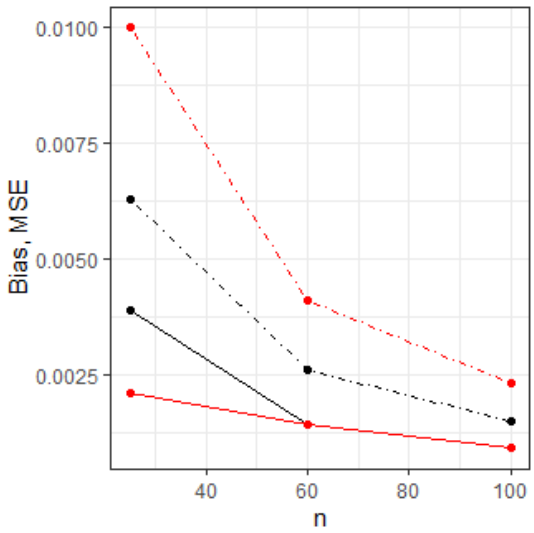
Analogicznie jak w przypadku różnicy i iloczynu zmiennych losowych następnym krokiem jest ocena jakości estymatora zdefiniowanego wzorami (4.65) oraz (4.66). Rozważane były próbki o rozmiarach $n \in \{25, 60, 100\}$ z rozkładu wykładniczego, Weibulla oraz Pareto. W każdym przypadku obliczono wartości z z zależności $F_{Z_3}(z) = u$, gdzie $u \in \{0.2, 0.4, 0.6, 0.8\}$, a $Z_3 = \frac{X}{Y}$. Dla każdej próbki obliczono wartość estymatora zdefiniowanego wzorami (4.65) lub (4.66) dla wyznaczonych z oraz $m \in \{5, 10, 15\}$. Następnie wyznaczono obciążenie oraz MSE otrzymanych estymacji. Dodatkowo, w każdym przypadku wygenerowano obciążenie i MSE klasycznej empirycznej dystrybuanty (ECDF) dla $\frac{X_i}{Y_i}$, gdzie $i = 1, 2, ..., n$, aby porównać wyniki z obciążeniem i MSE dla estymatora zdefiniowanego wzorami (4.65) i (4.66). W Tabelach 4.17-4.19 przedstawiono wyniki obciążenia i MSE dla wyżej odpisanych próbek. Ponadto, część wyników zebrano na Rysunkach 4.29 - 4.40. Na wspomnianych rysunkach przedstawiono wykresy obciążenia i MSE estymatora $\hat{F}_{Z_3}(z)$ dla $m = 15$ oraz dla ECDF. Ciągłe czarne i czerwone linie prezentują obciążenie dla, odpowiednio, estymatora $\hat{F}_{Z_3}(z)$ oraz ECDF. Linia przerywana czarna i czerwona reprezentuje MSE, odpowiednio, estymatora $\hat{F}_{Z_3}(z)$ oraz ECDF.

Tabela 4.17: Obciążenie i MSE estymatorów dla $X, Y \sim \exp(1; 0.5)$.

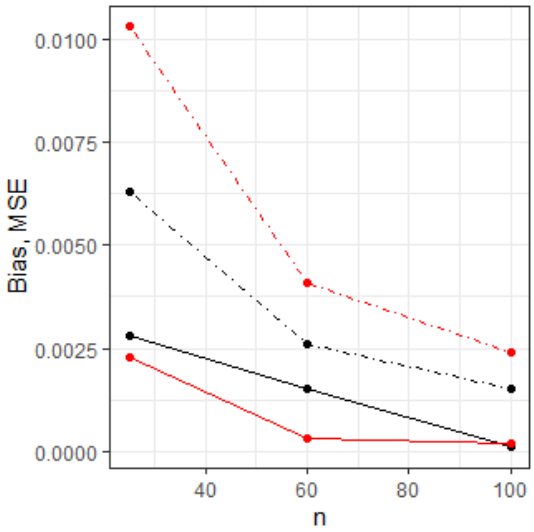
$n = 25$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0052	0.0039	0.0049	0.0040	0.0040	0.0040	0.0043	0.0066
0.4	0.0059	0.0058	0.0049	0.0063	0.0039	0.0063	0.0021	0.0100
0.6	0.0048	0.0059	0.0029	0.0063	0.0028	0.0063	0.0023	0.0103
0.8	0.0075	0.0031	0.0058	0.0032	0.0062	0.0032	0.0051	0.0058
$n = 60$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0046	0.0016	0.0008	0.0016	0.0004	0.0016	0.0008	0.0028
0.4	0.0029	0.0023	0.0015	0.0025	0.0014	0.0026	0.0014	0.0041
0.6	0.0047	0.0023	0.0020	0.0025	0.0015	0.0026	0.0003	0.0041
0.8	0.0064	0.0015	0.0038	0.0015	0.0033	0.0015	0.0025	0.0027
$n = 100$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0045	0.0009	0.0003	0.0009	0.0002	0.0009	0.0006	0.0016
0.4	0.0021	0.0013	0.0010	0.0015	0.0009	0.0015	0.0009	0.0023
0.6	0.0037	0.0013	0.0005	0.0015	0.0001	0.0015	0.0002	0.0024
0.8	0.0047	0.0009	0.0004	0.0009	0.0002	0.0009	0.0019	0.0016



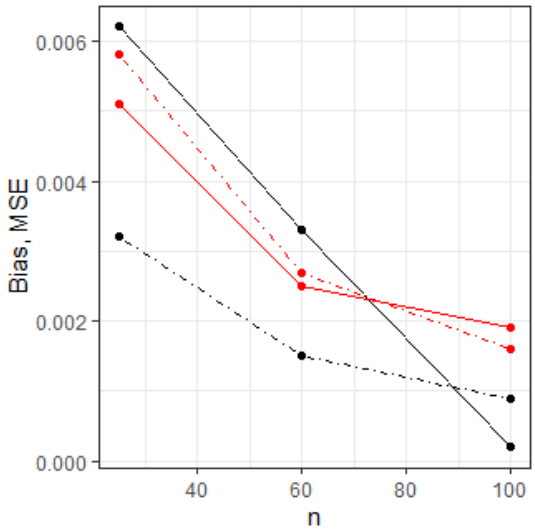
Rysunek 4.29: Obciążenie i MSE dla $F_{Z_3}(z) = 0.2$, gdy $X, Y \sim \exp(1; 0.5)$



Rysunek 4.30: Obciążenie i MSE dla $F_{Z_3}(z) = 0.4$, gdy $X, Y \sim \exp(1; 0.5)$



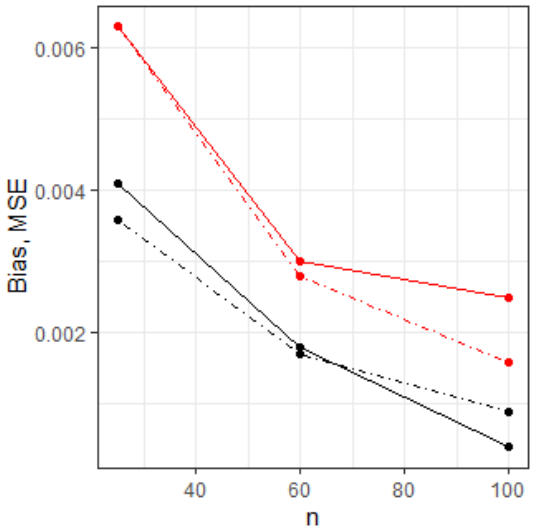
Rysunek 4.31: Obciążenie i MSE dla $F_{Z_3}(z) = 0.6$, gdy $X, Y \sim \exp(1; 0.5)$



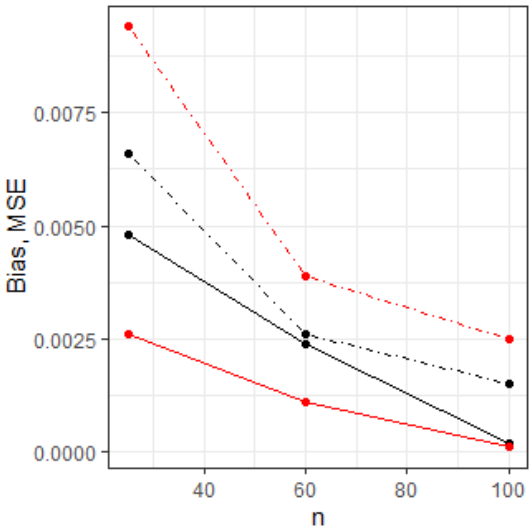
Rysunek 4.32: Obciążenie i MSE dla $F_{Z_3}(z) = 0.8$, gdy $X, Y \sim \exp(1; 0.5)$

Tabela 4.18: Obciążenie i MSE estymatorów dla $X, Y \sim \text{Weibull}(2; 1; 0.5)$.

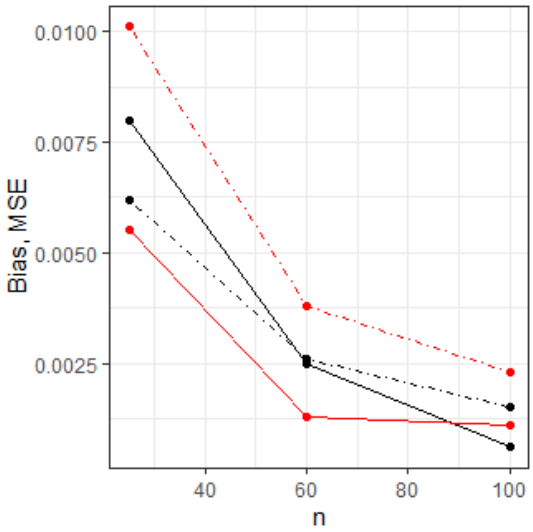
$n = 25$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0040	0.0036	0.0045	0.0036	0.0041	0.0036	0.0063	0.0063
0.4	0.0054	0.0065	0.0048	0.0066	0.0048	0.0066	0.0026	0.0094
0.6	0.0075	0.0059	0.0073	0.0061	0.0080	0.0062	0.0055	0.0101
0.8	0.0121	0.0034	0.0106	0.0033	0.0119	0.0033	0.0042	0.0060
$n = 60$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0059	0.0017	0.0024	0.0017	0.0018	0.0017	0.0030	0.0028
0.4	0.0048	0.0026	0.0027	0.0026	0.0024	0.0026	0.0011	0.0039
0.6	0.0055	0.0025	0.0028	0.0026	0.0025	0.0026	0.0013	0.0038
0.8	0.0084	0.0015	0.0045	0.0015	0.0041	0.0015	0.0015	0.0025
$n = 100$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0061	0.0010	0.0009	0.0010	0.0004	0.0009	0.0025	0.0016
0.4	0.0035	0.0015	0.0005	0.0015	0.0002	0.0015	0.0001	0.0025
0.6	0.0040	0.0015	0.0011	0.0015	0.0006	0.0015	0.0011	0.0023
0.8	0.0078	0.0010	0.0026	0.0009	0.0020	0.0009	0.0011	0.0015



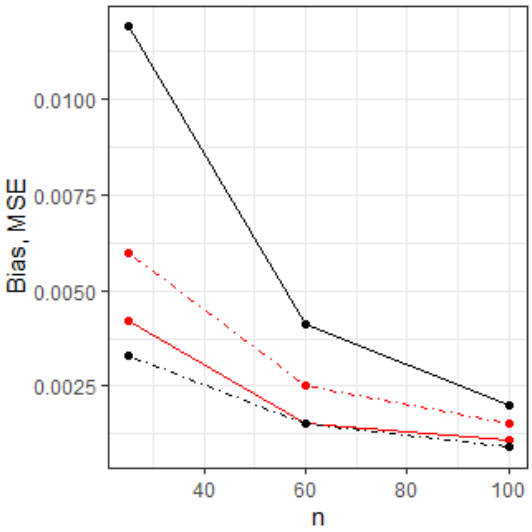
Rysunek 4.33: Obciążenie i MSE dla $F_{Z_3}(z) = 0.2$, gdy $X, Y \sim \text{Weibull}(2;1;0.5)$



Rysunek 4.34: Obciążenie i MSE dla $F_{Z_3}(z) = 0.4$, gdy $X, Y \sim \text{Weibull}(2;1;0.5)$



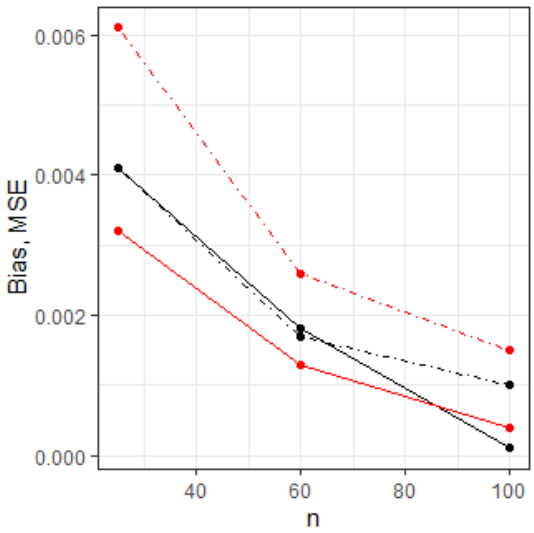
Rysunek 4.35: Obciążenie i MSE dla $F_{Z_3}(z) = 0.6$, gdy $X, Y \sim \text{Weibull}(2;1;0.5)$



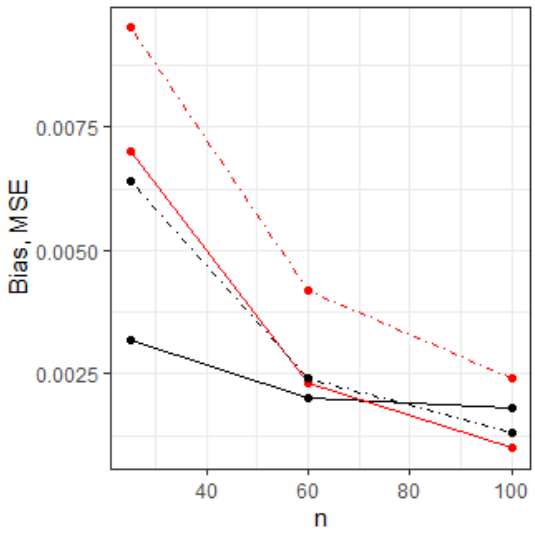
Rysunek 4.36: Obciążenie i MSE dla $F_{Z_3}(z) = 0.8$, gdy $X, Y \sim \text{Weibull}(2;1;0.5)$

Tabela 4.19: Obciążenie i MSE estymatorów dla $X,Y \sim \text{Pareto}(1;3)$.

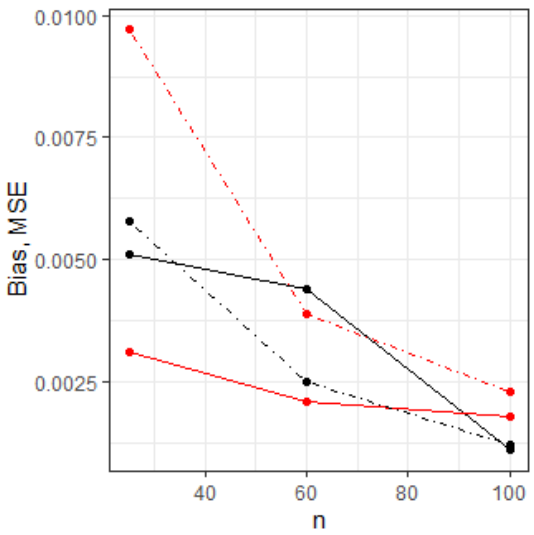
$n = 25$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0071	0.0038	0.0048	0.0041	0.0041	0.0041	0.0032	0.0061
0.4	0.0087	0.0053	0.0053	0.0064	0.0032	0.0064	0.0070	0.0095
0.6	0.0087	0.0050	0.0060	0.0056	0.0051	0.0058	0.0031	0.0097
0.8	0.0017	0.0036	0.0006	0.0038	0.0005	0.0039	0.0029	0.0061
$n = 60$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0026	0.0015	0.0024	0.0017	0.0018	0.0017	0.0013	0.0026
0.4	0.0038	0.0018	0.0026	0.0022	0.0020	0.0024	0.0023	0.0042
0.6	0.0051	0.0019	0.0048	0.0023	0.0044	0.0025	0.0021	0.0039
0.8	0.0007	0.0015	0.0006	0.0016	0.0001	0.0016	0.0011	0.0027
$n = 100$								
	$m = 5$		$m = 10$		$m = 15$		ECDF	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0004	0.0009	0.0002	0.0009	0.0001	0.0010	0.0004	0.0015
0.4	0.0023	0.0009	0.0019	0.0012	0.0018	0.0013	0.0010	0.0024
0.6	0.0022	0.0009	0.0014	0.0011	0.0011	0.0012	0.0018	0.0023
0.8	0.0003	0.0009	0.0001	0.0010	0.0001	0.0010	0.0009	0.0015



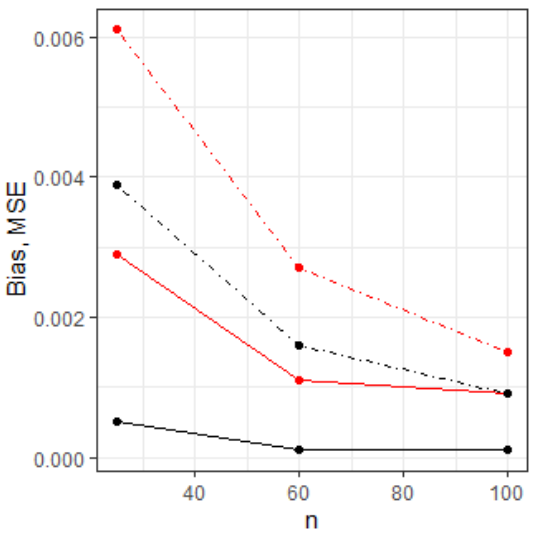
Rysunek 4.37: Obciążenie i MSE dla $F_{Z_3}(z) = 0.2$. gdy $X,Y \sim \text{Pareto}(1;3)$



Rysunek 4.38: Obciążenie i MSE dla $F_{Z_3}(z) = 0.4$, gdy $X,Y \sim \text{Pareto}(1;3)$



Rysunek 4.39: Obciążenie i MSE dla $F_{Z_3}(z) = 0.6$, gdy $X, Y \sim \text{Pareto}(1;3)$



Rysunek 4.40: Obciążenie i MSE dla $F_{Z_3}(z) = 0.8$, gdy $X, Y \sim \text{Pareto}(1;3)$

Podobnie jak w przypadku różnicy i iloczynu zmiennych losowych, można zauważyć podobną tendencję dla każdego z badanych rozkładów. Obserwujemy, że wraz ze wzrostem liczności próby MSE estymatora (4.65) lub (4.66) maleje i, co istotne, w każdym przypadku MSE estymatora (4.65) lub (4.66) jest wyraźnie mniejsze niż MSE wyznaczone dla ECDF. Ponadto obciążenie estymatora (4.65) lub (4.66) maleje wraz ze wzrostem liczności próby, a w wielu przypadkach obciążenie estymatora (4.65) lub (4.66) jest mniejsze niż obciążenie ECDF.

Ocena estymatora nieparametrycznego ($n_1 \neq n_2$)

Badania symulacyjne dla estymatora danego wzorem (4.65) lub (4.66) dla przypadku różnych rozmiarów próby przeprowadzono według tej samej procedury, co dla przypadku różnicy i iloczynu dwóch zmiennych losowych. Wykorzystano tę samą liczbę powtórzeń, rozmiary prób oraz wartości parametru m. Ponownie rozważano próby z rozkładu wykładniczego, Weibulla oraz Pareto. Następnie wyznaczono obciążenie (bias) oraz błąd średniokwadratowy (MSE) estymatora danego wzorem (4.65) lub (4.66). Dla porównania wyznaczono obciążenie i MSE dla estymatora opartego na wszystkich ilorazach par obserwacji (ozn. jako *wszystkie pary*). Otrzymane wyniki zestawiono w Tabelach 4.20 oraz 4.21.

Tabela 4.20: Obciążenie i MSE dla $X, Y \sim \exp(1; 0.5)$.

$n_1 = 30$, $n_2 = 25$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0060	0.0034	0.0057	0.0034	0.0054	0.0035	0.0040	0.0035
0.4	0.0092	0.0053	0.0089	0.0058	0.0082	0.0058	0.0060	0.0060
0.6	0.0050	0.0049	0.0042	0.0053	0.0025	0.0054	0.0025	0.0056
0.8	0.0091	0.0030	0.0088	0.0030	0.0086	0.0030	0.0031	0.0032
$n_1 = 60$, $n_2 = 65$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0058	0.0015	0.0014	0.0015	0.0012	0.0015	0.0017	0.0015
0.4	0.0066	0.0022	0.0051	0.0025	0.0045	0.0025	0.0037	0.0026
0.6	0.0041	0.0022	0.0009	0.0023	0.0007	0.0023	0.0007	0.0024
0.8	0.0069	0.0014	0.0032	0.0014	0.0031	0.0015	0.0006	0.0015
$n_1 = 105$, $n_2 = 100$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0054	0.0009	0.0007	0.0009	0.0001	0.0009	0.0001	0.0009
0.4	0.0042	0.0013	0.0008	0.0016	0.0002	0.0016	0.0009	0.0016
0.6	0.0038	0.0012	0.0002	0.0014	0.0001	0.0014	0.0005	0.0015
0.8	0.0062	0.0009	0.0015	0.0008	0.0007	0.0008	0.0004	0.0008

Tabela 4.21: Obciążenie i MSE dla $X, Y \sim \text{Weibull}(2;1;0.5)$.

$n_1 = 30$, $n_2 = 25$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0080	0.0035	0.0044	0.0035	0.0042	0.0035	0.0037	0.0035
0.4	0.0055	0.0062	0.0050	0.0065	0.0049	0.0064	0.0043	0.0064
0.6	0.0062	0.0056	0.0050	0.0056	0.0046	0.0056	0.0021	0.0057
0.8	0.0133	0.0037	0.0120	0.0035	0.0119	0.0035	0.0024	0.0036
$n_1 = 60$, $n_2 = 65$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0078	0.0017	0.0039	0.0016	0.0036	0.0016	0.0034	0.0016
0.4	0.0038	0.0023	0.0016	0.0025	0.0015	0.0024	0.0015	0.0025
0.6	0.0051	0.0024	0.0037	0.0025	0.0036	0.0025	0.0021	0.0025
0.8	0.0069	0.0015	0.0030	0.0014	0.0028	0.0014	0.0017	0.0015
$n_1 = 105$, $n_2 = 100$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0063	0.0010	0.0006	0.0009	0.0001	0.0009	0.0007	0.0009
0.4	0.0037	0.0016	0.0005	0.0016	0.0001	0.0016	0.0004	0.0017
0.6	0.0036	0.0014	0.0002	0.0015	0.0001	0.0015	0.0014	0.0015
0.8	0.0066	0.0010	0.0021	0.0009	0.0015	0.0009	0.0012	0.0009

Tabela 4.22: Obciążenie i MSE dla $X, Y \sim \text{Pareto}(1;3)$.

$n_1 = 30$, $n_2 = 25$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0061	0.0037	0.0054	0.0039	0.0051	0.0040	0.0035	0.0041
0.4	0.0090	0.0045	0.0082	0.0051	0.0061	0.0055	0.0016	0.0058
0.6	0.0158	0.0073	0.00145	0.0080	0.0078	0.0050	0.0067	0.0054
0.8	0.0159	0.0046	0.0147	0.0048	0.0043	0.0033	0.0033	0.0036
$n_1 = 60$, $n_2 = 65$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0010	0.0015	0.0005	0.0017	0.0002	0.0017	0.0010	0.0017
0.4	0.0045	0.0019	0.0035	0.0023	0.0027	0.0025	0.0010	0.0028
0.6	0.0078	0.0025	0.0068	0.0030	0.0017	0.0024	0.0015	0.0027
0.8	0.0077	0.0019	0.0073	0.0020	0.0029	0.0017	0.0029	0.0018
$n_1 = 105$, $n_2 = 100$								
	$m = 5$		$m = 10$		$m = 15$		wszystkie pary	
F_{Z_3}	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.0004	0.0009	0.0002	0.0010	0.0001	0.0010	0.0007	0.0011
0.4	0.0037	0.0010	0.0025	0.0013	0.0011	0.0015	0.0003	0.0017
0.6	0.0039	0.0012	0.0030	0.0015	0.0005	0.0014	0.0013	0.0016
0.8	0.0054	0.0010	0.0046	0.0011	0.0016	0.0010	0.0021	0.0011

Powyższe tabele pokazują, że dla dystrybuanty ilorazu niezależnych zmiennych losowych otrzymano analogiczne wyniki jak w przypadku różnicy i iloczynu zmiennych losowych. Wraz ze wzrostem wartości parametru m obciążenie proponowanego estymatora ulega wyraźnemu zmniejszeniu. Dla $m = 15$ wartości są porównywalne z estymatorem opartym na wszystkich parach obserwacji, a w niektórych przypadkach nawet od niego mniejsze. Jednocześnie wartości błędu średniokwadratowego pozostają na zbliżonym poziomie dla wszystkich analizowanych estymatorów i maleją wraz ze wzrostem liczebności prób. Otrzymane wyniki potwierdzają, że proponowany estymator zapewnia dokładność porównywalną z estymatorem porównawczym również w przypadku estymacji dystrybuanty ilorazu dwóch zmiennych losowych.

Podsumowanie

Kwantyle oraz estymatory kwantyli stanowią niezwykle istotne narzędzie w analizie danych, pozwalając na precyzyjne opisanie rozkładu zmiennej losowej nie tylko poprzez wartości średnie, lecz także w oparciu o jej całą strukturę. Dzięki nim możliwe jest badanie zróżnicowania danych, identyfikacja wartości skrajnych oraz lepsza analiza asymetrii rozkładu. Estymatory kwantyli odgrywają szczególną rolę w praktyce, ponieważ umożliwiają wiarygodne wnioskowanie na podstawie próby, nawet w sytuacjach, gdy rozkład populacji nie jest znany lub odbiega od klasycznych założeń. Z tego względu kwantyle i ich estymacja pozostają nie tylko ważnym elementem narzędzi matematycznych, lecz także praktycznym zagadnieniem wspierającym podejmowanie decyzji opartych na danych. W niniejszej pracy analizowano własności kwantyli oraz ich estymatorów ze szczególnym uwzględnieniem ich zachowania asymptotycznego, własności granicznych i wykorzystania do wyznaczania rozkładów.

Główne cele jakie realizowane były w niniejszej pracy, to uogólnienie znanej postaci reprezentacji Bahadura i wyznaczenie estymatora dystrybuanty wybranych funkcji zmiennych losowych z wykorzystaniem funkcji kwantylowych.

W ramach pierwszego celu uogólniono postać reprezentacji Bahadura na przypadek ośmiu różnych definicji estymatora kwantyla dla czterech różnych rodzajów struktur zależności w populacji, tj. rozważano zależność ujemnego stowarzyszenia, φ-mieszania, ρ-mieszania oraz ρ^*-mieszania. Ponadto w przypadku zmiennych losowych ρ-mieszających wykazano jednostajną asymptotyczną normalność estymatora kwantyla. Otrzymane wyniki teoretyczne dotyczące silnej zgodności i jednostajnej asymptotycznej normalności kwantyla z próby w przypadku zmiennych losowych ρ-mieszających zostały zilustrowane za pomocą symulacji komputerowych. Zaprezentowane wyniki zostały opublikowane w pracach Dudek i Kuczmaszewskiej [12], [13]. W pracy Dudek i Kuczmaszewskiej [11] uzyskano analogiczne wyniki dotyczące jednostajnej asymptotycznej normalności kwantyla z próby dla zmiennych losowych ρ^*-mieszających. Ponadto w rozprawie wyznaczono reprezentację dla estymatora kwantyla zdefiniowanego klasycznie wzorem $Q_n^4(p) = X_{\lfloor np \rfloor + 1}$ dla niezależnych zmiennych losowych, analogiczną do reprezentacji Bahadura w klasycznej przestrzeni probabilistycznej,

w terminach przestrzeni podliniowej wartości oczekiwanej, czyli tzw. przestrzeni z niepewnością prawdopodobieństwa. Należy podkreślić, że w całym dowodzie reprezentacji Bahadura w przestrzeni podliniowej wartości oczekiwanej niezbędne było wprowadzenie nowych narzędzi, które nie występowały w klasycznej przestrzeni probabilistycznej. Nowe narzędzia były konieczne m.in. ze względu na specyfikę zbioru $\mathscr{H}$ (odpowiednika zbioru zmiennych losowych) i brak możliwości (w ogólnym przypadku) wyznaczenia bezpośrednio podliniowej wartości oczekiwanej z funkcji indykatora I_A.

Realizując drugi cel pracy przedstawiono wzory aproksymujące dystrybuantę różnicy, iloczynu i ilorazu niezależnych zmiennych losowych dopełniając tym samym wyniki opublikowane w pracy Midhu i in. [32] (dotyczące sumy niezależnych, nieujemnych zmiennych losowych). Należy wyraźnie podkreślić, że wyniki dotyczące sumy zmiennych losowych nie mogą zostać przeniesione bezpośrednio na przypadek różnicy, iloczynu czy ilorazu zmiennych losowych. Konstrukcja wspomnianych wzorów aproksymujących opierała się na wykorzystaniu funkcji kwantylowych odpowiednich zmiennych losowych oraz na zastosowaniu numerycznej metody trapezów obliczania całek. Kluczowym problemem, który w niniejszej pracy został rozwiązany i który nie pojawiał się w pracy Midhu i in. [32] było wyznaczenie odpowiednika (we wzorze aproksymującym) kwantyla rzędu 1 dla rozkładów, których wartości są nieograniczone od góry, tzn. gdy $Q(1) = \infty$. Dla każdego wzoru aproksymującego został oszacowany błąd aproksymacji. Następnie na bazie wzorów aproksymujących został wyznaczony nieparametryczny estymator dystrybuanty różnicy, iloczynu i ilorazu zmiennych losowych. Przy konstruowaniu estymatorów dystrybuanty pojawiła się konieczność wyznaczenia odpowiednika kwantyla rzędu 0 oraz 1. W każdym przypadku została zbadana silna zgodność, a dla iloczynu zmiennych losowych został wyznaczony dokładny rząd zbieżności estymatora dystrybuanty do dystrybuanty rzeczywistej. Na koniec zostały przeprowadzone symulacje badające błąd aproksymacji, błąd średniokwadratowy i obciążenie estymatora. Wyniki dla nowych estymatorów porównano z wynikami dla klasycznej dystrybuanty empirycznej. W każdym przypadku stwierdzano mniejszy błąd średniokwadratowy dla nowych, zaproponowanych estymatorów. Wyniki przedstawione w niniejszej rozprawie dotyczące różnicy niezależnych zmiennych losowych zostały opublikowane w pracy Dudek i Kuczmaszewskiej [14].

Bibliografia

[1] Atkinson K. E., An Introduction to Numerical Analysis, John Wiley & Sons, 2008,

[2] Bahadur R. R., A note on quantiles in large samples, The Annals of Mathematical Statistics 37(3), 1966, str. 577-580,

[3] Beaulieu N. C., Abu-Dayya A. A., McLane P. J., Estimating the distribution of a sum of independent lognormal random variables, IEEE Transactions on Communicatins 43, 1995, str. 2869–2873,

[4] Buonocore A., Pirozzi E., Caputo L., A note on the sum of uniform random variables, Statistics & Probability Letters 79, 2009, str. 2092–2097,

[5] Caginalp C., Caginalp G., The quotient of normal random variables and application to asset price fat tails, Physica A: Statistical Mechanics and its Applications 499, 2018, str. 457-471,

[6] Chen P., Sung S. H., On complete convergence and complete moment convergence for weighted sums of ρ^*-mixing random variables, Journal of Inequalities and Applications 121, 2018,

[7] Chlebus M., Pomiar ryzyka rynkowego za pomocą miary Value at Risk-podejście dwuetapowe, Uniwersytet Warszawski, rozprawa doktorska, 2014,

[8] Dielman T., Lowry C., Pfaffenberger R., A comparison of quantile estimators. Communications in Statistics-Simulation and Computation 23(2), 1994, str. 355-371,

[9] Ding L., Li Y., The Berry-Esseen bounds of wavelet estimator for regression model whose errors form a linear process with a ρ-mixing, Journal of Inequalities and Applications 107, 2016, str. 1-12

[10] Douge L., The Bahadur representation of sample quantiles for associated sequences, Statistics Theory, 2022, https://doi.org/10.48550/arXiv.2206.05995

[11] Dudek D., Kuczmaszewska A., On the Bahadur Representation of Quantiles for a Sample from ρ^*-Mixing Structure Population, Advances in Science and Technology Research Journal 16(3), 2022, str. 316-330,

[12] Dudek D., Kuczmaszewska A., Some practical and theoretical issues related to the quantile estimators, Statistical Papers 65, 2024, str. 3917–3933,

[13] Dudek D., Kuczmaszewska A., The Bahadur representation for sample quantiles of ρ-mixing random variables, Statistics & Probability Letters 221, 2025,

[14] Dudek D., Kuczmaszewska A., A note on an approximation and estimation of distribution function of difference of random variables, Statistical Papers 66(151), 2025,

[15] Dufour J. M., Distribution and quantile functions, McGill University, 1995

[16] Fainleib A.S., A generalization of Esseen's inequality and its application in probabilistic number theory, Math USSR Izv. 2, 1968, str. 821-844,

[17] Galton F., Natural Inheritance, Macmillan, New York, 1889

[18] Gaunt R. E., Li S., Sutcliffe H. L., The Variance-Gamma Product Distribution, Results in Mathematics 80(208), 2025,

[19] Han P.S., Kong L.L., Zhao J.W., Zhou X.C., A general framework for quantile estimation with incomplete data, Journal of the Royal Statistical Society Series B 81(2), 2019, str. 305-333,

[20] Harrell F.E., Davis C.E., A new distribution-free quantile estimator, Biometrika 69(3), 1982, str. 635–640,

[21] Hu H.C., Bahadur representations of M-estimators and their applications in general linear models, Journal of Inequalities and Applications 2018(123), 2018,

[22] Joag-Dev K., Proschan F., Negative association of random variables with applications, The Annals of Statistics 11(1), 1983, str. 286–295.

[23] Joarder A. H., Omar M. H., On the Exact Distribution of the Difference Between Two Chi-Square Variables, Bulletin of the Malaysian Mathematical Sciences Society 45, 2022, str. 505-522,

[24] Koo J.Y., Lee Y., Kim Y., Park C., A Bahadur representation of the linear support vector machine, Journal of Machine Learning Research 9(44), 2008, str. 1343-1368,

[25] Kuczmaszewska A., Quantiles under the sub-linear expectations, Statistical Papers 66(82), 2025

[26] Li X., Yang W., Hu S., The Bahadur representation for sample quantile under NOD sequences, Journal of Nonparametric Statistics 23(1), 2011, str. 59-65,

[27] Ling N. X., The Bahadur representation for sample quantiles under negatively associated sequences, Statistics and Probability Letters 78(16), 2008, str. 2660-2663,

[28] Lo C.F., The Sum and Difference of Two Lognormal Random Variables, Journal of Applied Mathematics ID838397, 2012,

[29] Lomnicki Z. A., On the Distribution of Products of Random Variables, Journal of the Royal Statistical Society. Series B (Methodological) 29(3), 1967, str. 513-524,

[30] Ly S., Pho K. H., Ly S., Wong W. K., Determining Distribution for the Product of Random Variables by Using Copulas, Risks 7(1), 2019,

[31] Marinacci M., Limit laws for non-additive probabilities and their frequentist interpretation, Journal of Economic Theory 84(2), 1999, str. 145–195,

[32] Midhu N.N., Dewan I., Sudheesh K.K., Sreedevi E.P., On approximation and estimation of distribution function of sum of independent random variables, Statistical Papers 65(3), 2024, str. 1437-1467,

[33] Moschopoulos P.G., The distribution of the sum of independent gamma random variables, Annals of the Institute of Statistical Mathematics 37, 1985, str. 541–544,

[34] Nadarajah S., Jiang X., Chu J., A saddlepoint approximation to the distribution of the sum of independent non-identically beta random variables, Statistica Neerlandica 69, 2015, str. 102–114,

[35] Rozanov Y. A., Volkonski V. A., Some limit theorems for random function, Theory of Probability & Its Applications 4, 1959, str. 178-197,

[36] Peligrad M., On the Central Limit Theorem for ρ–Mixing Sequences od Random Variables, Annals of Probability 15(4), 1987, str. 1387-1394,

[37] Peng S., A new central limit theorem under sublinear expectations, 2008, https://arxiv.org/abs/0803.2656v1

[38] Peng S., Nonlinear Expectations and stochastic calculus under uncertainty: with robust CLT and G-Brownian Motion, 2019, https://doi.org/10.1007/978-3-662-59903-7

[39] Puri P. S., Rubin H., A Characterization Based on the Absolute Difference of Two I. I. D.Random Variables, The Annals of Mathematical Statistics 41, 1970, str. 2113-2122,

[40] Sen P. K., On the Bahadur representation of sample quantiles for sequences of φ–mixing random variables, Journal of Multivariate Analysis 2, 1972, str. 77-95,

[41] Sadooghi-Alvandi S. M., Nematollahi A. R., Habibi R., On the distribution of the sum of independent uniform random variables, Statistical Papers 50, 2009, str. 171–175,

[42] Shao Q. M., Maximal inequalities for partial sums of ρ-mixing sequences, The Annals of Probability 23(2), 1995, str. 948-965,

[43] Shao Q. M., A comparison theorem on moment inequalities between negatively associated and independent random variables, Journal of Theoretical Probability 13, 2000, str. 343-356,

[44] Sun S. X., The Bahadur representation for sample quantiles under weak dependence, Statistics and Probability Letters 76(12), 2006, str. 1238-1244,

[45] Takeuchi I., Le Q.V., Sears T.D., Smola A.J., Nonparametric quantile estimation, Journal of Machine Learning Research 7(45), 2006, str. 1231-1264,

[46] Utev S., Peligrad M., Maximal inequalities and an invariance principle for a class of weakly dependent random variables, Journal of Theoretical Probability 16, 2003, str. 101-115,

[47] Wang X. J., Hu S. H., Yang W. Z., The Bahadur representation for sample quantiles under strongly mixing sequence, Journal of Statistical Planning and Inference 141(2), 2011, str. 655-662,

[48] Wu Y., Yu W., Wang X., The Bahadur representation of sample quantiles for φ–mixing random variables and its application, Statistics 55(2), 2021, str. 426-444

[49] Xing G. D., Yang S. C., A remark on the Bahadur representation of sample quantiles for negatively associated sequences, Journal of the Korean Statistical Society 40(3), 2011, str. 277-280,

[50] Xing G., Yang S., On the Bahadur representation of sample quantiles for ψ-mixing sequences and its applications, Communications in Statistics–Theory and Methods 48, 2019, str. 1060-1072,

[51] Xu S., Ge L., Miao Y., On the Bahadur representation of sample quantiles and order statistics for NA sequences, Journal of the Korean Statistical Society 42(1), 2013, str. 1-7,

[52] Yang S.C., Uniformly asymptotic normality of the regression weighted estimator for negatively associated samples, Statistics & Probability Letters 62 (2), 2003, str. 101-110,

[53] Yang W., Hu S., Wang X., The Bahadur representation for sample quantiles under dependent sequence, Acta Mathematicae Applicatae Sinica, English Series 35, 2019, str. 521-531,

[54] Yoshihara K., The Bahadur representation of sample quantile for sequences of strongly mixing random variables, Statistic and Probability Letters 24(4), 1995, str. 299-304,

[55] Yuancheng S., Saralees N., Xiaodong S., On the distribution of quotient of random variables conditioned to the positive quadrant, Communications in Statistics - Theory and Methods 49(10), 2020, str. 2514-2528,

[56] Zelterman D., Smooth nonparametric estimation of the quantile function, Journal of Statistical Planning and Inference 26(3), 1990, str. 339–352,

[57] Zhang Q., Yang W., Hu S., On Bahadur representation for sample quantiles under α-mixing sequence, Statistical Papers 55, 2014, str. 285-299.

[58] Zhang L.X., Exponential inequalities under the sub-linear expectations with applications to laws of the iterated logarithm, Science China Mathematics 59, 2016, str. 2503–2526,

[59] Zhang L.X., Strong limit theorems for extended independent and extended negatively dependent random variables under non-linear expectations, 2016, https://arxiv.org/abs/1608.00710,

[60] Zhang L.X., Rosenthal's inequalities for independent and negatively dependent random variables under sub-linear expectations with application, Science China Mathematics 59, 2016, str. 751–768,

[61] Zhong H., Wu Q., Complete convergence and complete moment convergence for weighted sums of extended negatively dependent random variables under sub-linear expectation, Journal of Inequalities and Applications 2017(261), 2017.