

Adam Zagdański

Predykcja przedziałowa oraz
metody estymacji nieparametrycznej
dla procesów stochastycznych
wraz z zastosowaniami

Rozprawa doktorska przygotowana pod kierunkiem
dr hab. R. Róžańskiego, prof. nadzw. P.Wr.

Politechnika Wrocławska. Instytut Matematyki.

Wrocław 2003

Spis treści

Wstęp	3
1 Bootstrapowe przedziały predykcyjne z wykorzystaniem metody <i>sieve bootstrap</i>	10
1.1 Metoda <i>sieve bootstrap</i>	10
1.1.1 Założenia o modelu	10
1.1.2 Algorytm metody	11
1.1.3 Uwagi do założeń i algorytmu	13
1.1.4 Przegląd wyników	15
1.2 Predykcja przedziałowa	16
1.2.1 Optymalny predyktor liniowy	16
1.2.2 Przedziały predykcyjne	17
1.2.3 Gaussowskie przedziały predykcyjne	18
1.2.4 Bootstrapowe przedziały predykcyjne — przegląd wyników	18
1.2.5 Równoczesne przedziały predykcyjne	23
1.2.6 Przedziały predykcyjne z wykorzystaniem metody <i>sieve bootstrap</i>	24
1.3 Rezultaty dotyczące zgodności	26
1.3.1 Określenie zgodności metody bootstrap	26
1.3.2 Fakty pomocnicze	29
1.3.3 Zgodność metody <i>hybrid bootstrap</i>	42
1.3.4 Zgodność metody <i>bootstrap-t</i>	51
1.3.5 Zgodność jednostajna	60
1.3.6 Zgodność przedziałów predykcyjnych	64
1.4 Bootstrap wygładzony	70
1.4.1 Przegląd wyników	71
1.4.2 Zgodność wygładzonej metody bootstrap	73
1.4.3 Dobór szerokości okna h	74
1.4.4 Stabilizacja studentyzowanych (bootstrap-t) przedziałów ufności	75
1.5 Symulacje komputerowe	76

2	Zastosowanie metody sita do estymacji intensywności multiplikatywnych procesów punktowych	89
2.1	Metoda sita	89
2.2	Przegląd wyników	90
2.3	Estymacja intensywności w modelu multiplikatywnym	91
2.4	Zgodność i asymptotyczna normalność histogramowego estymatora sitowego.	92
2.5	Przypadek wielowymiarowy	96
2.6	Przykłady	97
2.7	Symulacje komputerowe	100
3	Zastosowanie metody sita do estymacji funkcji dryfu dla procesów Itô	105
3.1	Przegląd wyników	105
3.1.1	Estymacja dla procesów dyfuzyjnych	105
3.1.2	Modele semimartyngałowe	106
3.2	Estymacja funkcji dryfu	107
3.3	Zgodność i asymptotyczna normalność histogramowego estymatora sitowego.	108
3.4	Przykłady	112
3.5	Symulacje komputerowe	114
	Dodatek A: Wybrane zagadnienia teorii prognozy dla procesów stacjonarnych	125
	Literatura	127

Wstęp

Standardowe metody stosowane w przypadku estymacji skończenie wymiarowych parametrów, jak np. metoda największej wiarygodności czy metoda najmniejszych kwadratów, zazwyczaj zawodzą, kiedy próbuje się zastosować je bezpośrednio do zagadnień estymacji nieparametrycznej (nieskończenie wymiarowej).

Problem ten można dobrze zilustrować na przykładzie estymacji funkcji gęstości rozkładu. Załóżmy, że $(x_1, \dots, x_n)$ jest realizacją zmiennych losowych i.i.d. z rozkładu absolutnie ciągłego o nieznannej gęstości $f_0(x)$. Wówczas estymator największej wiarygodności funkcji f_0 maksymalizuje kryterium:

$$L(f) = \prod_{k=1}^n f(x_k). \quad (1)$$

Łatwo zauważyć jednak, że w przypadku, gdy nic więcej nie wiadomo o f_0 , maksimum (1) nie jest osiągalne na przestrzeni wszystkich gęstości.

Innym przykładem, dla którego zastosowanie klasycznych metod estymacji kończy się niepowodzeniem jest nieparametryczna estymacja funkcji regresji $m_0(x) = E(Y|X = x)$ na bazie obserwacji $(x_1, y_1), \dots, (x_n, y_n)$. Wówczas estymator m_0 wyznaczony metodą najmniejszych kwadratów minimalizuje

$$R(m) = \sum_{k=1}^n (y_k - m(x_k))^2. \quad (2)$$

W tym przypadku, minimum równe zero jest osiągalne dla dowolnej funkcji $\hat{m}$, takiej że: $\hat{m}(x_k) = y_k, k = 1, \dots, n$.

U. Grenander (1981) zaproponował następujące rozwiązanie przedstawionych problemów: optymalizujemy określone kryterium (np. maksymalizujemy funkcję wiarygodności lub minimalizujemy sumę kwadratów błędów) na pewnym skończenie wymiarowym podzbiorze przestrzeni parametrów $S(m(n))$ pozwalając, aby wymiar tego podzbioru rósł wraz ze wzrostem rozmiaru próby n . Ciąg podzbiorów $\{S(m(n))\}$, na bazie których konstruowany jest estymator Grenander nazwał „sitem”, samą procedurę estymacji określając natomiast jako „metodę sita” (*method of sieves*). Jak się okazało, metoda ta prowadzi, nawet w bardzo ogólnych przypadkach, do wyznaczenia zgodnych, nieparametrycznych estymatorów (np. Grenander (1981), Geman i Hwang (1982)). Poniżej przedstawimy nieco bardziej formalnie ideę metody sita.

Definicja 1 (Sito)

Rodzinę $\{S(m_n)\}$ podzbiorów przestrzeni parametrów A nazywamy *sitem* jeżeli:

- (i) Wymiar $m(n)$ rośnie wraz z n ,
- (ii) $\mathcal{S} = \bigcup_{m_n} S(m_n)$ jest zbiorem gęstym w A , tzn.

$$\overline{\bigcup_{m_n} S(m_n)} = A.$$

Uwaga 1

Często zamiast (i) zakłada się dodatkowo, że $\{S(m_n)\}$ jest rodziną monotoniczną, tzn.

$$S(m_n) \subseteq S(m_{n+1}).$$

Uwaga 2

Niekiedy jako parametr indeksujący sito przyjmuje się $\mu = \frac{1}{m_n}$, ($\mu > 0$) (ang. mesh size).

Estymator największej wiarygodności $\hat{\alpha}_{m_n}$ dla określonego podzbioru $S_{m_n} \subset A$ wyznaczamy tak, aby:

$$L_n(\hat{\alpha}_{m_n}) = \sup_{\alpha \in S_{m_n}} L_n(\alpha).$$

Jednym z najprostszych przykładów sita jest tzw. sito histogramowe, które dla problemu estymacji gęstości określamy jako:

$$S_{m_n} = \left\{ f : f \text{-gęstość rozkładu, przedziałami stała na } \left[\frac{k-1}{m_n}, \frac{k}{m_n} \right), k \in Z \right\}.$$

Dla przykładu wymienimy jeszcze inne stosowane w tym przypadku (estymacja gęstości) sito:

$$S_{m_n} = \{ f : f \text{-gęstość rozkładu, } |f'| \leq f m_n \text{ p.w.} \}.$$

Z kolei dla problemu nieparametrycznej estymacji funkcji regresji najczęściej używane są sita postaci:

$$S_{m_n} = \left\{ \alpha : \alpha \in \Psi, \int \left| \frac{d^p}{dx^p} \alpha(x) \right|^2 dx \leq m_n \right\},$$

gdzie Ψ oznacza zbiór funkcji mających absolutnie ciągłe pochodne rzędu $p - 1$.

Obszerny przegląd wykorzystywanych sit można znaleźć np. u Grenandera (1981) oraz w pracy Gemana i Hwanga (1982).

Zakładając, że mamy już estymator $\hat{\alpha}_{m_n}$ interesowała nas będzie analiza jego statystycznych własności. Szczególnie istotne jest zbadanie zgodności, co może być sformułowane na różne sposoby w zależności od struktury przestrzeni parametrów A . Kluczowym zagadnieniem w badaniu zgodności jest określenie szybkości wzrostu sita określonej przez m_n . W praktyce, musimy zagwarantować, aby ciąg m_n zmierzał do

nieskończoności „odpowiednio wolno”, gdy $n \rightarrow \infty$. Szczegółowe warunki (dotyczące przede wszystkim własności sita S_{m_n} oraz ciągu m_n), które gwarantują w ogólnym przypadku zgodność estymatorów skonstruowanych z wykorzystaniem metody sita przedstawiono m.in. we wspomnianej już pracy Geman i Hwanga (1982).

Metodę sita można zatem stosować do wielu zagadnień estymacji nieparametrycznej. Często otrzymane w ten sposób rezultaty są dualne do innych metod pozwalając spojrzeć na nie w nowym świetle (np. Geman i Hwang (1982) lub Walter i Blum (1984)). Najprostszym tego przykładem jest wykorzystanie sita histogramowego do estymacji gęstości rozkładu. W tym przypadku otrzymanym estymatorem jest dobrze znany histogram o szerokości okna $h = 1/m_n$. Wybierając odpowiednie sita możemy otrzymać także: projekcyjny estymator gęstości (tzn. estymator skonstruowany w oparciu o wielomiany ortogonalne), estymator największej wiarygodności z funkcją kary (ang. *Maximum Penalized Likelihood*) lub estymator oparty na wielomianowych funkcjach giętych w przypadku nieparametrycznej estymacji funkcji regresji.

Interesującym rezultatem jest także zastosowanie do problemu estymacji gęstości tzw. sita splotowego (Geman i Hwang (1982))

$$S_{m_n} = \left\{ f : f(x) = \int_{-\infty}^{\infty} \frac{m_n}{\sqrt{2\pi}} \exp \left\{ -\frac{m_n^2}{2}(x-y)^2 \right\} F(dy), \right\},$$

gdzie F – dowolna dystrybucja. Elementem zbioru S_{m_n} jest wówczas m.in. estymator jądrowy Rosenblatt-Parzena (dla jądra gaussowskiego). Co ciekawe jednak, maksymalizacja funkcji wiarygodności na zbiorze S_{m_n} prowadzi do estymatora innego niż estymator jądrowy.

Wprowadzona przez Grenandera (1981) metoda sita, której główną ideą jest aproksymacja modelu nieskończenie wymiarowego (nieparametrycznego) ciągiem skończenie wymiarowych modeli parametrycznych, znalazła również zastosowanie w analizie statystycznej szeregów czasowych.

P.Bühlmann (1997,1998) zaproponował algorytm *sieve bootstrap*, który prowadzi do konstrukcji replikacji stacjonarnych szeregów czasowych. W tym przypadku liniowy i odwracalny szereg czasowy, który możemy przedstawić jako proces $AR(\infty)$, tzn.:

$$\sum_{j=0}^{\infty} \phi_j X_{t-j} = \epsilon_t, \quad \phi_0 = 1, \tag{3}$$

(gdzie $\{\epsilon_t\}_{t \in \mathbb{Z}}$ jest ciągiem nieskorelowanych zmiennych losowych o $E[\epsilon_t] = 0$ oraz $\sum_{j=0}^{\infty} \phi_j^2 < \infty$) jest aproksymowany ciągiem procesów autoregresyjnych rzędu $p = p(n)$, który rośnie „odpowiednio wolno” wraz ze wzrostem rozmiaru próby, tzn. $p(n) \rightarrow \infty$ (gdy $n \rightarrow \infty$) oraz $p(n) = o(n)$.

Warto zaznaczyć, że spopularyzowana przez Efrona (1979) metoda bootstrap nie ma jednoznacznego uogólnienia na przypadek szeregów czasowych. Możemy wyróżnić dwie klasy metod: oparte na założeniu o postaci modelu (metody parametryczne) oraz wolne od założeń o modelu (metody nieparametryczne). Przegląd metod replikacji dla szeregów czasowych można znaleźć m.in. w pracach: Carlsteina

(1992), Shao i Tu (1995), Berkowitza i Kiliana (1996), Härdle, Horowitz i Kreiss (2001) lub Bühlmann (2002).

Za najbardziej ogólną nieparametryczną metodę replikacji dla szeregów uznawany jest tzw. bootstrap blokowy (*block bootstrap*, *moving blocks bootstrap*). Zasadniczą ideą tej metody jest podział danych na bloki obserwacji i konstrukcja replikacji bootstrapowych poprzez łączenie wyreplikowanych (losowanych ze zwracaniem) bloków. Metoda blokowa występuje w literaturze w kilku wariantach: bloki mogą być rozłączne (np. Hall (1985), Carlstein (1986)) lub zachodzące na siebie (Hall (1985), Künsch (1989), Politis i Romano (1993)). Wadą bootstrapu blokowego jest niewątpliwie fakt, że otrzymana replikacja nie odzwierciedla w pełni własności wyjściowego szeregu (nie jest stacjonarna i charakteryzuje się nienaturalnym zachowaniem w miejscach łączenia wyreplikowanych bloków).

Metoda *sieve bootstrap* stanowi istotną konkurencję dla metody blokowej. Z jednej strony, *sieve bootstrap* jest metodą ściśle nieparametryczną — nie postulujemy bowiem żadnego skończenia wymiarowego modelu dla danych. Z drugiej, posiada łatwość implementacji charakterystyczną dla metod parametrycznych. Jak zauważają Choi i Hall (2000) oraz Bühlmann (2002) metoda *sieve bootstrap* daje istotną poprawę w stosunku do bootstrapu blokowego dla przypadku liniowych szeregów czasowych. Ponadto odpowiedni wybór aproksymującego rzędu autoregresyjnego ($p(n)$) nie ma aż tak krytycznego znaczenia jak wybór długości bloku w przypadku bootstrapu blokowego.

Niniejsza rozprawa doktorska podzielona jest na trzy części. Rozdział 1 poświęcony został konstrukcji bootstrapowych przedziałów predykcyjnych z wykorzystaniem zaproponowanej przez P. Bühlmann (1997) metody *sieve bootstrap*. Jak wiadomo, prognozowanie to jeden z najbardziej popularnych obszarów zastosowań szeregów czasowych. Oprócz wyznaczenia optymalnych predyktorów (prognoz punktowych) konieczna jest także ocena ich dokładności, tak aby można było oszacować ryzyko związane z decyzjami podejmowanymi w oparciu o te prognozy. Dokładność prognoz może być wyrażona m.in. za pomocą średniokwadratowego błędu predykcji (PMSE). Do oceny dokładności prognoz można wykorzystać także predykcyjne przedziały ufności. Konstrukcja przedziałów predykcyjnych stanowi zatem ważny etap w procesie prognozowania i przeprowadzana jest głównie z zamiarem wskazania możliwej niepewności towarzyszącej prognozom punktowym.

Założenia o modelu, jak również sam algorytm metody oraz przegląd dotychczasowych rezultatów otrzymanych z wykorzystaniem procedury *sieve bootstrap* zamieszczono w Podrozdziale 1.1. Podrozdziały 1.2.2-1.2.5 zawierają m.in. przegląd wyników dotyczących konstrukcji przedziałów predykcyjnych, w tym klasyczne podejście Boxa-Jenkinsa oparte na aproksymacji gaussowskiej oraz metody konstrukcji wykorzystujące bootstrap (m.in. prace Stine'a (1987), Thombsa i Schucany'ego (1990), Cao (1997), Masarotto (1990) i Grigoletto (1998), Alonso, Peña i Romo (2002)).

W Podrozdziale 1.2.6 zaproponowano dwie nowe metody konstrukcji bootstrapowych przedziałów predykcyjnych z wykorzystaniem metody *sieve bootstrap*: przedziały hybrydowe (*hybrid bootstrap*) oraz przedziały studentyzowane (*bootstrap-t*).

Zaproponowane procedury stanowią uogólnienie algorytmu konstrukcji bezwarunkowych przedziałów predykcyjnych (zaproponowanego m.in. przez Stine'a (1987), Masarotto (1990), Grigoletto (1998) dla procesów $AR(p)$) na przypadek modeli liniowych reprezentowanych jako proces $AR(\infty)$.

Rezultaty teoretyczne związane ze zgodnością zaproponowanych metod konstrukcji przedziałów predykcyjnych zostały zamieszczone w Podrozdziale 1.3. Udowodniono m.in.: twierdzenia o zgodności metody *hybrid bootstrap* (Twierdzenie 1.3.4) oraz metody *bootstrap-t* (Twierdzenie 1.3.5) oraz przy dodatkowych założeniach twierdzenia o jednostajnej (w normie supremum) zgodności obu metod (Twierdzenia 1.3.7 oraz 1.3.8). W rozdziale tym podano również związki pomiędzy różnymi metodami dowodzenia zgodności metody bootstrap obejmującymi m.in. zgodność w metryce Lévy'ego oraz zgodność w sensie zaproponowanym przez Belyaeva (1997) oraz Belyaeva i Sjöstedt - de Luna (2000) określanym jako słaba aproksymacja ciągu rozkładów (*weakly approaching sequences*).

Oprócz sformułowania i uzasadnienia zgodności rozkładów bootstrapowych, w Podrozdziale 1.3.6 rozważono również problem zgodności skonstruowanych przedziałów predykcyjnych. Zasadnicze rezultaty związane ze zgodnością przedziałów ufności zawarto w Twierdzeniu 1.3.9 (Zgodność przedziałów hybrydowych) oraz Twierdzeniu 1.3.10 (Zgodność przedziałów studentyzowanych).

W Podrozdziale 1.4 rozważono modyfikację standardowej metody *sieve bootstrap* określaną jako bootstrap wygładzony (*smoothed bootstrap*). Główną ideą bootstrapu wygładzonego jest replikowanie residuów z rozkładu określonego przez wygładzony estymator dystrybuanty zamiast z rozkładu zadanego przez dystrybuantę empiryczną (jak w metodzie standardowej). Uzasadniony został rezultat dotyczący zgodności wygładzonej metody bootstrap (Lemat 1.4.1), który stanowi odpowiednik Lematu 1.3.6 (Bühlmann (1997)) udowodnionego dla standardowej metody *sieve bootstrap*. Przedstawione zostały również argumenty związane z wykorzystaniem bootstrapu wygładzonego do stabilizacji studentyzowanych przedziałów predykcyjnych.

Zachowanie zaproponowanych metod konstrukcji przedziałów predykcyjnych zostało zilustrowane poprzez symulacje komputerowe (Podrozdział 1.5). W badaniach symulacyjnych rozważono różne (zarówno gaussowskie jak i niegaussowskie) rozkłady błędów. Efektywność skonstruowanych przedziałów predykcyjnych porównano z klasyczną metodą Boxa-Jenkinsa poprzez wyznaczenie empirycznych prawdopodobieństw pokrycia oraz średniej długości przedziałów. Zbadano również wpływ wygładzonej metody bootstrap na stabilizację studentyzowanych przedziałów predykcyjnych w przypadku prób o małym rozmiarze.

W Rozdziale 2 rozważono zastosowanie metody sita w estymacji intensywności multiplikatywnych procesów punktowych. Multiplikatywny model intensywności wprowadził O.Aalen (1978), pokazując tym samym, w jaki sposób teoria wielowymiarowych procesów punktowych pozwala na określenie ogólnej struktury, dzięki której możemy analizować zarówno modele związane z klasyczną analizą przeżycia jak i bardziej złożone modele procesów Markowa.

W naszym przypadku estymator skonstruowano z wykorzystaniem sita histogra-

mowego (Podrozdział 2.1).

W Podrozdziale 2.2 przedstawiono krótki przegląd dotychczasowych rezultatów (m.in. prace: Karra (1987), Leśkowska i Róžańskiego (1989a) oraz Leśkowska (1989)).

Konstrukcja histogramowego estymatora intensywności została przeprowadzona w 2.3. W Podrozdziale 2.4 zawarte zostały zasadnicze wyniki teoretyczne, w tym twierdzenie o zgodności skonstruowanego estymatora (Twierdzenie 2.4.1) oraz twierdzenie o asymptotycznym rozkładzie normalnym (Twierdzenie 2.4.2). Otrzymane rezultaty stanowią uogólnienie wyników Leśkowska i Róžańskiego (1989a).

Rozważono również (Podrozdział 2.5) uogólnienie otrzymanych wyników na przypadek wielowymiarowy. W Podrozdziale 2.6 przedstawiono przykłady modeli multiplikatywnych, dla których możemy konstruować estymatory histogramowe intensywności, w tym m.in. klasyczny model analizy przeżycia uwzględniający cenzurowanie obserwacji oraz wielostanowy model Markowa.

Ostatni podrozdział Rozdziału 2 poświęcony jest symulacjom komputerowym, które ilustrują zachowanie histogramowego estymatora intensywności. W podrozdziale tym rozważono różne rozkłady spotykane w analizie przeżycia oraz zbadano wpływ wyboru ciągu $m(n)$ (określającego szybkość wzrostu sita) na wyniki empirycznego prawdopodobieństwa pokrycia wyznaczonego dla asymptotycznych przedziałów ufności (skonstruowanych w oparciu o twierdzenie o rozkładzie asymptotycznym estymatora histogramowego).

Ostatni, trzeci rozdział niniejszej pracy zawiera wyniki związane z zastosowaniem metody sita do estymacji funkcji dryfu dla procesów Itô. Jak już wspominaliśmy metoda sita stała się popularnym narzędziem estymacji nieparametrycznej, stosowanym również z powodzeniem do estymacji parametrów funkcyjnych procesów stochastycznych, takich jak intensywność procesu punktowego czy funkcja dryfu w przypadku procesów dyfuzyjnych i modeli semimartynałowych.

Podrozdział 3.1 zawiera przegląd wyników związanych z wykorzystaniem metody sita do estymacji funkcji dryfu dla modeli dyfuzyjnych i semimartynałowych (prace: Gemana i Hwanga (1982), Nguyena i Phama (1982), Leśkowska i Róžańskiego (1989b) oraz McKeague'a (1986)).

Konstrukcja estymatora największej wiarygodności dla funkcji dryfu procesów Itô została przedstawiona w Podrozdziale 3.2. Podobnie jak w Rozdziale 2 estymator ten został skonstruowany na bazie sita histogramowego.

Podstawowe własności asymptotyczne estymatora histogramowego funkcji dryfu zawarto w podrozdziale 3.3. Wyniki te obejmują: zgodność (Twierdzenie 3.3.1), asymptotyczną normalność (Twierdzenie 3.3.2) oraz przy dodatkowych warunkach również zgodność w L^1 (Twierdzenie 3.3.3). Otrzymane rezultaty stanowią uogólnienie wyników Leśkowska i Róžańskiego (1989b).

W Podrozdziale 3.4 przedstawiono przykłady wykorzystane następnie do ilustracji zachowania histogramowego estymatora funkcji dryfu z wykorzystaniem symulacji komputerowych (Podrozdział 3.5). W podrozdziale tym zbadano m.in. w jaki sposób sposób wybór ciągu $m(n)$ (odpowiedzialnego za szybkość wzrostu sita) wpływa na efektywność asymptotycznych przedziałów ufności oraz na kryteria dokładności estymatora, takie jak MISE i MIAE.

Wyniki przedstawione w Rozdziale 1 zostały w części opublikowane w pracach A.Zagdańskiego (1999,2001a). Rezultaty Rozdziału 2 zamieszczono w artykułach: R.Róžański i A.Zagdański (2001) oraz A.Zagdański (2001b). Rozdział 3 zawiera wyniki, które będą opublikowane w pracy R.Róžańskiego i A.Zagdańskiego (2002).

Symulacje komputerowe wykonano w pakiecie MATLAB z wykorzystaniem komputerów Wrocławskiego Centrum Sieciowo-Superkomputerowego oraz komputerów Instytutu Matematyki Politechniki Wrocławskiej.

Autor pragnie serdecznie podziękować Panu Prof. Romanowi Róžańskiemu za opiekę naukową oraz wiele cennych uwag pomocnych przy pisaniu tej rozprawy.

Rozdział 1

Bootstrapowe przedziały predykcyjne z wykorzystaniem metody *sieve bootstrap*

1.1 Metoda *sieve bootstrap*

1.1.1 Założenia o modelu

Niech $\{X_t\}_{t \in Z}$ będzie procesem stacjonarnym o wartościach rzeczywistych i wartości oczekiwanej $E[X_t] = \mu_X$. Jeżeli $\{X_t\}_{t \in Z}$ jest procesem całkowicie niedeterministycznym wówczas wiadomo z tw. Wolda (np. Anderson 1971 → dodatek A), że $\{X_t - \mu_X\}_{t \in Z}$ może być przedstawiony jako jednostronny proces ruchomej średniej rzędu nieskończonego ($MA(\infty)$), tzn.

$$X_t - \mu_X = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad \psi_0 = 1, \quad t \in Z, \quad (1.1)$$

gdzie $\{\epsilon_t\}_{t \in Z}$ jest ciągiem nieskorelowanych zmiennych losowych o $E[\epsilon_t] = 0$ i $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.

Będziemy wymagali również aby proces (1.1) był odwracalny, zawężając tym samym nieco klasę dopuszczalnych procesów stacjonarnych. Odpowiednie warunki gwarantujące odwracalność (sformułowane np. u Anderson'a 1971, Tw. 7.6.9 → dodatek A) pozwalają na przedstawienie $\{X_t\}_{t \in Z}$ jako procesu autoregresyjnego nieskończonego rzędu ($AR(\infty)$), tzn.

$$\sum_{j=0}^{\infty} \phi_j (X_{t-j} - \mu_X) = \epsilon_t, \quad \phi_0 = 1, \quad t \in Z, \quad (1.2)$$

gdzie $\sum_{j=0}^{\infty} \phi_j^2 < \infty$.

Przyjmując następujące oznaczenia:

$$\Phi(z) = \sum_{j=0}^{\infty} \phi_j z^j, \quad \phi_0 = 1, \quad z \in C,$$

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad \psi_0 = 1, \quad z \in C,$$

możemy przedstawić X_t w reprezentacji $MA(\infty)$ lub $AR(\infty)$:

$$\text{model } AR(\infty) : \quad \Phi(B)(X_t - \mu_X) = \epsilon_t$$

lub

$$\text{model } MA(\infty) : \quad (X_t - \mu_X) = \Psi(B)\epsilon_t,$$

gdzie B oznacza operator przesunięcia wstecz, tzn. $B X_t = X_{t-1}$. Przy powyższych oznaczeniach $\Psi(z) = 1/\Phi(z)$.

Oznaczmy również przez $\mathcal{F}_t = \sigma(\{\epsilon_s : s \leq t\})$ σ -ciało generowane przez $\{\epsilon_s\}_{s=-\infty}^t$.

Założenia dla modelu:

(A1') $X_t - \mu_X = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$, $\psi_0 = 1$ ($t \in Z$) gdzie $\{\epsilon_t\}_{t \in Z}$ ciąg stacjonarny i ergodyczny oraz $E[\epsilon_t | \mathcal{F}_{t-1}] \equiv 0$, $E[\epsilon_t^2 | \mathcal{F}_{t-1}] \equiv \sigma^2 < \infty$, $E|\epsilon_t|^s < \infty$ dla pewnego $s \geq 4$.

(A1) $X_t - \mu_X = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$, $\psi_0 = 1$ ($t \in Z$) gdzie $\{\epsilon_t\}_{t \in Z}$ ciąg i.i.d. oraz $E[\epsilon_t] = 0$, $E|\epsilon_t|^s < \infty$ dla pewnego $s \geq 4$.

(A2) $\Psi(z)$ jest oddzielona od zera dla $|z| \leq 1$, $\sum_{j=0}^{\infty} j^r |\psi_j| < \infty$ dla pewnego naturalnego r .

(B) $p = p(n) \rightarrow \infty$, $p(n) = o(n)$ ($n \rightarrow \infty$) oraz $\hat{\phi}_p = (\hat{\phi}_{1,n}, \dots, \hat{\phi}_{p,n})$ spełniają empiryczne równania Yule'a-Walkera, tzn.:

$$\hat{\Gamma}_p \hat{\phi}_p = -\hat{\gamma}_p,$$

gdzie $\hat{\Gamma}_p = [\hat{\gamma}(i-j)]_{i,j=1}^p$, $\hat{\gamma}_p = (\hat{\gamma}(1), \dots, \hat{\gamma}(p))'$, gdzie $\hat{\gamma}$ jest estymatorem funkcji autokowariancji zdefiniowanym jako:

$$\hat{\gamma}(j) = \frac{1}{n} \sum_{t=1}^{n-|j|} (X_{t,n} - \bar{X})(X_{t+|j|,n} - \bar{X}), \quad |j| \leq n-1,$$

gdzie $\bar{X} = \sum_{t=1}^n X_t$.

1.1.2 Algorytm metody

Zasadniczą ideą metody *sieve bootstrap* jest aproksymacja procesu X_t za pomocą procesów autoregresyjnych rzędu $p = p(n)$ rosnącego „odpowiednio wolno” wraz ze wzrostem rozmiaru próby n , tzn. $p(n) \rightarrow \infty$ gdy $(n \rightarrow \infty)$ ale $p(n) = o(n)$. Innymi słowy, aproksymujemy nieskończenie wymiarowy model nieparametryczny (który możemy przedstawić jako proces $AR(\infty)$) ciągiem skończenie wymiarowych modeli

parametrycznych – procedura dobrze znana jako metoda sita (Grenander (1981)). Poniżej przedstawiamy szczegółowo opis konstrukcji replikacji za pomocą metody *sieve bootstrap*.

Algorytm konstrukcji replikacji

Krok 1

Niech $X_1, \dots, X_n$ będzie realizacją procesu $\{X_t\}_{t \in \mathbb{Z}}$. Wykorzystując sugestię Bühlmann (1997) wybieramy aproksymujący rząd $p = p(n)$ wykorzystując kryterium Akaike (AIC) dla dopuszczalnego zakresu $[0, p_{\max}(n)]$ gdzie $p_{\max}(n)$ rośnie wraz ze wzrostem n . W praktyce przyjmujemy $p_{\max}(n) = 10 \log_{10}(n)$ (Wielkość standardowo stosowana w pakietach statystycznych).

Krok 2

Estymujemy współczynniki dla modelu $AR(p(n))$, tzn. $\phi_1, \dots, \phi_{p(n)}$ w oparciu o obserwacje $\{X_t\}_{t=1}^n$. Estymatory $\hat{\phi}_p = (\hat{\phi}_1, \dots, \hat{\phi}_{p(n)})'$ są wyznaczone metodą Yule'a-Walker'a (Brockwell, Davies, 1987, str.232-233).

$$\hat{\Gamma}_p \hat{\phi}_p = -\hat{\gamma}_p,$$

gdzie $\hat{\Gamma}_p$ oraz $\hat{\gamma}_p$ są określone jak w założeniu B.

Biorąc pod uwagę złożoność obliczeniową, bardziej efektywnie możemy wyznaczyć estymatory Yule'a-Walker'a wykorzystując rekurencyjny algorytm Durbin'a-Levinson'a :

$$\begin{aligned} \hat{\phi}_{11} &= \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)}, & \hat{\nu}_1 &= \hat{\gamma}(0)[1 - \hat{\phi}_{11}^2] \\ \hat{\phi}_{mm} &= \left(\hat{\gamma}(m) - \sum_{j=1}^{m-1} \hat{\phi}_{m-1,j} \hat{\gamma}(m-j) \right) / \hat{\nu}_{m-1} \\ \begin{bmatrix} \hat{\phi}_{m,1} \\ \vdots \\ \hat{\phi}_{m,m-1} \end{bmatrix} &= \begin{bmatrix} \hat{\phi}_{m-1,1} \\ \vdots \\ \hat{\phi}_{m-1,m-1} \end{bmatrix} - \hat{\phi}_{mm} \begin{bmatrix} \hat{\phi}_{m-1,m-1} \\ \vdots \\ \hat{\phi}_{m-1,1} \end{bmatrix} \\ \hat{\nu}_m &= \hat{\nu}_{m-1}(1 - \hat{\phi}_{mm}^2) \end{aligned}$$

Wykorzystanie powyższych zależności rekurencyjnych uwalnia nas od konieczności odwracania macierzy wymaganej przy bezpośrednim wyznaczeniu $\hat{\phi}_p$.

Krok 3

Wyznaczamy residua

$$\hat{\varepsilon}_{t,n} = \sum_{j=0}^{p(n)} \hat{\phi}_{j,n} X_{t-j}, \quad \hat{\phi}_{0,n} = 1, \quad t = p+1, \dots, n.$$

Krok 4

Konstruujemy replikację obserwacji. W tym celu centrujemy uprzednio wyznaczone residua

$$\tilde{\varepsilon}_{t,n} = \hat{\varepsilon}_{t,n} - \frac{1}{n-p} \sum_{t=p+1}^n \hat{\varepsilon}_{t,n}, \quad t = p+1, \dots, n,$$

a następnie losujemy residua ε_t^* z rozkładu określonego przez dystrybuantę empiryczną na bazie $\{\tilde{\varepsilon}_{t,n}\}_{t=p+1}^n$, tzn. ε_t^* i.i.d. $\sim \hat{F}_{\varepsilon,n}$, gdzie

$$\hat{F}_{\varepsilon,n}(u) = \frac{1}{n-p} \sum_{t=p+1}^n 1_{[\tilde{\varepsilon}_{t,n} \leq u]}.$$

Krok 5

Ostatecznie określamy replikację bootstrapową $\{X_1^*, \dots, X_n^*\}$ poprzez zależność rekurencyjną

$$\sum_{j=0}^{p(n)} \hat{\phi}_{j,n} X_{t-j}^* = \varepsilon_t^*, \quad \hat{\phi}_{0,n} = 1. \quad (1.3)$$

W praktyce, generujemy replikację $\{X_t^*\}$ rozpoczynając rekurencję od pewnych wartości początkowych, np. równych wyreplikowanym innowacjom ε_t^* .

1.1.3 Uwagi do założeń i algorytmu

- W metodzie *sieve bootstrap* przyjmujemy kluczowe założenie, że obserwacje $X_1, \dots, X_n$ są skończoną realizacją procesu $AR(\infty)$ jak w 1.2.

Zauważmy, że reprezentacja ta obejmuje ważną klasę modeli $ARMA(p, q)$:

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + \sum_{k=1}^q \psi_k \varepsilon_{t-k} + \varepsilon_t, \quad t \in Z,$$

z odwracalnym wielomianem ruchomej średniej, tzn. $\Psi(z) = \sum_{k=0}^q \psi_k z^k$, $z \in C$ ma wszystkie pierwiastki na zewnątrz koła jednostkowego $\{z \in C, |z| \leq 1\}$.

- Bühlmann (1997) zauważa, że alternatywnie moglibyśmy w algorytmie konstrukcji replikacji wykorzystać aproksymację ciągiem procesów ruchomej średniej. Aproksymacja autoregresyjna jako metoda liniowa jest bardziej popularna, szybsza i znana jako skuteczna technika w różnych sytuacjach (Berk (1974), An i inni (1982), Hannan (1987)).
- Jak podkreśla Bühlmann (1997) metoda *sieve bootstrap* jest łatwiejsza w zastosowaniu niż np. bootstrap blokowy w przypadku gdy mamy dane o nierównych odstępach czasowych lub szeregi o wielu brakujących obserwacjach.
- Do estymacji parametrów aproksymującego procesu autoregresyjnego możemy wykorzystać inną metodę niż metoda Yule’a-Walker’a. Niektórzy autorzy proponują zastąpienie metody Yule’a-Walker’a (Y-W) metodą najmniejszych kwadratów (np. Kilian, Berkowitz i Birgean (1999), Chang i Park (2001)). Metoda Y-W jest zalecana w przypadku małych prób ponieważ zawsze daje model przyczynowy ($MA(\infty)$) (Brockwell i Davis (1987)). Wraz ze wzrostem rozmiaru próby obie metody estymacji stają się równoważne.

- Założenie A2 jest spełnione dla modeli o wielomianowo zanikających współczynnikach $\{\psi_j\}_{j=0}^{\infty}$ (lub równoważnie $\{\phi_j\}_{j=0}^{\infty}$). Modele $ARMA(p, q)$ spełniają to założenie z wykładniczo zanikającymi współczynnikami $\{\psi_j\}$. Zauważmy także, że założenie A2 implikuje, że $\Phi(z)$ jest oddzielona od zera dla $|z| \leq 1$ oraz $\sum_{j=0}^{\infty} j^r |\phi_j| < \infty$.
- W literaturze proponowane są różne metody wyboru rzędu $p(n)$ dla aproksymującego procesu autoregresyjnego.

Podobnie jak Bühlmann (1997) wybieramy rząd $p(n)$, który minimalizuje kryterium AIC przy założeniu gaussowskich innowacji. Shibata (1980) pokazał optymalność AIC dla predykcji w modelu $AR(\infty)$. Ponadto zastosowanie kryterium AIC jest zalecane w przypadku estymacji wariancji estymatora $\hat{\theta} = \bar{X}_n$ (Bühlmann (1997,2002)).

Warto zauważyć, że do wyboru rzędu $p(n)$ możemy wykorzystać narzędzia graficzne takie jak *prewhitening* (Bühlmann (1997)). W tym przypadku dopasowujemy model autoregresyjny dla kilku wartości p , wyznaczamy residua i konstruujemy estymator gęstości spektralnej. Wybieramy wówczas taką wartość p , dla której estymowane spektrum jest bliskie stałej.

Alonso, Peña i Romo (2001) proponują aby do estymacji rzędu $p(n)$ aproksymującego modelu autoregresyjnego wykorzystać kryterium AICC (wersja kryterium AIC o skorygowanym obciążeniu), które nakłada większą karę dla modeli dużego rzędu w odróżnieniu do kryterium AIC. Ponadto, jak pokazali Hurvich i Tsai (1989) zmiana wartości maksymalnego dopuszczalnego rzędu, tzn. $p_{max}(n)$ ma mniejszy wpływ na *AICC* niż na kryteria *AIC* czy *BIC*. Alonso, Peña i Romo przyjęli $p_{max}(n) = n/10$ wielkość rekomendowaną przez Bhansali'ego (1983)

Bühlmann (2002) stwierdził, że dotychczasowe badania symulacyjne sugerują, że efektywność metody *sieve bootstrap* jest stosunkowo nieczuła ze względu na wybór aproksymującego rzędu autoregresyjnego, jeżeli tylko wybrany rząd jest rozsądny.

- Badania teoretyczne i symulacyjne wykazały, że dla klasy liniowych i odwracalnych szeregów czasowych metoda *sieve bootstrap* charakteryzuje się dużą dokładnością, zazwyczaj lepszą niż bardziej ogólna nieparametryczna metoda replikacji jaką jest bootstrap blokowy.

Bühlmann (1997) pokazał, że w przypadku estymatora średniej $\hat{\theta} = \sum_{t=1}^n X_t/n$ wykorzystując kryterium AIC do określenia aproksymującego rzędu autoregresyjnego $p(n)$ otrzymujemy:

$$nVar^*(\hat{\theta}^*) - nVar(\hat{\theta}) = O_P(n^{-(v-2)/(2v)})$$

jeżeli tylko współczynniki autoregresyjne $\{\phi_j\}_{j \in \mathbb{N}}$ maleją odpowiednio szybko, tzn. $\phi_j \leq const \cdot j^{-v}$ ($v > 2$). W szczególności, jeżeli ϕ_j maleją wykładniczo

otrzymujemy:

$$nVar^*(\hat{\theta}^*) - nVar(\hat{\theta}) = O_P(n^{-1/2+\kappa})$$

dla dowolnego $\kappa > 0$.

Powyższe wyniki pokazują, że metoda *sieve bootstrap* adaptuje się automatycznie w zależności od stopnia zanikania w wyjściowej strukturze zależności. Ta pożądana własność nie występuje w przypadku bootstrapu blokowego.

Metoda *sieve bootstrap* okazała się efektywna nie tylko w przypadku estymacji wariancji.

1.1.4 Przegląd wyników

Otrzymaną replikację szeregu możemy wykorzystać w wielu różnych zagadnieniach związanych z analizą statystyczną szeregów czasowych. W szczególności, jeżeli mamy statystykę $T_n = T_n(X_1, \dots, X_n)$ będącą mierzalną funkcją n obserwacji, możemy określić odpowiadającą jej statystykę bootstrapową T_n^* następująco: $T_n^* = T_n(X_1^*, \dots, X_n^*)$. Przedstawimy teraz przegląd wyników otrzymanych z wykorzystaniem metody *sieve bootstrap*.

Bühlmann (1997) udowodnił zgodność metody *sieve bootstrap* dla wielu różnych statystyk T_n , w tym dla średniej arytmetycznej i pewnej klasy estymatorów nieliniowych postaci:

$$T_n = f\{(n - m + 1)^{-1} \sum_{t=1}^{n-m+1} g(X_t, \dots, X_{t+m-1})\},$$

gdzie $g = (g_1, \dots, g_q)'$ oraz $f : R^q \rightarrow R^{\tilde{q}}$, ($q, \tilde{q} \geq 1$) spełniają określone warunki gładkości. Klasa ta obejmuje m.in. próbkową autokowariancję, autokorelację, częściową autokorelację oraz estymatory Yule'a-Walker'a dla modeli autoregresyjnych.

Bühlmann (1998) rozważył także uogólnienie metody *sieve bootstrap* na przypadek niestacjonarnych szeregów czasowych postaci:

$$Y_t = s(t) + Z_t, \quad t \in Z,$$

gdzie $s(t)$ jest deterministycznym trendem a Z_t jest szumem stacjonarnym o średniej zero. Zaproponowana metoda została wykorzystana następnie do konstrukcji nieparametrycznych przedziałów ufności dla funkcji trendu.

Berkowitz, Birgean i Kilian (1999) przedstawili interesujące porównanie nieparametrycznych metod replikacji dla szeregów czasowych spotykanych w ekonomii. Badania symulacyjne oparte były na szeregach dotyczących: nominalnej stopy procentowej (3-miesięczne treasury bills), stopy wzrostu produkcji przemysłowej, stopy inflacji cen konsumpcyjnych, stopy wymiany Yen-Dolar. Kryterium wykorzystanym do porównania różnych metod było empiryczne prawdopodobieństwo pokrycia dla przedziałów ufności skonstruowanych dla gęstości spektralnej oraz funkcji odpowiedzi impulsowej.

Pośród rozważanych w porównaniu nieparametrycznych metod replikacji, obejmujących m.in. różne wersje bootstrapu blokowego, bootstrap stacjonarny, bootstrap oparty na dekompozycji Cholesky'ego znalazły się także dwie wersje metody sieve bootstrap: „encompassing sieve bootstrap” oraz „AIC sieve bootstrap”. W przypadku metody „encompassing sieve bootstrap” dopasowujemy aproksymujący model $AR(p_{max}(n))$, gdzie $p_{max}(n)$ jest górnym ograniczeniem czyli maksymalnym dopuszczalnym rzędem dla aproksymującego modelu autoregresyjnego. Z kolei w metodzie „AIC sieve bootstrap” wybieramy rząd aproksymującego procesu autoregresyjnego $p(n)$ w oparciu o kryterium AIC, pod warunkiem, że $p(n) \leq p_{max}(n)$.

Najbardziej dokładną zarówno w przypadku funkcji odpowiedzi impulsowej jak i gęstości spektralnej okazała się metoda „encompassing sieve bootstrap”, która daje dokładne bootstrapowe przedziały ufności dla małych prób, pod warunkiem, że zadbamy o odpowiedni dobór rzędu aproksymacji autoregresyjnej.

Berkowitz, Birgean i Kilian sugerują, że zastosowanie „oszczędnych” (*parsimonious*) aproksymujących modeli autoregresyjnych otrzymanych z wykorzystaniem kryterium doboru rzędu modelu takich jak np. AIC, jak sugerował to Bühlmann (1997), nie wydaje się być obiecujące dla modeli nieskończenie wymiarowych. Nawet kryterium AIC, znane jako najbardziej liberalne w tej klasie wykazuje tendencje do niedoszacowania w przypadku małych prób. Autorzy opracowania podkreślają także potrzebę rozwijania dalszych badań dotyczących wyboru rzędu modelu aproksymującego dla określonego rozmiaru próby, w szczególności opracowanie pewnego rodzaju metod kciuka, które pozwoliłyby praktykom na stosowanie w swoich badaniach nieparametrycznych metod replikacji.

Choi i Hall (2000) stosując algorytm *sieve bootstrap* pokazali własności drugiego rzędu dla przypadku przedziałów ufności skonstruowanych poprzez kalibrację przedziałów pierwszego rzędu z wykorzystaniem podwójnego bootstrapu.

Chang i Park (2001) rozważyli wykorzystanie metody *sieve bootstrap* do testowania pierwiastków jednostkowych dla ogólnych procesów liniowych. Otrzymany test stanowi bootstrapową wersję testu ADF (Augmented-Dickey-Fuller) zaproponowanego przez Said'a i Dickey'a (1984). Udowodniono własności asymptotyczne przy istotnie słabszych założeniach niż te przyjęte przez Said'a i Dickey'a. Symulacje pokazują, że wykorzystanie metody *sieve bootstrap* przynosi znaczącą poprawę rozmiaru testu dla skończonych prób.

1.2 Predykcja przedziałowa

1.2.1 Optymalny predyktor liniowy

Wiadomo (np. Brockwell i Davies, 1987, str. 159-162), że dla procesu stacjonarnego o średniej 0, najlepszą (w sensie odległości średniokwadratowej) kombinacją liniową $1, X_1, \dots, X_n$ prognozującą X_{n+h} , ($h \geq 1$) jest projekcja ortogonalna X_{n+h} na domkniętą podprzestrzeń liniową $\overline{\text{span}}\{X_1, \dots, X_n\}$. Możemy zatem przedstawić

h -krokowy predyktor jako:

$$P_n X_{n+h} = P_{\overline{sp}} \{X_1, \dots, X_n\} X_{n+h}.$$

Powyższy predyktor możemy wyznaczyć z odpowiednich równań projekcyjnych. Bardziej efektywne jest jednak wykorzystanie algorytmu rekurencyjnego np. algorytmu innowacji (Brockwell i Davies, 1987, str.167-168), tzn.

$$P_n X_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}), \quad (1.4)$$

gdzie jednokrokowe prognozy $\hat{X}_{m+1}$ są dane przez

$$\hat{X}_{m+1} = \begin{cases} 0 & , m = 0 \\ \sum_{j=1}^m \theta_{mj} (X_{m+1-j} - \hat{X}_{m+1-j}) & , m \geq 1 \end{cases}$$

natomiast współczynniki θ_{mj} wyznaczamy z zależności rekurencyjnej

$$\begin{cases} v_0 &= \gamma(0), \\ \theta_{m,m-k} &= v_k^{-1} (\gamma(m-k) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{m,m-j} v_j), \quad k = 0, 1, \dots, m-1, \\ v_m &= \gamma(0) - \sum_{j=0}^{m-1} \theta_{m,m-j}^2 v_j. \end{cases}$$

Ponadto, średniokwadratowy błąd predykcji (PMSE), mierzący niepewność związaną z odpowiednimi prognozami możemy wyrazić jako

$$\sigma_n^2(h) := E(X_{n+h} - P_n X_{n+h})^2 = \gamma(0) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 v_{n+h-1-j}. \quad (1.5)$$

1.2.2 Przedziały predykcyjne

Oprócz wyznaczenia optymalnych predyktorów konieczna jest także ocena ich dokładności, tak aby można było oszacować ryzyko związane z decyzjami podejmowanymi w oparciu o te prognozy. Jak wspomniano, dokładność prognoz może być wyrażona m.in. za pomocą średniokwadratowego błędu predykcji (PMSE). Do oceny dokładności prognoz można wykorzystać także predykcyjne przedziały ufności.

Konstrukcja przedziałów predykcyjnych stanowi zatem ważny etap w procesie prognozowania i przeprowadzana jest głównie z zamiarem wskazania możliwej niepewności towarzyszącej prognozom punktowym.

Jak zauważa Chatfield (1998) wyznaczenie prognoz przedziałowych może być godne szczególnej uwagi dla praktyków zajmujących się zagadnieniami związanymi z planowaniem i podejmowaniem decyzji. Wśród argumentów przemawiających za potrzebą konstrukcji przedziałów predykcyjnych wymieniona jest m.in. możliwość zaplanowania różnych strategii i przeanalizowania różnych scenariuszy dla całego zakresu możliwych wyników wskazywanych przez przedział predykcyjny, a także możliwość bardziej gruntownego porównania prognoz uzyskanych różnymi metodami.

Podamy teraz formalną definicję przedziału predykcyjnego.

Definicja 1.2.1 (Przedział predykcyjny)

Przedziałem predykcyjnym nazywamy losowy przedział skonstruowany na bazie przeszłych obserwacji $\underline{X} = (X_1, X_2, \dots, X_n)$ dla przyszłej (nieznanej) wartości X_{n+h} , $h \geq 1$. Przedział predykcyjny dla X_{n+h} na poziomie ufności $1 - 2\alpha$ definiujemy jako

$$I(h, \underline{X}) = [L(\underline{X}), R(\underline{X})],$$

tak aby $P(L(\underline{X}) \leq X_{n+h} \leq R(\underline{X})) = 1 - 2\alpha$.

Warto zauważyć, że w literaturze występuje pewna niejednoznaczność w stosowanej terminologii. Przedziały predykcyjne określane są jako: *forecast limits* (Wei (1990)), *prediction bounds* (Brockwell i Davis (1987)), *forecast region* (Hyndman (1995)). Najpowszechniej używane jest jednak określenie *prediction intervals*.

Chatfield (1998) zauważa także, że autorzy monografii na temat analizy szeregów czasowych poświęcają zbyt mało uwagi zagadnieniom związanym z prognozą przedziałową ograniczając się najczęściej do klasycznej metody Boxa-Jenkinsa.

1.2.3 Gaussowskie przedziały predykcyjne

Dla stacjonarnego procesu gaussowskiego możemy konstruować przedziały predykcyjne wykorzystując fakt (Brockwell i Davies, 1987, str.175), że błąd predykcji $X_{n+h} - P_n X_{n+h}$ ma rozkład normalny o średniej zero i wariancji $\sigma_n^2(h)$.

Gaussowski przedział predykcyjny (na poziomie ufności $1 - 2\alpha$) jest zatem następującej postaci:

$$I_G(h) = [P_n X_{n+h} - \Phi_{1-\alpha} \sigma_n(h), P_n X_{n+h} + \Phi_{1-\alpha} \sigma_n(h)], \quad (1.6)$$

gdzie $\Phi_{1-\alpha}$ oznacza kwantyl rzędu $(1 - \alpha)$ standardowego rozkładu normalnego.

Gaussowskie przedziały predykcyjne, powszechne znane w literaturze jako przedziały predykcyjne Boxa-Jenkinsa, konstruowane są więc przy założeniu, że rozkład zakłóceń jest normalny i ewentualne odstępstwa od normalności mogą mieć niekorzystny wpływ na ich dokładność. Ponadto, stosując metodę Boxa-Jenkinsa nie uwzględniamy zmienności związanej z estymacją parametrów modelu. Wszystko to może powodować, że tak skonstruowane przedziały będą charakteryzowały się mniejszym niż nominalne prawdopodobieństwem pokrycia.

W przypadku niegaussowskich szeregów czasowych do konstrukcji przedziałów predykcyjnych możemy wykorzystać metodę bootstrap.

1.2.4 Bootstrapowe przedziały predykcyjne — przegląd wyników

Przedstawimy teraz przegląd wyników dotyczących konstrukcji bootstrapowych przedziałów predykcyjnych.

Stine (1987) zaproponował konstrukcję bootstrapowych przedziałów predykcyjnych dla procesów autoregresyjnych znanego rzędu ($AR(p)$). Przedstawimy teraz pokrótce ideę algorytmu Stine'a:

- Zakładając, że dysponujemy obserwacjami $\underline{X} = (X_1, \dots, X_n)$ będącymi realizacją procesu:

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t$$

możemy wyznaczyć estymatory $\hat{\underline{\phi}} = (\hat{\phi}_1, \dots, \hat{\phi}_p)$ wykorzystując metodę najmniejszych kwadratów.

- Wyznaczamy residua:

$$\hat{\epsilon}_i = X_i - \hat{\delta} - \hat{\phi}_1 X_{i-1} - \dots - \hat{\phi}_p X_{i-p} \quad i = p+1, p+2, \dots, n$$

oraz estymator wariancji

$$\hat{\sigma}^2 = \frac{1}{n-2p} \sum_{i=p+1}^n \hat{\epsilon}_i^2.$$

- Generujemy replikację obserwacji $\underline{X}^* = (X_1^*, \dots, X_n^*)$ rozpoczynając od losowo wybranego początkowego bloku $X_0^* = (X_{-p+1}^*, \dots, X_0^*)$ i wykorzystując następnie zależność rekurencyjną:

$$X_i^* = \hat{\delta} + \hat{\phi}_1 X_{i-1}^* + \dots + \hat{\phi}_p X_{i-p}^* + \epsilon_i^* \quad i = 1, 2, \dots, n,$$

gdzie $\{\epsilon_i^*\}$ są replikowane z rozkładu określonego przez następujący estymator dystrybucyjny (wykorzystany zamiast standardowo stosowanej w tym przypadku dystrybucyjnej empirycznej):

$$F_n = \begin{cases} \frac{1}{2} + \frac{\#\left\{\sqrt{\frac{n-p}{n-2p}} |\hat{\epsilon}_i| \leq x\right\}}{2(n-p)}, & x \geq 0, \\ 1 - F_n(-x), & x < 0 \end{cases}$$

Jak podkreśla Stine estymator ten posiada pewne własności minimaksowe, a skalowanie powoduje, że ϵ_t^* ma pożądaną wariancję. Zauważmy, że $E^*(\epsilon_t^*) = 0$ oraz $E^*(\epsilon_t^*)^2 = \hat{\sigma}^2$.

- Estymujemy nieznaną rozkład błędu predykcji

$$\Delta_n(h) = X_{n+h} - \hat{X}_{n+h}$$

odpowiednim rozkładem bootstrapowego błędu predykcji

$$\Delta_n^*(h) = X_{n+h}^* - \hat{X}_{n+h}^*,$$

gdzie: $\hat{X}_{n+h}$ – estymator optymalnego predyktora liniowego, $\hat{X}_{n+h}^*$ jest optymalnym predyktorem liniowym dla X_{n+h}^* skonstruowanym na bazie $X_1^*, \dots, X_n^*$.

- W praktyce do estymacji kwantyli rozkładu błędu predykcji możemy wykorzystać metodę Monte Carlo zakładając, że dysponujemy odpowiednio dużą liczbą replikacji bootstrapowych:

$$\Delta^{*(b)} = X_{n+h}^{*(b)} - \hat{X}_{n+h}^{*(b)}, \quad b = 1, \dots, B.$$

Zakładając, że rozkład błędu jest symetryczny i o skończonych momentach, tzn. $P(\epsilon_t \leq x) = F_\epsilon(x) = 1 - F_\epsilon(-x)$ oraz $E(|\epsilon_t|^k) < \infty$, dla $k = 1, 2, \dots$ Stine udowodnił zgodność skonstruowanych przedziałów predykcyjnych.

Stine rozważył dodatkowo również wykorzystanie metody bootstrap do estymacji średniokwadratowego błędu predykcji (PMSE).

Thombs i Schucany (1990) wykorzystali metodę bootstrap do konstrukcji warunkowych przedziałów predykcyjnych dla modeli autoregresyjnych znanego rzędu ($AR(p)$). Główną ideą jest w tym przypadku aproksymacja warunkowego rozkładu przyszłej wartości X_{t+h} pod warunkiem $X_{t-p+1}, \dots, X_t$. Do generowania replikacji bootstrapowych Thombs i Schucany wykorzystali oprócz standardowej (*forward*) również wsteczną (*backward*) reprezentację modelu autoregresyjnego.

Zasadnicze etapy algorytmu Thombsa i Schucany'ego są następujące:

- Zakładając, że dysponujemy obserwacjami $\underline{X} = (X_{t-n+1}, X_{t-n+2}, \dots, X_t)$ będącymi realizacją procesu $AR(p)$ estymujemy parametry modelu $(\hat{\delta}, \hat{\phi}_1, \dots, \hat{\phi}_p)$ wykorzystując metodę najmniejszych kwadratów.
- Wyznaczamy scentrowane i przeskalowane residua wsteczne (*backward*):

$$\tilde{e}_i = \left[\frac{n-p}{n-2p} \right]^{1/2} \left(\hat{e}_i - \frac{1}{n-p} \sum_{j=1}^{n-p} \hat{e}_j \right),$$

gdzie

$$\hat{e}_i = X_i - \hat{\delta} - \hat{\phi}_1 X_{i+1} - \dots - \hat{\phi}_p X_{i+p} \quad i = 1, 2, \dots, t-p.$$

Niech $\hat{F}_e$ oznacza dystrybuantę empiryczną określoną na zbiorze residuów $\{\tilde{e}_i\}_{i=1, \dots, t-p}$

- Konstruujemy replikację bootstrapową obserwacji. W tym celu określamy p ostatnich wartości w replikacji równym ostatnim wartościom wyjściowego szeregu

$$X_j^* = X_j \quad \text{dla } j = t-p+1, \dots, t.$$

Wartości replikacji dla wcześniejszych chwil j otrzymujemy następująco:

$$X_j^* = \hat{\delta} + \hat{\phi}_1 X_{j+1}^* + \dots + \hat{\phi}_p X_{j+p}^* + \tilde{e}_j^*, \quad j = 1, 2, \dots, t-p,$$

gdzie bootstrapowe replikacje residuów $\tilde{e}_j^*$ są generowane z rozkładu o dystrybuancie $\hat{F}_e$.

- Estymujemy parametry modelu $AR(p)$ na bazie replikacji

$$(X_1^*, \dots, X_{t-p}^*, X_{t-p+1}^*, \dots, X_t^*)$$

otrzymując replikacje bootstrapowe estymatorów $\hat{\delta}^*, \hat{\phi}_1^*, \dots, \hat{\phi}_p^*$.

- Generujemy bootstrapowe residua *forward* $\{\tilde{a}_j^*\}$ z rozkładu o dystrybuancie empirycznej $\hat{F}_a$ określonej na zbiorze scentrowanych i przeskalowanych residuów $\{\tilde{a}_i\}$, $i = p+1, \dots, t$, tzn.

$$\tilde{a}_i = \left[\frac{n-p}{n-2p} \right]^{1/2} \left(\hat{a}_i - \frac{1}{n-p} \sum_{j=1}^{n-p} \hat{a}_j \right),$$

gdzie

$$\hat{a}_i = X_i - \hat{\delta} - \hat{\phi}_1 X_{i-1} - \dots - \hat{\phi}_p X_{i-p} \quad i = p+1, p+2, \dots, t.$$

- Konstruujemy replikacje przyszłych wartości wykorzystując zależność rekurencyjną

$$X_{t+j}^* = \hat{\delta}^* + \hat{\phi}_1^* X_{t+j-1} + \dots + \hat{\phi}_p^* X_{t+j-p} + \tilde{a}_{t+j}^*, \quad j = 1, 2, \dots, h.$$

- Wyznaczamy oszacowanie rozkładu warunkowego na bazie B replikacji, tzn.:

$$G_B^*(x) = \frac{\#\{X_{t+h}^{*b} \leq x\}}{B}.$$

- Wyznaczamy warunkowy przedział bootstrapowy dla X_{t+h} na poziomie ufności $(1 - 2\alpha)$.

$$[L_B^*(\underline{y}), U_B^*(\underline{y})] = [Q_B^*(\alpha), Q_B^*(1 - \alpha)],$$

gdzie $Q_B^* = G_B^{*-1}$.

Należy zaznaczyć, że teoretyczne rozkłady F_a i F_e (odpowiednio dla błędów *forward* i *backward*) są takie same jeżeli są one gaussowskie. Dla szeregów niegaussowskich powinniśmy jednak pamiętać o ich rozróżnieniu. Ponadto, w przypadku niegaussowskim, residua *backward* nie są niezależne ale jedynie nieskorelowane. Rozwiązaniem tego problemu byłoby replikowanie jedynie residuów *forward*, a następnie wykorzystanie związku pomiędzy tymi dwoma rodzajami błędów do generowania replikacji residuów *backward*.

Thombs i Schucany przyjmując naturalne założenia, że $\{X_j\}$ jest stacjonarnym procesem autoregresyjnym, dla którego $E(a_t) = 0$ oraz $E\{|a_t^k|\} < \infty$ (dla pewnego $k > 2$) pokazali słabą zbieżność bootstrapowego rozkładu warunkowego $G_b^*(\cdot)$ do

teoretycznego rozkładu warunkowego $G_{X_{t+h}|\underline{X}}(\cdot)$ gdy $n \rightarrow \infty$, a w konsekwencji również zgodność skonstruowanych bootstrapowych przedziałów predykcyjnych.

Cao (1997) zaproponował modyfikację algorytmu Thombsa i Schucany'ego, która pozwala na poprawę efektywności obliczeniowej. Zasadniczą ideą tej modyfikacji jest rezygnacja z konstrukcji replikacji *backward*, a w konsekwencji nie uwzględnianie zmienności związanej z estymacją parametrów. W odróżnieniu do algorytmu Thombsa i Schucany'ego nie są więc konstruowane replikacje bootstrapowe wyjściowych obserwacji $\underline{X}_1^*, \dots, \underline{X}_B^*$, które odzwierciedlają strukturę korelacyjną prognozowanego szeregu, ale jedynie replikowane są przyszłe (nieznane) wartości szeregu.

W tym przypadku replikacje przyszłych wartości otrzymujemy następująco:

$$X_{t+j}^* = \hat{\delta} + \hat{\phi}_1 X_{t+j-1} + \dots + \hat{\phi}_p X_{t+j-p} + \tilde{a}_{t+j}^*, \quad j = 1, 2, \dots, h,$$

gdzie $\{\tilde{a}_{t+j}^*\}$ są replikacjami bootstrapowymi residuów, wygenerowanymi z rozkładu o dystrybucji $\hat{F}_a$ (określonej na zbiorze scentrowanych i przeskalowanych residuów $\{\tilde{a}_i\}$, $i = p+1, \dots, t$, wyznaczonych jak w algorytmie Thombsa i Schucany'ego).

Cao udowodnił (w tym samym sensie jak Thombs i Schucany), że zaproponowana metoda konstrukcji przedziałów jest asymptotycznie zgodna. Przeprowadzone symulacje komputerowe wykazały istotną poprawę efektywności obliczeniowej w porównaniu z metodą Thombsa i Schucany'ego.

Dodatkowo, Cao otrzymał też obiecujące wyniki symulacyjne, w przypadku gdy zastąpimy dystrybuantę $\hat{F}_a$ jądrowym estymatorem $\hat{F}_{a,h}$ przy generowaniu replikacji residuów $\tilde{a}^*$. Procedura ta znana jest jako bootstrap wygładzony (ang. *smoothed bootstrap*). Bardziej szczegółowo problem ten będziemy omawiali w rozdziale 1.4.

Masarotto (1990) wykorzystał metodę bootstrap do konstrukcji przedziałów predykcyjnych dla procesu $AR(p)$ o skończonym, ale nieznanym p , w oparciu o optymalny predyktor liniowy. Przyjmując założenie, że $E(\epsilon_t^4) < \infty$ oraz rząd modelu jest estymowany poprzez minimalizację kryterium postaci:

$$D_n(k) = \sum_{i=1}^k \log [1 - \hat{\phi}_{ni}^2] + kf(n),$$

gdzie $f(n) \rightarrow 0$ gdy $n \rightarrow \infty$, Masarotto pokazał asymptotyczną zgodność skonstruowanych przedziałów predykcyjnych. Warto zauważyć, że postać D_n obejmuje m.in. takie kryteria jak: Akaike AIC, kryterium Hannana i Quinna oraz kryterium Schwartza.

Grigoletto (1998) wprowadził modyfikację metody Masarotto, która pozwala na istotną redukcję wariancji estymatorów kwantyli rozkładu błędu predykcji.

Kim (2001) rozważył konstrukcję przedziałów predykcyjnych dla wektorowych modeli autoregresyjnych znanego rzędu (modele VAR) z wykorzystaniem podejścia bootstrap-after-bootstrap.

Warto zauważyć, że metoda bootstrap-after-bootstrap, zastosowana po raz pierwszy przez Kiliana (1998a) do konstrukcji przedziałów ufności dla funkcji odpowiedzi impulsowej dla modeli VAR, ma wbudowaną procedurę korekcji obciążenia. Zastosowanie tej korekcji pozwala na uzyskanie poprawy w stosunku do standardowej

metody bootstrap, która — szczególnie w przypadku prób o małym rozmiarze — mogła produkować obciążone replikacje, z powodu obciążenia obecnego przy estymacji parametrów modelu autoregresyjnego.

Kim stosuje, podobnie jak Thombs i Schucany (1990), podejście warunkowe (warunkowanie ze względu na p - ostatnich obserwacji) wykorzystując do konstrukcji replikacji wsteczną (*backward*) reprezentację modelu autoregresyjnego. Aby uogólnić konstrukcję na przypadek wielowymiarowy zastosowana została metoda Bonferro-niego (Lütkepohl (1991)).

Przeprowadzono szereg symulacji komputerowych, które potwierdziły, że w większości przypadków zaproponowana przez Kima metoda konstrukcji przedziałów predykcyjnych zachowuje się znacznie lepiej niż metoda oparta na aproksymacji gaussowskiej oraz standardowa metoda bootstrap (Thombsa i Schucany'ego). Poprawa widoczna jest szczególnie w przypadku prób o małym i umiarkowanym rozmiarze oraz w przypadku modeli autoregresyjnych o pierwiastkach charakterystycznych bliższych jednostkowemu.

Niewątpliwie wadą zaproponowanego podejścia jest ograniczenie możliwości jego stosowania do modeli AR znanego rzędu. Ponadto, podobnie jak w przypadku algorytmu Thombsa i Schucany'ego, asymptotyczna poprawność algorytmu wymaga przyjęcia założenia o normalności innowacji, ponieważ residua *backward* są niezależne jedynie w przypadku kiedy residua *forward* są gaussowskie. Rozwiązaniem alternatywnym w przypadku niegaussowskich innowacji byłoby replikowanie residuów *forward*, o których z założenia wiemy że są niezależne, a następnie generowanie replikacji residuów *backward* wykorzystując związek pomiędzy tymi dwoma rodzajami residuów (Breidt i inni (1992)).

Zauważmy, że przedstawione powyżej metody konstrukcji bootstrapowych przedziałów predykcyjnych oparte były na założeniu o skończeniu wymiarowej postaci modelu, tzn. przyjmowaliśmy, że model $AR(p)$ jest dobrze dopasowany dla naszych danych.

Alonso, Peña i Romo (2002) rozważyli ostatnio uogólnienie metody Thombsa i Schucany'ego (1990) oraz Cao (1997) na przypadek ogólnej klasy procesów liniowych. Nieparametryczne warunkowe przedziały predykcyjne w tym przypadku skonstruowane są z wykorzystaniem metody *sieve bootstrap*. Przeprowadzone symulacje Monte Carlo pokazują, że zaproponowana metoda charakteryzuje się lepszymi wynikami dotyczącymi średniego pokrycia i długości przedziałów i stanowi alternatywę dla standardowej (gaussowskiej) procedury konstrukcji przedziałów predykcyjnych.

1.2.5 Równoczesne przedziały predykcyjne

Do tej pory, mówiąc o konstrukcji przedziałów predykcyjnych ograniczaliśmy się do przypadku przedziału predykcyjnego dla pojedynczej obserwacji i ustalonego horyzontu czasowego h . Bardziej złożonym problemem jest wyznaczenie równoczesnego obszaru (regionu) predykcyjnego dla określonego zbioru przyszłych obserwacji. W tym przypadku mogą wystąpić dwie sytuacje: prognoza przedziałowa ma być wyznaczona dla pojedynczej zmiennej w różnych horyzontach czasowych, tzn.

$\{X_{n+h}, h = 1, \dots, K\}$ albo też dla K zmiennych w tym samym (ustalonym) horyzoncie czasowym ($\underline{X}_{n+h} = X_{n+h,1}, \dots, X_{n+h,K}$).

Do konstrukcji równoczesnych przedziałów predykcyjnych możemy wykorzystać metodę Bonferroniego (Lütkepohl (1991)), która oparta jest na następującej nierówności probabilistycznej:

$$P\left(\bigcap_{i=1}^K E_i\right) \geq 1 - \sum_{i=1}^K P(E_i^c),$$

gdzie $E_i, i = 1, \dots, K$ są zdarzeniami: $\{X_{n+i} \in I(i)\}$ (odpowiednio $\{X_{n+h,i} \in I_i(h)\}$) oraz E_i^c jest dopełnieniem zdarzenia E_i .

Jeżeli skonstruujemy przedziały predykcyjne na poziomie ufności $[1 - \frac{\alpha}{K}] \cdot 100\%$ dla wszystkich wartości $\{X_{n+h}, h = 1, \dots, K\}$ (odpowiednio dla każdej z K składowych wektora $\underline{X}_{n+h}$) otrzymamy łączny obszar predykcyjny, który z prawdopodobieństwem co najmniej $(1 - \alpha)100\%$ zawierał będzie K zmiennych jednocześnie.

Dla gaussowskich szeregów czasowych możemy konstruować dokładne obszary predykcyjne wykorzystując odpowiednie kwantyle wielowymiarowego rozkładu normalnego.

Cheung, Wu i Chan (1998) przedstawili interesujące porównanie różnych metod konstrukcji równoczesnych obszarów predykcyjnych dla modeli ARIMA. Porównanie obejmuje metodę dokładną oraz dwie metody aproksymacyjne oparte na nierównościach Bonferroniego wyższego rzędu oraz nierównościach typu produktoowego. Przedstawiono też – ważne dla praktyków – argumenty przemawiające za potrzebą konstrukcji obszarów predykcyjnych szczególnie w takich obszarach zastosowań jak ekonomia i business.

Kim (2001) zaproponował ostatnio metodę konstrukcji bootstrapowych równoczesnych przedziałów predykcyjnych dla wektorowych modeli autoregresyjnych (VAR) wykorzystując nierówność Bonferroniego. Przeprowadzone badania symulacyjne potwierdziły efektywność zaproponowanej procedury.

1.2.6 Przedziały predykcyjne z wykorzystaniem metody *sieve bootstrap*

Wykorzystując metodę *sieve bootstrap* rozważymy teraz uogólnienie algorytmu konstrukcji bezwarunkowych przedziałów predykcyjnych (zaproponowanego m.in. przez Stine'a (1987), Masarotto(1990) i Grigoletto(1998) dla procesów $AR(p)$) na przypadek modeli liniowych reprezentowanych jako proces $AR(\infty)$.

Przedział predykcyjny będziemy konstruowali aproksymując nieznaną rozkład błędu predykcyjnego $\Delta_n(h)$ lub odpowiednio rozkład unormowanego błędu predykcyjnego $T_n(h) = \Delta_n(h)/\hat{\sigma}_n(h)$ dla estymatora optymalnego predyktora liniowego skonstruowanego na bazie obserwacji $X_1, \dots, X_n$ z uwzględnieniem aproksymacji $AR(p(n))$.

Główną ideą wykorzystania metody *sieve bootstrap* w kontekście prognozowania jest wygenerowanie replikacji $X_1^*, \dots, X_n^*$ na podstawie obserwowanego szeregu

$X_1, \dots, X_n$, a następnie rozszerzenie tej replikacji do przyszłej chwili $n+h$. Replikacje przyszłych wartości X_{n+h}^* możemy łatwo wyznaczyć wykorzystując aproksymację autoregresyjną ($AR(p(n))$):

$$X_{n+h}^* = \hat{\phi}_1^* X_{n+h-1}^* + \dots + \hat{\phi}_{p(n)}^* X_{n+h-p(n)}^* + \epsilon_{n+h}^*,$$

gdzie $\hat{\phi}_1^*, \dots, \hat{\phi}_{p(n)}^*$ – bootstrapowe replikacje estymatorów wyznaczone na bazie $X_1^*, \dots, X_n^*$.

Rozważymy teraz dwie metody konstrukcji przedziałów predykcyjnych

- **Hybrydowe przedziały predykcyjne**

W tym przypadku aproksymujemy nieznany rozkład błędu predykcji

$$\Delta_n(h) = X_{n+h} - \hat{P}_n X_{n+h}$$

odpowiednim rozkładem bootstrapowym

$$\Delta_n^*(h) := X_{n+h}^* - \hat{P}_n^* X_{n+h}^*,$$

gdzie:

$\hat{P}_n X_{n+h}$ – estymator optymalnego predyktora liniowego dla X_{n+h} wyznaczony na podstawie obserwacji $X_1, \dots, X_n$ z uwzględnieniem aproksymacji $AR(p(n))$,
 $\hat{P}_n^* X_{n+h}^*$ – optymalny predyktor liniowy dla X_{n+h}^* wyznaczony na podstawie obserwacji $X_1^*, \dots, X_n^*$.

Metoda ta określana jest w literaturze jako *hybrid bootstrap* (Shao i Tu (1995)). Oznaczając przez q_α^* i $q_{1-\alpha}^*$ odpowiednie kwantyle rozkładu $\Delta_n^*(h)$ otrzymujemy bootstrapowy przedział predykcyjny

$$I_B(h) = [\hat{P}_n X_{n+h} + q_\alpha^*, \hat{P}_n X_{n+h} + q_{1-\alpha}^*]. \quad (1.7)$$

W praktyce estymujemy q_α^* i $q_{1-\alpha}^*$ w oparciu o B replikacji (gdzie B odpowiednio duże).

- **Studentyzowane przedziały predykcyjne**

Innym algorytmem, który możemy wykorzystać do konstrukcji przedziałów predykcyjnych jest metoda studentyzacji (*bootstrap-t*). Nieznany rozkład statystyki

$$T_n(h) = \frac{X_{n+h} - \hat{P}_n X_{n+h}}{\hat{\sigma}_n(h)},$$

jest w tym przypadku estymowany rozkładem bootstrapowym

$$T_n^*(h) = \frac{X_{n+h}^* - \hat{P}_n^* X_{n+h}^*}{\hat{\sigma}_n^*(h)},$$

gdzie $\hat{\sigma}_n^2(h)$ oraz $\hat{\sigma}_n^{*2}(h)$ są odpowiednimi błędami średniokwadratowymi predykcji (PMSE).

Studentyzowany przedział predykcyjny jest ostatecznie postaci:

$$I_{B-t}(h) = [\hat{P}_n X_{n+h} + t_\alpha^* \hat{\sigma}_n(h), \hat{P}_n X_{n+h} + t_{1-\alpha}^* \hat{\sigma}_n(h)], \quad (1.8)$$

gdzie t_α^* and $t_{1-\alpha}^*$ są kwantylami $T_n^*(h)$, które możemy estymować jak poprzednio wykorzystując odpowiednio dużą liczbę replikacji bootstrapowych.

1.3 Rezultaty dotyczące zgodności

1.3.1 Określenie zgodności metody bootstrap

W rozdziale 1.2 zaproponowaliśmy wykorzystanie metody *sieve bootstrap* do aproksymacji nieznanego rozkładu błędu predykcji $\Delta_n(h)$ (lub odpowiednio unormowanego błędu predykcji $T_n(h)$). Uzasadnienie poprawności tej aproksymacji wiąże się najczęściej z koniecznością udowodnienia twierdzeń o zgodności. Przedstawimy teraz pokrótce ogólną koncepcję zgodności metody bootstrap.

Definicja 1.3.1 (Słaba i mocna zgodność rozkładów)

Niech ρ - oznacza metrykę na $\mathcal{F}_{\mathbb{R}^s} = \{\text{wszystkie rozkłady na } \mathbb{R}^s\}$.

Oznaczmy także:

$$H_n(x) = P(T_n \leq x),$$

$$H_n^*(x) = P^*(T_n^* \leq x),$$

gdzie „gwiazdka” — jak zwykle w przypadku metody bootstrap — oznacza rozkład warunkowy (pod warunkiem obserwacji $X_1, \dots, X_n$).

Wówczas:

- H_n^* jest zgodnym estymatorem H_n w metryce ρ (słabo ρ - zgodnym) jeżeli

$$\rho(H_n^*, H_n) \xrightarrow{P} 0 \quad \text{gdy } n \rightarrow \infty.$$

- H_n^* jest mocno zgodnym estymatorem H_n (mocno ρ - zgodnym) jeżeli

$$\rho(H_n^*, H_n) \rightarrow 0 \quad \text{p.w. } \text{ gdy } n \rightarrow \infty.$$

Do najczęściej stosowanych metryk ρ należą: metryka supremum ρ_∞ , metryka Lévy’ego ρ_L oraz metryka Mallowsa d_r .

Definicja 1.3.2 (Metryka supremum)

$$\rho_\infty(F, G) = \sup_s |F(s) - G(s)|$$

Definicja 1.3.3 (Metryka Lévy'ego)

$$\rho_L(F, G) := \inf\{\epsilon > 0 : F(x - \epsilon) - \epsilon \leq G(x) \leq F(x + \epsilon) + \epsilon, \quad x \in \mathbb{R}\}.$$

Definicja 1.3.4 (Metryka Mallowsa)

Niech $\mathcal{F}_{r,s} = \{G \in \mathcal{F}_{\mathbb{R}^s} : \int |x|^r dG(x) < \infty\}$. Dla dwóch rozkładów F oraz G z $\mathcal{F}_{r,s}$ określamy ich odległość Mallowsa jako:

$$d_r(F, G) = \inf_{\tau_{X,Y}} (E\|X - Y\|^r)^{(1/r)},$$

gdzie $\tau_{X,Y}$ jest zbiorem wszystkich możliwych rozkładów par (X, Y) , których rozkładami brzegowymi są odpowiednio F i G .

Informacje o wzajemnych relacjach pomiędzy różnymi rodzajami metryk probabilistycznych można znaleźć np. w pracy Gibbsa i Su (2002).

Stosunkowo niedawno Belyaev (1997) oraz Belyaev i Sjöstedt - de Luna (2000) zaproponowali inną definicję zgodności rozkładów określaną jako słaba aproksymacja ciągu rozkładów (*weakly approaching sequences*).

Definicja 1.3.5 (Słaba aproksymacja ciągu rozkładów)

Niech $\{X_n, Z_n\}$ oraz $\{Y_n\}$ będą dwoma ciągami zmiennych losowych, $X_n, Y_n \in \mathcal{R}$, $Z_n \in \mathcal{Z}_n$, gdzie X_n, Z_n są zdefiniowane na tej samej przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathcal{P})$. Wówczas ciąg rozkładów warunkowych $\{\mathcal{L}(X_n|Z_n)\}$ słabo aproksymuje ciąg $\{\mathcal{L}(Y_n)\}$ według prawdopodobieństwa, jeżeli dla dowolnej funkcji ciągłej i ograniczonej $h(\cdot) \in C_b(\mathcal{X})$:

$$E[h(X_n)|Z_n] - E[h(Y_n)] \xrightarrow{P} 0.$$

Będziemy dalej oznaczać:

$$\mathcal{L}(X_n|Z_n) \xleftrightarrow{wa(P)} \mathcal{L}(Y_n) \quad \text{gdy } n \rightarrow \infty.$$

Uwaga 1.3.1

W przypadku gdy mamy dwa ciągi rozkładów bezwarunkowych $\mathcal{L}(X_n)$ i $\mathcal{L}(Y_n)$ przyjmujemy następującą definicję:

$$\mathcal{L}(X_n) \xleftrightarrow{wa} \mathcal{L}(Y_n) \quad \text{gdy } n \rightarrow \infty$$

wtedy i tylko wtedy gdy:

$$\forall h(\cdot) \in C_b(\mathcal{X}) \quad E[h(X_n)] - E[h(Y_n)] \longrightarrow 0.$$

Definicję słabej aproksymacji ciągu rozkładów można także uogólnić na przypadek dwóch ciągów rozkładów warunkowych (Belyaev i Sjöstedt - de Luna (2000)).

Poniżej przedstawiamy najważniejsze rezultaty udowodnione przez Belyaeva i Sjöstedt - de Luna (2000), które podają związki nowo wprowadzonego pojęcia ze zbieżnością funkcji charakterystycznych oraz zbieżnością według rozkładu.

Zasadniczą rolę w cytowanych poniżej wynikach odgrywa założenie o ciasności ciągu $\{Y_n\}$.

Definicja 1.3.6 (Ciąg ciasny)

Ciąg zmiennych losowych $\{Y_n\}$ będziemy nazywali ciasnym (ang. tight) wtedy i tylko wtedy, gdy:

$$\forall \epsilon > 0 \exists \text{ zbiór zwarty } C_\epsilon \subset \mathcal{X}, \text{ taki że: } P(Y_n \notin C_\epsilon) < \epsilon \text{ dla } n = 1, 2, \dots$$

Ustalmy następujące oznaczenia dla funkcji charakterystycznych:

$$\begin{aligned} \phi_{X_n|Z_n}(t) &= E(e^{itX_n}|Z_n) \\ \phi_{Y_n}(t) &= E(e^{itY_n}) \end{aligned}$$

Twierdzenie 1.3.1 (Twierdzenie o ciągłości)

Niech $\{X_n, Z_n\}$ oraz $\{Y_n\}$ będą dwoma ciągami zmiennych losowych jak w Definicji 1.3.5. Ponadto, niech $\{Y_n\}$ będzie ciągiem ciasnym. Wówczas

$$\mathcal{L}(X_n|Z_n) \xrightarrow{wa(P)} \mathcal{L}(Y_n) \text{ gdy } n \rightarrow \infty,$$

wtedy i tylko wtedy, gdy dla dowolnego $t \in \mathbb{R}^k$

$$\phi_{X_n|Z_n}(t) - \phi_{Y_n}(t) \xrightarrow{P} 0 \text{ gdy } n \rightarrow \infty.$$

Lemat 1.3.1 (Podciągowa charakterystyka bliskości rozkładów)

Niech Y_n będzie ciągiem ciasnym. Wówczas następujące warunki są równoważne:

- (i) $\mathcal{L}(X_n) \xrightarrow{wa} \mathcal{L}(Y_n)$ gdy $n \rightarrow \infty$
- (ii) Dla dowolnego podciągu $\{n_k\}$, takiego że $\mathcal{L}(Y_{n_k}) \xrightarrow{w} \mathcal{L}(Y_0)$ dla pewnej zmiennej losowej Y_0 , zachodzi również $\mathcal{L}(X_{n_k}) \xrightarrow{w} \mathcal{L}(Y_0)$, gdy $k \rightarrow \infty$.
- (iii) Dla dowolnego podciągu $\{n_k\}$ istnieje podciąg $\{n_{k(l)}\}$ oraz zmienna losowa Y_0 , takie że: $\mathcal{L}(Y_{n_{k(l)}}) \xrightarrow{w} \mathcal{L}(Y_0)$ oraz $\mathcal{L}(X_{n_{k(l)}}) \xrightarrow{w} \mathcal{L}(Y_0)$, gdy $l \rightarrow \infty$.

Lemat 1.3.2 (Zbieżność w metryce Lévy'ego)

Niech $\{Y_n\}$ będzie ciągiem ciasnym.

Wówczas

$$\mathcal{L}(X_n) \xrightarrow{wa} \mathcal{L}(Y_n) \iff \rho_L(F_n, G_n) \rightarrow 0 \text{ gdy } n \rightarrow \infty,$$

gdzie ρ_L oznacza metrykę Lévy'ego natomiast F_n i G_n oznaczają dystrybuanty rozkładów odpowiednio X_n oraz Y_n .

Zauważmy, że aby zastosować Lematy 1.3.1 oraz 1.3.2 do dowodzenia zgodności metody bootstrap wymagane będzie ich uogólnienie na przypadek rozkładów warunkowych. Problem ten będziemy rozważali w rozdziale 1.3.2.

1.3.2 Fakty pomocnicze

Stochastyczne relacje $O_{P^*}(\cdot)$ i $o_{P^*}(\cdot)$

Dla badania asymptotyki aproksymacji bootstrapowej wygodnie nam będzie zdefiniować i określić podstawowe własności dla stochastycznych relacji $o_{P^*}(\cdot)$ i $O_{P^*}(\cdot)$, które są odpowiednikami relacji $o_P(\cdot)$ i $O_P(\cdot)$ (zdefiniowanymi np. u Serflinga (1991)).

Definicja 1.3.7 („ $O_{P^*}(1)$ według prawdopodobieństwa”)

Mówimy, że

$$X_n^* = O_{P^*}(1) \text{ wg prawdopodobieństwa}$$

wtedy i tylko wtedy, gdy:

$$\forall_{\epsilon, \eta > 0} \exists_{\delta_{\epsilon, \eta} > 0} P(P^*(|X_n^*| > \delta_{\epsilon, \eta}) > \epsilon) < \eta \text{ dla } n = 1, 2, \dots$$

Definicja 1.3.8 („ $o_{P^*}(1)$ według prawdopodobieństwa”)

Mówimy, że

$$X_n^* = o_{P^*}(1) \text{ według prawdopodobieństwa}$$

wtedy i tylko wtedy, gdy:

$$\forall_{\epsilon, \delta, \eta > 0} \exists_N P(P^*(|X_n^*| > \delta) > \epsilon) < \eta \text{ dla } n > N.$$

Zauważmy, że zgodnie z przyjętą definicją:

$$X_n^* = o_{P^*}(1) \text{ wg p-stwa} \iff X_n^* \xrightarrow{P^*} 0 \text{ wg p-stwa.}$$

Definicje 1.3.7 oraz 1.3.8 możemy łatwo uogólnić na przypadek dowolnego ciągu a_n , tzn. będziemy przyjmowali, że:

$$X_n^* = O_{P^*}(a_n) \text{ wg. p-stwa} \iff \frac{X_n^*}{a_n} = O_{P^*}(1) \text{ wg. p-stwa}$$

oraz

$$X_n^* = o_{P^*}(a_n) \text{ wg. p-stwa} \iff \frac{X_n^*}{a_n} = o_{P^*}(1) \text{ wg. p-stwa.}$$

Fakt 1.3.1 Przypuśćmy, że:

$$(i) \quad X_n^* = O_{P^*}(1) \text{ wg. prawdopodobieństwa,}$$

$$(ii) \quad Y_n^* \xrightarrow{P^*} 0 \text{ wg. prawdopodobieństwa.}$$

Wówczas

$$X_n^* Y_n^* \xrightarrow{P^*} 0 \text{ wg. prawdopodobieństwa.}$$

Symbolicznie możemy zapisać: $o_{P^*}(1) O_{P^*}(1) = o_{P^*}(1)$ według prawdopodobieństwa.

DOWÓD FAKTU 1.3.1

Mamy pokazać, że:

$$\forall_{\epsilon, \delta, \eta > 0} \exists_N P(P^*(|X_n^* Y_n^*| > \delta) > \epsilon) < \eta \quad \text{dla } n > N.$$

Z założenia $X_n^* = O_{P^*}(1)$ wg. prawdopodobieństwa, tzn.

$$\forall_{\epsilon', \eta' > 0} \exists_{\delta_{\epsilon', \eta'} > 0} P(P^*(|X_n^*| > \delta_{\epsilon', \eta'}) > \epsilon') < \eta' \quad \text{dla } n = 1, 2, \dots$$

Niech ϵ, δ oraz η będą dowolnymi, ustalonymi, dodatnimi liczbami rzeczywistymi. Przyjmijmy $\epsilon' = \epsilon/2$ oraz $\eta' = \eta/2$. Wówczas istnieje $\delta_{\epsilon', \eta'} = \delta_{\epsilon/2, \eta/2}$, takie że:

$$P(P^*(|X_n^*| > \delta_{\epsilon', \eta'}) > \epsilon/2) < \eta/2 \quad \text{dla } n = 1, 2, \dots$$

Zauważmy, że:

$$\begin{aligned} P^*(|X_n^* Y_n^*| > \delta) &= P^*(|X_n^* Y_n^*| > \delta, |X_n^*| > \delta_{\epsilon', \eta'}) + \\ &\quad + P^*(|X_n^* Y_n^*| > \delta, |X_n^*| \leq \delta_{\epsilon', \eta'}) \\ &\leq P^*(|X_n^*| > \delta_{\epsilon', \eta'}) + P^*(|Y_n^*| > \delta/\delta_{\epsilon', \eta'}). \end{aligned} \quad (1.9)$$

Z (1.9) wynika, że:

$$\begin{aligned} P(P^*(|X_n^* Y_n^*| > \delta) > \epsilon) &\leq \\ &\leq P(P^*(|X_n^*| > \delta_{\epsilon', \eta'}) + P^*(|Y_n^*| > \delta/\delta_{\epsilon', \eta'}) > \epsilon). \end{aligned} \quad (1.10)$$

Dalej możemy oszacować:

$$\begin{aligned} P(P^*(|X_n^*| > \delta_{\epsilon', \eta'}) + P^*(|Y_n^*| > \delta/\delta_{\epsilon', \eta'}) > \epsilon) &\leq \\ &\leq P(P^*(|X_n^*| > \delta_{\epsilon', \eta'}) > \epsilon/2 \text{ lub } P^*(|Y_n^*| > \delta/\delta_{\epsilon', \eta'}) > \epsilon/2) \\ &\leq P(P^*(|X_n^*| > \delta_{\epsilon', \eta'}) > \epsilon/2) + P(P^*(|Y_n^*| > \delta/\delta_{\epsilon', \eta'}) > \epsilon/2) \\ &\leq \eta/2 + P(P^*(|Y_n^*| > \delta/\delta_{\epsilon', \eta'}) > \epsilon/2). \end{aligned} \quad (1.11)$$

Na mocy założenia (ii) dla dostatecznie dużych n możemy zapisać:

$$P(P^*(|Y_n^*| > \delta/\delta_{\epsilon', \eta'}) > \epsilon/2) < \eta/2. \quad (1.12)$$

Ostatecznie zatem (1.10), (1.11) i (1.12) implikuje, że dla dowolnych $\epsilon, \delta, \eta > 0$

$$P(P^*(|X_n^* Y_n^*| > \delta) > \epsilon) < \eta/2 + \eta/2 = \eta$$

dla dostatecznie dużych n , co kończy dowód Faktu 1.3.1.

□

Fakt 1.3.2 *Niech*

$$(i) \quad X_n^* \xrightarrow{P^*} 0 \quad \text{wg. prawdopodobieństwa,}$$

$$(ii) \quad Y_n \xrightarrow{P} 0.$$

Wówczas

$$X_n^* + Y_n \xrightarrow{P^*} 0 \quad \text{wg. prawdopodobieństwa.}$$

Zapisując symbolicznie: $o_{P^*}(1) + o_P(1) = o_{P^*}(1)$ według prawdopodobieństwa.

DOWÓD FAKTU 1.3.2.

Mamy pokazać, że:

$$\forall_{\epsilon, \delta, \eta > 0} \exists_N P(P^*(|X_n^* + Y_n| > \delta) > \epsilon) < \eta \quad \text{dla } n > N.$$

Niech ϵ, δ, η będą dowolnymi, ustalonymi, dodatnimi liczbami rzeczywistymi. Za-
uważmy, że:

$$\begin{aligned} P(P^*(|X_n^* + Y_n| > \delta) > \epsilon) &\leq \\ &\leq P(P^*(|X_n^*| + |Y_n| > \delta) > \epsilon) \\ &\leq P(P^*(|X_n^*| > \delta/2 \vee |Y_n| > \delta/2) > \epsilon) \\ &\leq P(P^*(|X_n^*| > \delta/2) + P^*(|Y_n| > \delta/2) > \epsilon) \\ &\leq P(P^*(|X_n^*| > \delta/2) > \epsilon/2 \vee P^*(|Y_n| > \delta/2) > \epsilon/2) \\ &\leq P(P^*(|X_n^*| > \delta/2) > \epsilon/2) + P(P^*(|Y_n| > \delta/2) > \epsilon/2) \\ &= I_1 + I_2. \end{aligned}$$

Oszacujmy teraz kolejno składniki I_1 oraz I_2 .

Z założenia (i) dla dostatecznie dużego n mamy:

$$I_1 = P(P^*(|X_n^*| > \delta/2) > \epsilon/2) < \eta/2.$$

Stosując dla drugiego składnika nierówność Markowa otrzymujemy:

$$\begin{aligned} I_2 &= P(P^*(|Y_n| > \delta/2) > \epsilon/2) \\ &= P(E^* 1_{\{|Y_n| > \delta/2\}} > \epsilon/2) \\ &= P(1_{\{|Y_n| > \delta/2\}} > \epsilon/2) \\ &\leq \frac{P(|Y_n| > \delta/2)}{\epsilon/2}. \end{aligned}$$

Z założenia (ii) dla odpowiednio dużego n mamy, że

$$I_2 < \eta/2.$$

Otrzymaliśmy więc, że dla dowolnych $\epsilon, \delta, \eta > 0$:

$$P(P^*(|X_n^* + Y_n| > \delta) > \epsilon) < \eta/2 + \eta/2 = \eta$$

dla dostatecznie dużego n , co kończy uzasadnienie Faktu 1.3.2.

□

Fakt 1.3.3 Załóżmy, że:

$$(i) \quad |P^*(X_n^* \leq u) - P(X \leq u)| \xrightarrow{P} 0,$$

dla dowolnego u będącego punktem ciągłości dystrybuanty F_X ,

$$(ii) \quad Y_n^* \xrightarrow{P^*} 0 \quad \text{wg. prawdopodobieństwa,}$$

$$(iii) \quad V_n \xrightarrow{P} 0.$$

Wówczas

$$|P^*(X_n^* + Y_n^* \leq u) - P(X + V_n \leq u)| \xrightarrow{P} 0.$$

DOWÓD FAKTU 1.3.3

Dla dowolnego $\epsilon > 0$ możemy zapisać

$$\begin{aligned} P^*(X_n^* + Y_n^* \leq u) &= \\ &= P^*(X_n^* + Y_n^* \leq u, |Y_n^*| < \epsilon) + P^*(X_n^* + Y_n^* \leq u, |Y_n^*| \geq \epsilon) \\ &\leq P^*(X_n^* + Y_n^* \leq u, |Y_n^*| < \epsilon) + P^*(|Y_n^*| \geq \epsilon) \\ &\leq P^*(X_n^* \leq u + \epsilon) + P^*(|Y_n^*| \geq \epsilon). \end{aligned}$$

Z założenia (ii) otrzymujemy dla odpowiednio dużego n :

$$P^*(|Y_n^*| \geq \epsilon) \leq \kappa \quad \text{wg. p-stwa}, \quad (1.13)$$

gdzie $\kappa > 0$ dowolne.

Otrzymaliśmy zatem:

$$P^*(X_n^* + Y_n^* \leq u) \leq P^*(X_n^* \leq u + \epsilon) + \kappa \quad \text{wg. p-stwa.} \quad (1.14)$$

Z drugiej strony mamy:

$$\begin{aligned} P^*(X_n^* \leq u - \epsilon) &= \\ &= P^*(X_n^* \leq u - \epsilon, |Y_n^*| \geq \epsilon) + P^*(X_n^* \leq u - \epsilon, |Y_n^*| < \epsilon) \\ &\leq P^*(|Y_n^*| \geq \epsilon) + P^*(X_n^* \leq u - \epsilon, |Y_n^*| < \epsilon) \\ &\leq P^*(|Y_n^*| \geq \epsilon) + P^*(X_n^* \leq u - \epsilon, Y_n^* < \epsilon) \\ &\leq P^*(|Y_n^*| \geq \epsilon) + P^*(X_n^* + Y_n^* \leq u). \end{aligned}$$

Wykorzystując ponownie (1.13) otrzymujemy:

$$P^*(X_n^* + Y_n^* \leq u) \geq P^*(X_n^* \leq u - \epsilon) - \kappa \quad \text{wg. p-stwa} \quad (1.15)$$

Postępując analogicznie możemy pokazać, że:

$$P(X + V_n \leq u) \leq P(X \leq u + \epsilon) + \kappa. \quad (1.16)$$

oraz

$$P(X + V_n \leq u) \geq P(X \leq u - \epsilon) - \kappa. \quad (1.17)$$

Wybierając ϵ tak aby $u \pm \epsilon$ były punktami ciągłości dystrybuanty F_X otrzymujemy na mocy założenia (i), że:

$$|P^*(X_n^* \leq u \pm \epsilon) - P(X \leq u \pm \epsilon)| \leq \kappa \quad \text{wg. p-stwa.} \quad (1.18)$$

Wykorzystując teraz kolejno oszacowania (1.14), (1.17) oraz (1.18) otrzymujemy:

$$\begin{aligned} P^*(X_n^* + Y_n^* \leq u) - P(X + V_n \leq u) &\leq \\ &\leq P(X \leq u + \epsilon) - P(X \leq u - \epsilon) + 3\kappa \text{ wg. p-stwa.} \end{aligned} \quad (1.19)$$

Z drugiej strony stosując kolejno (1.15), (1.16) i (1.18) dostajemy:

$$\begin{aligned} P^*(X_n^* + Y_n^* \leq u) - P(X + V_n \leq u) &\geq \\ &\geq P(X \leq u - \epsilon) - P(X \leq u + \epsilon) - 3\kappa \text{ wg. p-stwa.} \end{aligned} \quad (1.20)$$

Ponieważ u jest punktem ciągłości dystrybuanty F_X oraz ϵ i κ były dowolnie małe, dodatnie, otrzymujemy ostatecznie:

$$P^*(X_n^* + Y_n^* \leq u) - P(X + V_n \leq u) \xrightarrow{P} 0.$$

co kończy dowód Lematu 1.3.3.

□

Lemat 1.3.3 (*Modyfikacja Lematu Słuckiego*)

Przypuśćmy, że:

- (i) $X_n^* \xrightarrow{d^*} X$ wg. prawdopodobieństwa,
- (ii) $Y_n^* \xrightarrow{P^*} c$ wg. prawdopodobieństwa,
(gdzie $c \neq 0$ jest pewną stałą).

Wówczas

$$\frac{X_n^*}{Y_n^*} \xrightarrow{d^*} \frac{X}{c} \text{ wg. prawdopodobieństwa.}$$

DOWÓD LEMATU 1.3.3

Dla wygody dowodu założmy, że $c > 0$. Dowód przypadku $c < 0$ przebiega analogicznie.

Niech u będzie punktem ciągłości dystrybuanty rozkładu zmiennej losowej X/c . Dla dowolnego ϵ , takiego że $0 < \epsilon < c$ możemy napisać:

$$\begin{aligned} P^*\left(\frac{X_n^*}{Y_n^*} \leq u\right) &= \\ &= P^*\left(\frac{X_n^*}{Y_n^*} \leq u, |Y_n^* - c| \leq \epsilon\right) + P^*\left(\frac{X_n^*}{Y_n^*} \leq u, |Y_n^* - c| > \epsilon\right). \end{aligned} \quad (1.21)$$

Bez utraty ogólności rozważań możemy przyjąć dalej, że $u > 0$. Otrzymujemy wówczas:

$$\begin{aligned} P^*\left(\frac{X_n^*}{c-\epsilon} \leq u, |Y_n^* - c| \leq \epsilon\right) &\leq P^*\left(\frac{X_n^*}{Y_n^*} \leq u, |Y_n^* - c| \leq \epsilon\right) \\ &\leq P^*\left(\frac{X_n^*}{c+\epsilon} \leq u, |Y_n^* - c| \leq \epsilon\right). \end{aligned} \quad (1.22)$$

Możemy przedstawić lewą i prawą stronę w oszacowaniu (1.22) jako:

$$\begin{aligned} P^*(X_n^* \leq u(c - \epsilon), |Y_n^* - c| \leq \epsilon) &= \\ &= P^*(X_n^* \leq u(c - \epsilon)) - P^*(X_n^* \leq u(c - \epsilon), |Y_n^* - c| > \epsilon) \end{aligned} \quad (1.23)$$

oraz

$$\begin{aligned} P^*(X_n^* \leq u(c + \epsilon), |Y_n^* - c| \leq \epsilon) &= \\ &= P^*(X_n^* \leq u(c + \epsilon)) - P^*(X_n^* \leq u(c + \epsilon), |Y_n^* - c| > \epsilon). \end{aligned} \quad (1.24)$$

Wykorzystując (1.21)–(1.24) otrzymujemy, że:

$$\begin{aligned} P^*(X_n^* < u(c - \epsilon)) - P^*(X_n^* < u(c - \epsilon), |Y_n^* - c| > \epsilon) &\leq \\ &\leq P^*\left(\frac{X_n^*}{Y_n^*} \leq u\right) - P^*\left(\frac{X_n^*}{Y_n^*} \leq u, |Y_n^* - c| > \epsilon\right) \leq \\ &\leq P^*(X_n^* \leq u(c + \epsilon)) - P^*(X_n^* \leq u(c + \epsilon), |Y_n^* - c| > \epsilon). \end{aligned} \quad (1.25)$$

Zauważmy teraz, że założenie (ii) ($Y_n^* \xrightarrow{P^*} c$ wg. p-stwa) implikuje, że:

$$\begin{aligned} P^*\left(\frac{X_n^*}{Y_n^*} \leq u, |Y_n^* - c| > \epsilon\right) &\leq P^*(|Y_n^* - c| > \epsilon) = o_P(1), \\ P^*(X_n^* \leq u(c - \epsilon), |Y_n^* - c| > \epsilon) &\leq P^*(|Y_n^* - c| > \epsilon) = o_P(1), \\ P^*(X_n^* \leq u(c + \epsilon), |Y_n^* - c| > \epsilon) &\leq P^*(|Y_n^* - c| > \epsilon) = o_P(1). \end{aligned} \quad (1.26)$$

Stosując (1.26) do oszacowania (1.25) otrzymujemy:

$$\begin{aligned} P^*(X_n^* \leq u(c - \epsilon)) - o_P(1) &\leq P^*\left(\frac{X_n^*}{Y_n^*} \leq u\right) - o_P(1) \\ &\leq P^*(X_n^* \leq u(c + \epsilon)) - o_P(1). \end{aligned} \quad (1.27)$$

Dobierając ϵ , tak aby $u(c - \epsilon)$ oraz $u(c + \epsilon)$ były punktami ciągłości dystrybuanty rozkładu zmiennej losowej X , otrzymujemy na mocy założenia (i), że:

$$P^*(X_n^* \leq u(c - \epsilon)) - o_P(1) \xrightarrow{P} P(X \leq u(c - \epsilon))$$

oraz

$$P^*(X_n^* \leq u(c + \epsilon)) - o_P(1) \xrightarrow{P} P(X \leq u(c + \epsilon)).$$

Z drugiej strony, ponieważ ϵ może być dowolnie małe i u było z założenia punktem ciągłości dystrybuanty zmiennej losowej X/c otrzymujemy ostatecznie, że:

$$P^*\left(\frac{X_n^*}{Y_n^*} \leq u\right) \xrightarrow{P} P\left(\frac{X}{c} \leq u\right),$$

co kończy dowód Lematu 1.3.3.

□

Charakteryzacja podciągowa bliskości rozkładów

Wykorzystując omówioną w poprzednim podrozdziale koncepcję słabej aproksymacji ciągu rozkładów (Belyaev (1997), Belyaev i Sjöstedt-de Luna (2000)) udowodnimy teraz rezultat będący uogólnieniem Lematu 1 (Belyaev i Sjöstedt-de Luna (2000)) na przypadek rozkładów warunkowych. Wynik ten będzie wykorzystany następnie w dowodzie Lematu 1.3.5 oraz w dowodzie twierdzenia o zgodności metody *bootstrap-t*.

Lemat 1.3.4 *Niech Y_n będzie ciągiem ciasnym. Wówczas następujące warunki są równoważne:*

- (i) $\mathcal{L}(X_n|Z_n) \xrightarrow{wa(P)} \mathcal{L}(Y_n)$ gdy $n \rightarrow \infty$
- (ii) Dla dowolnego podciągu $\{n_k\}$, takiego że $\mathcal{L}(Y_{n_k}) \xrightarrow{w} \mathcal{L}(Y_0)$ dla pewnej zmiennej losowej Y_0 , zachodzi również $\mathcal{L}(X_{n_k}|Z_{n_k}) \xrightarrow{w} \mathcal{L}(Y_0)$ według prawdopodobieństwa, gdy $k \rightarrow \infty$.
- (iii) Dla dowolnego podciągu $\{n_k\}$ istnieje podciąg $\{n_{k(l)}\}$ oraz zmienna losowa Y_0 , takie że: $\mathcal{L}(Y_{n_{k(l)}}) \xrightarrow{w} \mathcal{L}(Y_0)$ oraz $\mathcal{L}(X_{n_{k(l)}}|Z_{n_{k(l)}}) \xrightarrow{w} \mathcal{L}(Y_0)$ według prawdopodobieństwa gdy $l \rightarrow \infty$.

DOWÓD LEMATU 1.3.4.

„(i) $\Rightarrow$ (ii)”

Z założenia Y_n jest ciągiem ciasnym, a zatem z tw. o ciągłości (Tw. 2, Belyaev (1997), Belyaev i Sjöstedt-de Luna (2000)) otrzymujemy, że dla dowolnego $t \in R$:

$$\phi_{X_n}(t|Z_n) - \phi_{Y_n}(t) \xrightarrow{P} 0 \text{ gdy } n \rightarrow \infty, \quad (1.28)$$

gdzie $\phi_{X_n}(\cdot|Z_n)$, $\phi_{Y_n}(\cdot)$ oznaczają odpowiednie funkcje charakterystyczne.

Rozważmy teraz podciąg $\{n_k\}$, dla którego $\mathcal{L}(Y_{n_k}) \xrightarrow{w} \mathcal{L}(Y_0)$ gdy $k \rightarrow \infty$, dla pewnej zmiennej losowej Y_0 . Zbieżność słaba rozkładów implikuje zbieżność funkcji charakterystycznych dla dowolnego $t \in R$:

$$\phi_{Y_{n_k}}(t) \longrightarrow \phi_{Y_0}(t) \text{ gdy } k \rightarrow \infty. \quad (1.29)$$

Zauważmy, że z (1.28) wynika również:

$$\phi_{X_{n_k}}(t|Z_{n_k}) - \phi_{Y_{n_k}}(t) \xrightarrow{P} 0 \text{ gdy } k \rightarrow \infty. \quad (1.30)$$

(1.29) i (1.30) implikuje, że :

$$\phi_{X_{n_k}}(t|Z_{n_k}) \xrightarrow{P} \phi_{Y_0}(t) \text{ gdy } k \rightarrow \infty,$$

a to na mocy twierdzenia o ciągłości (Tw.2, Belyaev i Sjöstedt-de Luna (2000)) oznacza, że:

$$\mathcal{L}(X_{n_k}|Z_{n_k}) \xrightarrow{w} \mathcal{L}(Y_0) \text{ wg. prawdopodobieństwa, gdy } k \rightarrow \infty.$$

„(ii) $\Rightarrow$ (iii)”

Ponieważ z założenia $\{Y_n\}$ jest ciągiem ciasnym, to z tw. Prohorowa (Billingsley, (1968)) otrzymujemy, że dla dowolnego podciągu $\{n_k\}$ istnieje podciąg $\{n_k(l)\}$, taki że:

$$\mathcal{L}(Y_{n_k(l)}) \xrightarrow{w} \mathcal{L}(Y_0).$$

Z (ii) otrzymujemy, że również

$$\mathcal{L}(X_{n_k(l)}|Z_{n_k(l)}) \xrightarrow{w} \mathcal{L}(Y_0) \quad \text{wg. p-stwa gdy } l \rightarrow \infty.$$

„(iii) $\Rightarrow$ (i)”

Zgodnie z Definicją 2 (Belyaev i Sjöstedt-de Luna (2000)) mamy pokazać, że $E[h(X_n)|Z_n] - E[h(Y_n)] \xrightarrow{P} 0$ dla dowolnej funkcji $h(\cdot)$ ciągłej i ograniczonej.

Niech $h(\cdot) \in C_b(\mathcal{X})$. Oznaczmy:

$$A_n := E[h(X_n)|Z_n] - E[h(Y_n)].$$

Z (iii) otrzymujemy, że dla dowolnego podciągu $\{n_k\}$ istnieje podciąg $\{n_k(l)\}$, taki że:

$$E[h(X_{n_k(l)})|Z_{n_k(l)}] \xrightarrow{P} E[h(Y_0)] \quad (1.31)$$

oraz

$$E[h(Y_{n_k(l)})] \longrightarrow E[h(Y_0)] \quad (1.32)$$

gdy $l \rightarrow \infty$. Z (1.31) i (1.32) wynika natomiast, że

$$A_{n_k(l)} = E[h(X_{n_k(l)})|Z_{n_k(l)}] - E[h(Y_{n_k(l)})] \xrightarrow{P} 0 \quad \text{gdy } l \rightarrow \infty.$$

Z Tw. 1.6.1 (Serfling, (1991)) istnieje podciąg $\{n_k(l_m)\}$, taki że:

$$A_{n_k(l_m)} \longrightarrow 0 \quad \text{z p-stwem 1, gdy } m \rightarrow \infty.$$

Otrzymaliśmy więc, że dla dowolnego podciągu $\{n_k\}$ istnieje podciąg $\{n_k(l_m)\}$ taki że

$$A_{n_k(l_m)} \longrightarrow 0 \quad \text{z p-stwem 1, gdy } m \rightarrow \infty,$$

a to na mocy Tw.20.5 (Billingsley, (1987)) oznacza że $A_n \xrightarrow{P} 0$.

□

Zbieżność w metryce Lévy'ego

Kolejny rezultat związany jest z równoważnością słabej aproksymacji ciągu rozkładów ze zbieżnością w metryce Lévy'ego. Wynik ten stanowi uogólnienie Lematu 10 (Belyaev i Sjöstedt-de Luna (2000)) na przypadek rozkładów warunkowych.

Lemat 1.3.5 *Niech $\{Y_n\}$ będzie ciągiem ciasnym.*

Wówczas

$$\mathcal{L}(X_n|Z_n) \xrightarrow{wa(P)} \mathcal{L}(Y_n) \iff \rho_L(F_n^*, G_n) \xrightarrow{P} 0 \text{ gdy } n \rightarrow \infty,$$

gdzie ρ_L oznacza metrykę Lévy'ego natomiast F_n^* i G_n oznaczają dystrybuantę rozkładu warunkowego X_n pod warunkiem Z_n oraz odpowiednio dystrybuantę rozkładu Y_n .

DOWÓD LEMATU 1.3.5.

„ $\Rightarrow$ ”

Założmy, nie wprost, że

$$\rho_L(F_n^*, G_n) \not\xrightarrow{P} 0.$$

Wówczas istnieje podciąg $\{n_k\}$ oraz $\delta > 0$, dla których:

$$d(\rho_L(F_{n_k}^*, G_{n_k}), 0) > \delta, \tag{1.33}$$

gdzie $d(\cdot, \cdot)$ oznacza metrykę zbieżności według prawdopodobieństwa (np. Loève, str. 175, p.9).

Na mocy ciasności ciągu $\{Y_n\}$ otrzymujemy z tw. Prohorowa, że dla podciągu $\{n_k\}$ istnieje podciąg $\{n_k(l)\}$ taki że

$$\mathcal{L}(Y_{n_k(l)}) \xrightarrow{w} \mathcal{L}(Y_0) \text{ gdy } l \rightarrow \infty$$

dla pewnej zmiennej losowej Y_0 .

Z Lematu 1.3.4(ii) otrzymujemy, że również:

$$\mathcal{L}(X_{n_k(l)}|Z_{n_k(l)}) \xrightarrow{w} \mathcal{L}(Y_0) \text{ wg. prawdopodobieństwa, gdy } l \rightarrow \infty.$$

Oznaczając przez G_0 dystrybuantę rozkładu zmiennej Y_0 możemy teraz oszacować:

$$\rho_L(F_{n_k(l)}^*, G_{n_k(l)}) \leq \rho_L(F_{n_k(l)}^*, G_0) + \rho_L(G_{n_k(l)}, G_0).$$

Ponieważ zbieżność słaba jest równoważna zbieżności w metryce Lévy'ego otrzymujemy ostatecznie, że

$$\rho_L(F_{n_k(l)}^*, G_{n_k(l)}) \xrightarrow{P} 0$$

co jest sprzeczne z (1.33).

„ $\Leftarrow$ ”

Przypuśćmy, nie wprost, że istnieje taka funkcja ciągła i ograniczona $h_0(\cdot)$, dla której

$$E[h_0(X_n)|Z_n] - E[h_0(Y_n)] \not\xrightarrow{P} 0.$$

Wówczas istnieje podciąg $\{n_k\}$ oraz $\delta > 0$, dla których:

$$d(E[h_0(X_{n_k})|Z_{n_k}] - E[h_0(Y_{n_k})], 0) > \delta, \quad (1.34)$$

gdzie $d(\cdot, \cdot)$ oznacza, jak poprzednio, metrykę zbieżności według prawdopodobieństwa.

Z tw. Prohorowa, otrzymujemy że dla podciągu $\{n_k\}$ istnieje podciąg $\{n_k(l)\}$, taki że

$$\mathcal{L}(Y_{n_k(l)}) \xrightarrow{w} \mathcal{L}(Y_0) \quad \text{gdy } l \rightarrow \infty$$

dla pewnej zmiennej losowej Y_0 o dystrybucji G_0 . Z tego wynika w szczególności, że:

$$E[h_0(Y_{n_k(l)})] \rightarrow E[h_0(Y_0)] \quad \text{gdy } l \rightarrow \infty. \quad (1.35)$$

Ponieważ

$$\rho_L(F_{n_k(l)}^*, G_0) \leq \rho_L(F_{n_k(l)}^*, G_{n_k(l)}) + \rho_L(G_{n_k(l)}, G_0),$$

oraz zbieżność słaba jest równoważna zbieżności w metryce Lévy'ego otrzymujemy, że również $\rho_L(F_{n_k(l)}^*, G_0) \xrightarrow{P} 0$ gdy $l \rightarrow \infty$, a to implikuje że:

$$E[h_0(X_{n_k(l)})|Z_{n_k(l)}] \xrightarrow{P} E[h_0(Y_0)] \quad \text{gdy } l \rightarrow \infty. \quad (1.36)$$

Z (1.35) i (1.36) wynika następnie, że:

$$E[h_0(X_{n_k(l)})|Z_{n_k(l)}] - E[h_0(Y_{n_k(l)})] \xrightarrow{P} 0 \quad \text{gdy } l \rightarrow \infty,$$

co daje sprzeczność z (1.34).

□

Uogólnienie twierdzenia Craméra-Wolda

Sformułujemy teraz uogólnienie klasycznego twierdzenia Craméra-Wolda na przypadek ciągu wektorów losowych słabo aproksymujących się (ang. *weakly approaching*).

Tw. to pozwala sprowadzić problem uzasadnienia zgodności rozkładów wielowymiarowych do badania zgodności rozkładów kombinacji liniowej składowych. Wynik ten może być wykorzystany do uzasadnienia zgodności wielowymiarowych (równoczesnych) przedziałów predykcyjnych.

Twierdzenie 1.3.2 *Niech $\underline{X}_n = (X_{n1}, \dots, X_{nk})$ oraz $\underline{Y}_n = (Y_{n1}, \dots, Y_{nk})$ będą dwoma ciągami wektorów losowych. Załóżmy ponadto, że $\underline{Y}_n$ jest ciągiem ciasnym. Wówczas:*

$$\begin{aligned} \mathcal{L}(\underline{X}_n|Z_n) &\xleftrightarrow{wa(P)} \mathcal{L}(\underline{Y}_n) \iff \\ &\iff \mathcal{L}(\lambda_1 X_{n,1} + \dots + \lambda_k X_{n,k}|Z_n) \xleftrightarrow{wa(P)} \mathcal{L}(\lambda_1 Y_{n,1} + \dots + \lambda_k Y_{n,k}) \end{aligned}$$

dla dowolnego wektora $\underline{\lambda} = (\lambda_1, \dots, \lambda_k)$.

DOWÓD Tw. 1.3.2

„ $\Rightarrow$ ”

Ponieważ $\underline{Y}_n$ jest z założenia ciągiem ciasnym, to z tw. o ciągłości dla rozkładów słabo aproksymujących się (Tw. 2, Belyaev (1997), Belyaev i Sjöstedt (2000)), wiadomo że:

$$\phi_{\underline{X}_n|Z_n}(\underline{t}) - \phi_{\underline{Y}_n}(\underline{t}) \xrightarrow{P} 0 \text{ gdy } n \rightarrow \infty, \quad (1.37)$$

dla dowolnego $\underline{t} \in R^k$, gdzie

$$\phi_{\underline{X}_n|Z_n}(\underline{t}) = E(\exp(i \sum_{j=1}^k t_j X_{n,j}) | Z_n)$$

oraz

$$\phi_{\underline{Y}_n}(\underline{t}) = E(\exp(i \sum_{j=1}^k t_j Y_{n,j})).$$

Weźmy $\underline{t} = \underline{\lambda} \cdot s$, gdzie $\underline{\lambda} \in R^k$ i $s \in R$ dowolne.

Wówczas

$$\begin{aligned} \phi_{\underline{X}_n|Z_n}(\underline{t}) - \phi_{\underline{Y}_n}(\underline{t}) &= \\ &= E(\exp(is \sum_{j=1}^k \lambda_j X_{n,j}) | Z_n) - E(\exp(is \sum_{j=1}^k \lambda_j Y_{n,j})) \\ &= \phi_{\lambda_1 X_{n,1} + \dots + \lambda_k X_{n,k} | Z_n}(s) - \phi_{\lambda_1 Y_{n,1} + \dots + \lambda_k Y_{n,k}}(s). \end{aligned} \quad (1.38)$$

(1.37) i (1.38) oznaczają, że dla dowolnego $s \in R$:

$$\phi_{\lambda_1 X_{n,1} + \dots + \lambda_k X_{n,k} | Z_n}(s) - \phi_{\lambda_1 Y_{n,1} + \dots + \lambda_k Y_{n,k}}(s) \xrightarrow{P} 0.$$

Stosując ponownie tw. o ciągłości otrzymujemy, że

$$\mathcal{L}(\lambda_1 X_{n,1} + \dots + \lambda_k X_{n,k} | Z_n) \xrightarrow{wa(P)} \mathcal{L}(\lambda_1 Y_{n,1} + \dots + \lambda_k Y_{n,k}).$$

„ $\Leftarrow$ ”

Zauważmy, że na mocy cytowanego już twierdzenia o ciągłości:

$$\phi_{\lambda_1 X_{n,1} + \dots + \lambda_k X_{n,k} | Z_n}(s) - \phi_{\lambda_1 Y_{n,1} + \dots + \lambda_k Y_{n,k}}(s) \xrightarrow{P} 0$$

dla dowolnego $s \in R$.

W szczególności dla $s = 1$ otrzymujemy:

$$E(\exp(i \sum_{j=1}^k \lambda_j X_{n,j}) | Z_n) - E(\exp(i \sum_{j=1}^k \lambda_j Y_{n,j})) = \phi_{\underline{X}_n|Z_n}(\underline{\lambda}) - \phi_{\underline{Y}_n}(\underline{\lambda}) \xrightarrow{P} 0,$$

Ponieważ $\underline{\lambda} \in R^k$ było dowolne, to stosując jeszcze raz tw. o ciągłości (Tw. 2, Belyaev (1997), Belyaev i Sjöstedt (2000)), otrzymujemy, że:

$$\mathcal{L}(\underline{X}_n | Z_n) \xrightarrow{wa(P)} \mathcal{L}(\underline{Y}_n).$$

□

Twierdzenie o ciągłości

Kolejne twierdzenie ustanawia związek pomiędzy punktową zgodnością ciągów dystrybuant, a zbieżnością funkcji charakterystycznych.

Twierdzenie 1.3.3 *Jeżeli:*

(i) ciąg Y_n jest ciasny,

(ii) $P^*(X_n^* \leq u) - P(Y_n \leq u) \xrightarrow{P} 0$ dla $u \in E$,

gdzie: E jest zbiorem gęstym w R oraz miara Lebesgue'a zbioru E^c jest równa 0, $P^*(\cdot)$ oznacza jak poprzednio rozkład warunkowy (pod warunkiem obserwacji),

to

$$\phi_{X_n^*}(t) - \phi_{Y_n}(t) \xrightarrow{P} 0 \quad \text{dla dowolnego } t \in R.$$

Ponadto, zbieżność ta jest jednostajna względem t w każdym skończonym przedziale wartości t .

DOWÓD TWIERDZENIA 1.3.3

Przyjmijmy następujące oznaczenia:

$$\begin{aligned} F_n^*(\cdot) &= P^*(X_n^* \leq \cdot), \\ G_n(\cdot) &= P(Y_n \leq \cdot), \\ \phi_{X_n^*}(t) &= \int_{-\infty}^{\infty} e^{itx} dF_n^*(x), \\ \phi_{Y_n}(t) &= \int_{-\infty}^{\infty} e^{itx} dG_n(x). \end{aligned}$$

Z założenia (i), Y_n jest ciągiem ciasnym, tzn.

$$\forall \kappa > 0 \quad \exists \delta_\kappa > 0 \quad P(|Y_n| > \delta_\kappa) < \kappa \quad \text{dla } n = 1, 2, \dots$$

Niech $\kappa = \epsilon/12$, gdzie $\epsilon > 0$ - dowolne, ustalone. Wówczas istnieje δ_ϵ takie że

$$P(|Y_n| > \delta_\epsilon) < \epsilon/12 \quad \text{dla } n = 1, 2, \dots$$

Ponieważ E jest z założenia zbiorem gęstym, możemy wziąć $\delta'_\epsilon, \delta''_\epsilon \in E$ takie, że $\delta'_\epsilon < -\delta_\epsilon$ oraz $\delta''_\epsilon > \delta_\epsilon$. Wówczas otrzymujemy

$$P(Y_n < \delta'_\epsilon \vee Y_n > \delta''_\epsilon) \leq P(|Y_n| > \delta_\epsilon) < \epsilon/12.$$

Przedstawmy teraz:

$$\phi_{X_n^*}(t) = \int_{-\infty}^{\delta'_\epsilon} e^{itx} dF_n^*(x) + \int_{\delta'_\epsilon}^{\delta''_\epsilon} e^{itx} dF_n^*(x) + \int_{\delta''_\epsilon}^{\infty} e^{itx} dF_n^*(x) = I_{n1}^* + I_{n2}^* + I_{n3}^*,$$

oraz

$$\begin{aligned} \phi_{Y_n}(t) &= \int_{-\infty}^{\delta'_\epsilon} e^{itx} dG_n(x) + \int_{\delta'_\epsilon}^{\delta''_\epsilon} e^{itx} dG_n(x) + \int_{\delta''_\epsilon}^{\infty} e^{itx} dG_n(x) \\ &= I_{n1} + I_{n2} + I_{n3}. \end{aligned}$$

Rozważmy różnicę

$$I_{n2}^* - I_{n2} = \int_{\delta'_\epsilon}^{\delta''_\epsilon} e^{itx} dF_n^*(x) - \int_{\delta'_\epsilon}^{\delta''_\epsilon} e^{itx} dG_n(x).$$

Stosując wzór na całkowanie przez części otrzymujemy:

$$I_{n2}^* - I_{n2} = e^{itx}(F_n^*(x) - G_n(x)) \Big|_{\delta'_\epsilon}^{\delta''_\epsilon} - it \int_{\delta'_\epsilon}^{\delta''_\epsilon} (F_n^*(x) - G_n(x)) e^{itx} dx.$$

Dalej mamy

$$\begin{aligned} |I_{n2}^* - I_{n2}| &\leq \\ &\leq |F_n^*(\delta'_\epsilon) - G_n(\delta'_\epsilon)| + |F_n^*(\delta''_\epsilon) - G_n(\delta''_\epsilon)| + |t| \int_{\delta'_\epsilon}^{\delta''_\epsilon} |F_n^*(x) - G_n(x)| dx \end{aligned}$$

Ponieważ $\delta'_\epsilon, \delta''_\epsilon \in E$ to na mocy założenia (ii) otrzymujemy, że dla odpowiednio dużego n :

$$|F_n^*(\delta'_\epsilon) - G_n(\delta'_\epsilon)| < \epsilon/9, \quad \text{wg. p-stwa} \quad (1.39)$$

oraz

$$|F_n^*(\delta''_\epsilon) - G_n(\delta''_\epsilon)| < \epsilon/9, \quad \text{wg. p-stwa.} \quad (1.40)$$

Niech $T_1 < t < T_2$, gdzie T_1, T_2 - dowolne, ustalone. Oznaczmy $K = \max(|T_1|, |T_2|)$. Wówczas

$$|t| \int_{\delta'_\epsilon}^{\delta''_\epsilon} |F_n^*(x) - G_n(x)| dx \leq K \int_{\delta'_\epsilon}^{\delta''_\epsilon} |F_n^*(x) - G_n(x)| dx \quad (1.41)$$

Ponieważ $|F_n^*(x) - G_n(x)|$ jest jednostajnie ograniczone względem n oraz z założenia zbiega do 0 według prawdopodobieństwa (dla prawie wszystkich x), to z wersji tw. o zbieżności zmajoryzowanej dla ciągu procesów stochastycznych (Andersen i inni, (1993), Proposition II.5.3) otrzymujemy, że dla dostatecznie dużego n :

$$K \int_{\delta'_\epsilon}^{\delta''_\epsilon} |F_n^*(x) - G_n(x)| dx < \epsilon/9 \quad \text{wg. p-stwa.} \quad (1.42)$$

Ostatecznie zatem (1.39),(1.40),(1.41) i (1.42) implikuje:

$$|I_{n2}^* - I_{n2}| \leq \epsilon/9 + \epsilon/9 + \epsilon/9 = \epsilon/3 \quad \text{wg. p-stwa.}$$

Rozważmy teraz

$$I_{n1}^* - I_{n1} = \int_{-\infty}^{\delta'_\epsilon} e^{itx} dF_n^*(x) - \int_{-\infty}^{\delta'_\epsilon} e^{itx} dG_n(x).$$

Możemy oszacować

$$|I_{n1}^* - I_{n1}| \leq \int_{-\infty}^{\delta'_\epsilon} |e^{itx}| dF_n^*(x) + \int_{-\infty}^{\delta'_\epsilon} |e^{itx}| dG_n(x) = F_n^*(\delta'_\epsilon) + G_n(\delta'_\epsilon).$$

Z założenia o ciasności Y_n otrzymujemy, że:

$$\begin{aligned} G_n(\delta'_\epsilon) &= P(Y_n < \delta'_\epsilon) \leq P(Y_n < \delta'_\epsilon \vee Y_n > \delta''_\epsilon) \\ &\leq P(|Y_n| > \delta_\epsilon) \\ &< \epsilon/12. \end{aligned} \tag{1.43}$$

Z drugiej strony, ponieważ $\delta'_\epsilon \in E$ z założenia (ii) otrzymujemy również, że

$$F_n^*(\delta'_\epsilon) \leq |F_n^*(\delta'_\epsilon) - G_n(\delta'_\epsilon)| + G_n(\delta'_\epsilon) < \epsilon/6 \quad \text{wg. p-stwa.} \tag{1.44}$$

Otrzymane oszacowania (1.43) i (1.44) pozwalają nam napisać:

$$|I_{n1}^* - I_{n1}| \leq \epsilon/6 + \epsilon/6 = \epsilon/3 \quad \text{wg. p-stwa.}$$

Postępując analogicznie jak powyżej możemy pokazać, że:

$$|I_{n3}^* - I_{n3}| < \epsilon/3 \quad \text{wg. p-stwa.}$$

Ostatecznie zatem otrzymaliśmy, że dla dostatecznie dużego n i dowolnego $t \in (T1, T2)$:

$$|\phi_{X_n^*}(t) - \phi_{Y_n}(t)| \leq |I_{n1}^* - I_{n1}| + |I_{n2}^* - I_{n2}| + |I_{n3}^* - I_{n3}| \leq \epsilon \quad \text{wg. p-stwa,}$$

co kończy dowód Twierdzenia 1.3.3.

□

1.3.3 Zgodność metody *hybrid bootstrap*

Sformułujemy teraz lematy pomocnicze wykorzystane następnie w dowodzie zgodności metody *hybrid bootstrap*.

Lemat 1.3.6 (*Bühlmann (1997), Lemat 5.4*)

Niech spełnione będą założenia $A1'$ dla $s = 4$, $A2$ dla $r = 1$ oraz B dla $p(n) = o((n/\log(n))^{1/2})$. Wówczas dla dowolnego $t \in Z$:

$$\epsilon_t^* \xrightarrow{d^*} \epsilon_t \quad \text{według prawdopodobieństwa.}$$

Lemat 1.3.7 *Przypuśćmy, że spełnione są założenia $A2$ dla $r \geq 1$ ($r \in N$) oraz B dla $p(n) = o((n/\log(n))^{1/(2r+2)})$. Wówczas*

$$\sum_{j=p(n)+1}^{\infty} \phi_j X_{n+1-j} + \sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+1-j} = o_P(1).$$

DOWÓD LEMATU 1.3.7

Oznaczmy

$$\sum_{j=p(n)+1}^{\infty} \phi_j X_{n+1-j} + \sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+1-j} = S_1 + S_2.$$

Oszacujmy pierwszy składnik.

$$\begin{aligned} E|S_1| &= E|\sum_{j=p(n)+1}^{\infty} \phi_j X_{n+1-j}| \\ &\leq E(\sum_{j=p(n)+1}^{\infty} |\phi_j| |X_{n+1-j}|) \\ &\leq \sqrt{EX_t^2 \sum_{j=p(n)+1}^{\infty} |\phi_j|}. \end{aligned}$$

Z założenia A2, wynika że $\sum_{j=0}^{\infty} j^r |\psi_j| < \infty$, a stąd (np. Lemat 2.1, Bühlmann (1995)) otrzymujemy również, że $\sum_{j=0}^{\infty} j^r |\phi_j| < \infty$.

Dalej mamy

$$\sum_{j=p(n)+1}^{\infty} j^r |\phi_j| > \sum_{j=p(n)+1}^{\infty} (p(n))^r |\phi_j|,$$

a to oznacza, że $\sum_{j=p(n)+1}^{\infty} |\phi_j| = o(p(n)^{-r})$. Ostatecznie otrzymujemy więc $E|S_1| = o(p(n)^{-r})$.

Wykorzystując teraz nierówność Markowa dostajemy, że dla dowolnego $\epsilon > 0$

$$P(|S_1| > \epsilon) \leq E|S_1|/\epsilon,$$

czyli $S_1 = o_P(p(n)^{-r})$.

Oszacujmy teraz S_2 .

$$\begin{aligned} |S_2| &= |\sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+1-j}| \\ &= |\sum_{j=1}^{p(n)} (\phi_j - \phi_{j,n} + \phi_{j,n} - \hat{\phi}_j) X_{n+1-j}| \\ &\leq |\sum_{j=1}^{p(n)} (\phi_j - \phi_{j,n}) X_{n+1-j}| + |\sum_{j=1}^{p(n)} (\phi_{j,n} - \hat{\phi}_j) X_{n+1-j}| \\ &= I_1 + I_2, \end{aligned}$$

gdzie $\underline{\phi}_{p(n)} = (\phi_{1,n}, \dots, \phi_{p(n),n})'$ spełnia teoretyczne równania Yule'a-Walkera, tzn.

$$\Gamma_{p(n)} \underline{\phi}_{p(n)} = -\underline{\gamma}_{p(n)},$$

gdzie $\Gamma_{p(n)} = [\gamma(i-j)]_{i,j=1}^{p(n)}$, $\underline{\gamma}_{p(n)} = (\gamma(1), \dots, \gamma(p(n)))'$.

Do oszacowania I_1 wykorzystamy rozszerzoną nierówność Baxtera (Deistler i Hannan (1988), Tw. 6.6.12, str. 271)

$$\sum_{j=0}^{p(n)} |\phi_{j,n} - \phi_j| \leq c \sum_{j=p(n)+1}^{\infty} |\phi_j|,$$

gdzie c -stała zależna od modelu teoretycznego.

Możemy więc zapisać:

$$\begin{aligned} E I_1 &= E \left| \sum_{j=1}^{p(n)} (\phi_j - \phi_{j,n}) X_{n+1-j} \right| \\ &\leq \sum_{j=1}^{p(n)} E |X_{n+1-j}| \cdot |\phi_{j,n} - \phi_j| \\ &\leq \sqrt{E X_t^2} \cdot c \cdot \sum_{j=p(n)+1}^{\infty} |\phi_j| = o(p(n)^{-r}). \end{aligned}$$

Aby oszacować I_2 wykorzystamy nierówność Cauchy'ego oraz tw. 2.1 (Hannan i Kavalieris (1986)):

$$\begin{aligned} \max_{1 \leq j \leq p} |\hat{\phi}_j - \phi_{j,n}| &= O_{p.w.} \left(\sqrt{\log n/n} \right). \\ I_2 &= \left| \sum_{j=1}^{p(n)} (\phi_{j,n} - \hat{\phi}_j) X_{n+1-j} \right| \\ &\leq \sqrt{\sum_{j=1}^{p(n)} (\phi_{j,n} - \hat{\phi}_j)^2} \cdot \sqrt{\sum_{j=1}^{p(n)} X_{n+1-j}^2} \\ &\leq \sqrt{\sum_{j=1}^{p(n)} \left(\max_{1 \leq j \leq p(n)} |\hat{\phi}_j - \phi_{j,n}| \right)^2} \cdot \sqrt{\sum_{j=1}^{p(n)} X_{n+1-j}^2} \\ &= \sqrt{p(n)} \max_{1 \leq j \leq p(n)} |\hat{\phi}_j - \phi_{j,n}| \sqrt{\sum_{j=1}^{p(n)} X_{n+1-j}^2}. \end{aligned}$$

Oszacujmy teraz $\sqrt{\sum_{j=1}^{p(n)} X_{n+1-j}^2}$.

Niech $\epsilon > 0$ dowolne. Weźmy $\eta = \sqrt{E X_t^2}/\epsilon$. Wówczas z nierówności Czebyszewa otrzymujemy:

$$P \left(\frac{\sqrt{\sum_{j=1}^{p(n)} X_{n+1-j}^2}}{\sqrt{p(n)}} \geq \eta \right) \leq \epsilon,$$

a to oznacza że

$$\sqrt{\sum_{j=1}^{p(n)} X_{n+1-j}^2} = O_P(\sqrt{p(n)}).$$

Ostatecznie zatem

$$I_2 = O_P \left(p(n) (\log(n)/n)^{1/2} \right) = O_P \left((n/\log(n))^{\frac{-r}{2(r+1)}} \right).$$

co kończy dowód lematu 1.3.7. □

Lemat 1.3.8 (*Proposition 3.5, Alonso, Peña i Romo (2001)*)

Niech będą spełnione założenia: A1 dla $s = 4$, A2 dla $r > 1$ ($r \in N$) oraz B dla $p(n) = o((n/\log(n))^{\frac{1}{2r+2}})$. Ponadto, niech $\underline{\phi}_{p(n)} = (\phi_{1,n}, \dots, \phi_{p(n),n})$ spełniają teoretyczne równania Yule'a-Walkera, tzn.:

$$\Gamma_{p(n)} \underline{\phi}_{p(n)} = -\underline{\gamma}_{p(n)},$$

gdzie $\Gamma_{p(n)} = [\gamma(i-j)]_{i,j=1}^{p(n)}$, $\underline{\gamma}_{p(n)} = (\gamma(1), \dots, \gamma(p(n)))'$.

Wówczas

$$\max_{1 \leq j \leq p(n)} |\hat{\phi}_j^* - \phi_{j,n}| \xrightarrow{P^*} 0 \quad \text{według prawdopodobieństwa.}$$

Fakt 1.3.4 Przy założeniach jak w Lemacie 1.3.8

$$\max_{1 \leq j \leq p(n)} |\hat{\phi}_j^* - \phi_j| \xrightarrow{P^*} 0 \quad \text{według prawdopodobieństwa.}$$

DOWÓD FAKTU 1.3.4

Możemy oszacować

$$\max_{1 \leq j \leq p(n)} |\hat{\phi}_j^* - \phi_j| \leq \max_{1 \leq j \leq p(n)} |\hat{\phi}_j^* - \phi_{j,n}| + \max_{1 \leq j \leq p(n)} |\phi_{j,n} - \hat{\phi}_{j,n}| + \max_{1 \leq j \leq p(n)} |\hat{\phi}_{j,n} - \phi_j|.$$

W dowodzie Proposition 3.5 (Alonso, Peña i Romo (2001)) pokazane zostało, że:

$$\max_{1 \leq j \leq p(n)} |\hat{\phi}_j^* - \phi_{j,n}| = O_{P^*}((n/\log(n))^{-\frac{(r-1)}{(2r+2)}}) \quad \text{według prawdopodobieństwa.}$$

Ponadto, Hannan i Kavalieris ((1986), Tw. 2.1) udowodnili, że:

$$\max_{1 \leq j \leq p(n)} |\phi_{j,n} - \hat{\phi}_{j,n}| = O(\sqrt{\log(n)/n}) \quad p.w.$$

oraz

$$\max_{1 \leq j \leq p(n)} |\hat{\phi}_{j,n} - \phi_j| = O(\sqrt{\log(n)/n}) + o(p(n)^{-r}) \quad p.w.$$

Ostatecznie zatem

$$\max_{1 \leq j \leq p(n)} |\hat{\phi}_j^* - \phi_j| = O_{P^*}((n/\log(n))^{-\frac{(r-1)}{(2r+2)}}) \quad \text{według prawdopodobieństwa}$$

co kończy uzasadnienie Faktu 1.3.4.

□

Uwaga 1.3.2

Dla uproszczenia notacji będziemy oznaczali estymator optymalnego predyktora liniowego wyznaczonego na bazie $X_1, \dots, X_n$ z uwzględnieniem aproksymacji $AR(p(n))$ jako $\hat{X}_{n+h}$, zamiast stosowanego w rozdziale 1.2 oznaczenia $\hat{P}_n X_{n+h}$. Analogicznie zamiast $\hat{P}_n^* X_{n+h}^*$ przyjmujemy oznaczenie $\hat{X}_{n+h}^*$.

Kolejne dwa fakty podają zależności rekurencyjne dla błędu predykcji i odpowiednio bootstrapowego błędu predykcji.

Fakt 1.3.5 (Wzór rekurencyjny dla błędu predykcji)

Dla dowolnego, ustalonego horyzontu h ($h \in N$) oraz dla dostatecznie dużego n :

$$\begin{aligned} X_{n+h} - \hat{X}_{n+h} &= \sum_{j=1}^{h-1} \hat{\phi}_j (X_{n+h-j} - \hat{X}_{n+h-j}) + \sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+h-j} + \\ &\quad + \sum_{j=p(n)+1}^{\infty} \phi_j X_{n+h-j} + \epsilon_{n+h}. \end{aligned}$$

DOWÓD FAKTU 1.3.5

Wykorzystując reprezentację $AR(\infty)$ dla szeregu X_t możemy zapisać:

$$X_{n+h} = \sum_{j=1}^{\infty} \phi_j X_{n+h-j} + \epsilon_{n+h}. \quad (1.45)$$

Ponieważ dla dostatecznie dużego n mamy $p(n) > h$, możemy przedstawić predyktor $\hat{X}_{n+h}$ w zależności od prognoz dla wcześniejszych kroków $1, 2, \dots, h-1$ następująco:

$$\hat{X}_{n+h} = \hat{\phi}_1 \hat{X}_{n+h-1} + \dots + \hat{\phi}_{h-1} \hat{X}_{n+1} + \hat{\phi}_h X_n + \dots + \hat{\phi}_{p(n)} X_{n+h-p(n)}. \quad (1.46)$$

Odejmując (1.46) od (1.45) otrzymujemy już tezę Faktu 1.3.5.

□

Fakt 1.3.6 (Wzór rekurencyjny dla bootstrapowego błędu predykcji)

Dla dowolnego, ustalonego horyzontu h ($h \in N$) oraz dla dostatecznie dużego n :

$$X_{n+h}^* - \hat{X}_{n+h}^* = \sum_{j=1}^{h-1} \hat{\phi}_j^* (X_{n+h-j}^* - \hat{X}_{n+h-j}^*) + \epsilon_{n+h}^*.$$

DOWÓD FAKTU 1.3.6

Postępując analogicznie jak w dowodzie Faktu 1.3.5 możemy przedstawić:

$$X_{n+h}^* = \hat{\phi}_1^* X_{n+h-1}^* + \dots + \hat{\phi}_{p(n)}^* X_{n+h-p(n)}^* + \epsilon_{n+h}^*, \quad (1.47)$$

oraz predyktor dla X_{n+h}^* skonstruowany na bazie obserwacji $X_1^*, \dots, X_n^*$:

$$\hat{X}_{n+h}^* = \hat{\phi}_1^* \hat{X}_{n+h-1}^* + \dots + \hat{\phi}_{h-1}^* \hat{X}_{n+1}^* + \hat{\phi}_h^* X_n^* + \dots + \hat{\phi}_{p(n)}^* X_{n+h-p(n)}^*. \quad (1.48)$$

Odejmując teraz (1.48) od (1.47) otrzymujemy już wzór rekurencyjny na bootstrapowy błąd predykcji.

□

Fakt 1.3.7 (Stochastyczna ograniczoność błędu predykcji)

Dla dowolnego, ustalonego $h = 1, 2, \dots$

$$X_{n+h} - \hat{X}_{n+h} = O_P(1).$$

DOWÓD FAKTU 1.3.7

Do uzasadnienia Faktu 1.3.7 wykorzystamy rozumowanie indukcyjne.

($h = 1$)

Wykorzystując Lemat 1.3.7 otrzymujemy

$$\begin{aligned} X_{n+1} - \hat{X}_{n+1} &= \sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+1-j} + \sum_{j=p(n)+1}^{\infty} \phi_j X_{n+1-j} + \epsilon_{n+1} \\ &= o_P(1) + \epsilon_{n+1} \\ &= O_P(1). \end{aligned}$$

(h – dowolne)

Przyjmijmy założenie indukcyjne:

$$X_{n+k} - \hat{X}_{n+k} = O_P(1) \quad \text{dla } 1 \leq k \leq h-1.$$

Korzystając z zależności rekurencyjnej (Fakt 1.3.5) oraz Lematu 1.3.7 otrzymujemy dla horyzontu h :

$$\begin{aligned} X_{n+h} - \hat{X}_{n+h} &= \sum_{i=1}^{h-1} \hat{\phi}_i (X_{n+h-i} - \hat{X}_{n+h-i}) + \sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+h-j} + \\ &\quad + \sum_{j=p(n)+1}^{\infty} \phi_j X_{n+h-j} + \epsilon_{n+h} \\ &= \sum_{i=1}^{h-1} \hat{\phi}_i (X_{n+h-i} - \hat{X}_{n+h-i}) + o_P(1) + \epsilon_{n+h}. \end{aligned}$$

Wykorzystując dalej założenie indukcyjne dla $(X_{n+h-i} - \hat{X}_{n+h-i})$, gdzie $(1 \leq i \leq h-1)$ otrzymujemy ostatecznie, że:

$$X_{n+h} - \hat{X}_{n+h} = O_P(1).$$

□

Fakt 1.3.8 (*Stochastyczna ograniczoność bootstrapowego błędu predykcji*)

Dla dowolnego, ustalonego $h = 1, 2, \dots$

$$X_{n+h}^* - \hat{X}_{n+h}^* = O_{P^*}(1) \quad \text{według prawdopodobieństwa.}$$

DOWÓD FAKTU 1.3.8

Przyjmijmy, że $h = 1$. Niech $\epsilon > 0$ będzie dowolne, ustalone. Wówczas wykorzystując nierówności Czebyszewa oraz Lemat 5.3 (Bühlmann, (1997)) otrzymujemy:

$$\begin{aligned} P^*(|X_{n+1}^* - \hat{X}_{n+1}^*| > \delta_\epsilon) &= P^*(|\epsilon_{n+1}^*| > \delta_\epsilon) \\ &\leq E^*(\epsilon_{n+1}^*)^2 / (\delta_\epsilon^2) \\ &\leq E(\epsilon_1)^2 / (\delta_\epsilon^2) + o_P(1). \end{aligned}$$

Biorąc $\delta_\epsilon = \sqrt{E\epsilon_1^2}/\epsilon$ otrzymujemy ostatecznie, że:

$$\forall \epsilon > 0 \quad \exists \delta_\epsilon > 0 \quad P^*(|X_{n+1}^* - \hat{X}_{n+1}^*| > \delta_\epsilon) \leq \epsilon \quad \text{wg. p-stwa,}$$

a to oznacza, że:

$$X_{n+1}^* - \hat{X}_{n+1}^* = O_{P^*}(1) \quad \text{według prawdopodobieństwa.}$$

Uzasadnienie Faktu 1.3.8 dla dowolnego h otrzymamy przeprowadzając rozumowanie indukcyjne, podobnie jak w dowodzie Faktu 1.3.7.

□

Fakt 1.3.9 Dla dowolnego, ustalonego horyzontu h ($h \in N$) możemy przedstawić błąd predykcji jako:

$$X_{n+h} - \hat{X}_{n+h} = d_{1,h}(\underline{\phi}_{h-1}) \epsilon_{n+1} + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{n+h-1} + \epsilon_{n+h} + o_P(1),$$

gdzie $d_{1,h}, \dots, d_{h-1,h}$ - funkcje ciągłe oraz $\underline{\phi}_{h-1} = (\phi_1, \dots, \phi_{h-1})$.

DOWÓD FAKTU 1.3.9.

W uzasadnieniu wykorzystamy indukcję względem h .

($h=1$)

W tym przypadku mamy:

$$\begin{aligned} X_{n+1} - \hat{X}_{n+1} &= \epsilon_{n+1} + (\sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+1-j} + \sum_{j=p(n)+1}^{\infty} \phi_j X_{n+1-j}) \\ &= \epsilon_{n+1} + o_P(1) \end{aligned}$$

przy czym ostatnia równość jest konsekwencją Lematu 1.3.7.

(h -dowolne)

Przyjmijmy założenie indukcyjne, że dla horyzontu k , takiego że $1 \leq k \leq h-1$, możemy przedstawić błąd predykcji $X_{n+k} - \hat{X}_{n+k}$ jako:

$$\begin{aligned} X_{n+k} - \hat{X}_{n+k} &= \\ &= \tilde{d}_{1,k}(\phi_1, \dots, \phi_{k-1}) \epsilon_{n+1} + \dots + \tilde{d}_{k-1,k}(\phi_1, \dots, \phi_{k-1}) \epsilon_{n+k-1} + \epsilon_{n+k} + o_P(1) \end{aligned}$$

Dalej, korzystając kolejno z zależności rekurencyjnej (Fakt 1.3.5) oraz z Lematu 1.3.7 otrzymujemy:

$$\begin{aligned} X_{n+h} - \hat{X}_{n+h} &= \sum_{j=1}^{h-1} \hat{\phi}_j (X_{n+h-j} - \hat{X}_{n+h-j}) + \sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+h-j} + \\ &\quad + \sum_{j=p(n)+1}^{\infty} \phi_j X_{n+h-j} + \epsilon_{n+h} \\ &= \sum_{j=1}^{h-1} \hat{\phi}_j (X_{n+h-j} - \hat{X}_{n+h-j}) + \epsilon_{n+h} + o_P(1) \\ &= \sum_{j=1}^{h-1} \phi_j (X_{n+h-j} - \hat{X}_{n+h-j}) + \\ &\quad + \sum_{j=1}^{h-1} (\hat{\phi}_j - \phi_j) (X_{n+h-j} - \hat{X}_{n+h-j}) + \epsilon_{n+h} + o_P(1) \\ &= \sum_{j=1}^{h-1} \phi_j (X_{n+h-j} - \hat{X}_{n+h-j}) + \epsilon_{n+h} + o_P(1), \end{aligned}$$

przy czym ostatnia równość jest konsekwencją stochastycznej ograniczoności błędu predykcji (Fakt 1.3.7) oraz wynikającego z Tw. 2.1 Hannana i Kavalierisa (1986) faktu, że $\max_{1 \leq j \leq p(n)} |\phi_j - \hat{\phi}_j| = o_P(1)$.

Wykorzystując teraz założenie indukcyjne dla każdego ze składników $X_{n+h-j} - \hat{X}_{n+h-j}$, otrzymujemy ostatecznie tezę dla horyzontu h , co kończy dowód Faktu 1.3.9.

□

Fakt 1.3.10 Dla dowolnego, ustalonego horyzontu h ($h \in N$) możemy przedstawić bootstrapowy błąd predykcji jako:

$$\begin{aligned} X_{n+h}^* - \hat{X}_{n+h}^* &= \\ &= d_{1,h}(\underline{\phi}_{h-1}) \epsilon_{n+1}^* + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{n+h-1}^* + \epsilon_{n+h}^* + o_{P^*}(1), \end{aligned}$$

gdzie $d_{1,h}, \dots, d_{h-1,h}$ - funkcje ciągłe, (jak w reprezentacji błędu predykcji (Fakt 1.3.9)) oraz $\underline{\phi}_{h-1} = (\phi_1, \dots, \phi_{h-1})$.

DOWÓD FAKTU 1.3.10

Zauważmy najpierw, że dla $h = 1$ mamy:

$$X_{n+1}^* - \hat{X}_{n+1}^* = \epsilon_{n+1}^*.$$

Z kolei, korzystając z Faktu 1.3.6 dla dowolnego h możemy napisać:

$$\begin{aligned} X_{n+h}^* - \hat{X}_{n+h}^* &= \sum_{j=1}^{h-1} \hat{\phi}_j^* (X_{n+h-j}^* - \hat{X}_{n+h-j}^*) + \epsilon_{n+h}^* \\ &= \sum_{j=1}^{h-1} (\hat{\phi}_j^* - \phi_j) (X_{n+h-j}^* - \hat{X}_{n+h-j}^*) + \\ &\quad + \sum_{j=1}^{h-1} \phi_j (X_{n+h-j}^* - \hat{X}_{n+h-j}^*) + \epsilon_{n+h}^* \\ &= \sum_{j=1}^{h-1} \phi_j (X_{n+h-j}^* - \hat{X}_{n+h-j}^*) + \epsilon_{n+h}^* + o_{P^*}(1). \end{aligned}$$

W ostatniej równości wykorzystaliśmy Fakt 1.3.4, tzn. :

$$\max_{1 \leq j \leq p(n)} |\hat{\phi}_j^* - \phi_j| \xrightarrow{P^*} 0$$

oraz Fakty 1.3.8 i 1.3.1. Rozumowanie indukcyjne analogiczne jak w dowodzie Faktu 1.3.9 kończy uzasadnienie Faktu 1.3.10.

□

Przejdźmy teraz do sformułowania i udowodnienia zasadniczego rezultatu dotyczącego zgodności metody *hybrid bootstrap*.

Twierdzenie 1.3.4 (Zgodność metody *hybrid bootstrap*)

Przypuśćmy, że spełnione są założenia: A1 dla $s = 4$, A2 dla $r \geq 2$ oraz B dla $p(n) = o((\frac{n}{\log n})^{\frac{1}{2r+2}})$. Wówczas

$$|P^*(X_{n+h}^* - \hat{X}_{n+h}^* \leq u) - P(X_{n+h} - \hat{X}_{n+h} \leq u)| = o_P(1) \quad (1.49)$$

dla dowolnego u będącego punktem ciągłości dystrybucyjności zmiennej losowej $d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1} + \epsilon_h$, gdzie $\underline{\phi}_{h-1} = (\phi_1, \dots, \phi_{h-1})$.

DOWÓD TWIERDZENIA 1.3.4.

Dla przejrzystości dowodu, przedstawiamy go w dwóch krokach, dla $h = 1$ oraz dla dowolnego h .

KROK I: $h = 1$

Wykorzystując reprezentację $AR(\infty)$ dla procesu X_t mamy

$$X_{n+1} = \sum_{j=1}^{\infty} \phi_j X_{n+1-j} + \epsilon_{n+1}.$$

Estymator optymalnego predyktora jednokrokowego dla X_{n+1} skonstruowany na bazie aproksymacji procesem $AR(p(n))$ wyraża się jako:

$$\hat{X}_{n+1} = \hat{\phi}_1 X_n + \dots + \hat{\phi}_{p(n)} X_{n-p(n)+1}.$$

Analogicznie możemy przedstawić najlepszy predyktor liniowy dla X_{n+1}^* , skonstruowany na bazie obserwacji $X_1^*, \dots, X_n^*$, gdzie $X_{n+1}^* = \sum_{j=1}^{p(n)} \hat{\phi}_j^* X_{n+1-j}^* + \epsilon_{n+1}^*$:

$$\hat{X}_{n+1}^* = \hat{\phi}_1^* X_n^* + \dots + \hat{\phi}_{p(n)}^* X_{n+1-p(n)}^*.$$

Wykorzystując Lemat 1.3.7 oraz fakt, że ϵ_t jest ciągiem *i.i.d.* i ϵ_t^* jest warunkowo ciągiem *i.i.d.* otrzymujemy:

$$\begin{aligned} & |P^*(X_{n+1}^* - \hat{X}_{n+1}^* \leq u) - P(X_{n+1} - \hat{X}_{n+1} \leq u)| = \\ & = |P^*(\epsilon_{n+1}^* \leq u) - P(\epsilon_{n+1} \leq u - (\sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+1-j} + \\ & \quad + \sum_{j=p(n)+1}^{\infty} \phi_j X_{n+1-j}))| \\ & = |P^*(\epsilon_{n+1}^* \leq u) - P(\epsilon_{n+1} \leq u - o_P(1))| \\ & = |P^*(\epsilon_1^* \leq u) - P(\epsilon_1 \leq u - o_P(1))| \\ & \leq |P^*(\epsilon_1^* \leq u) - P(\epsilon_1 \leq u)| + |P(\epsilon_1 \leq u) - P(\epsilon_1 + o_P(1) \leq u)|. \end{aligned}$$

Stosując Lemat 5.4 (Bühlmann, (1997)) oraz Lemat Słuckiego otrzymujemy, że dla dowolnego u będącego punktem ciągłości rozkładu ϵ_t

$$|P^*(X_{n+1}^* - \hat{X}_{n+1}^* \leq u) - P(X_{n+1} - \hat{X}_{n+1} \leq u)| = o_P(1),$$

co kończy dowód KROKU I.

KROK II: h dowolne

Wykorzystując reprezentację błędu predykcji (Fakt 1.3.9) oraz bootstrapowego błędu predykcji (Fakt 1.3.10) możemy napisać:

$$\begin{aligned} & P^*(X_{n+h}^* - \hat{X}_{n+h}^* \leq u) - P(X_{n+h} - \hat{X}_{n+h} \leq u) = \\ & = P^*(d_{1,h}(\underline{\phi}_{h-1}) \epsilon_{n+1}^* + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{n+h-1}^* + \epsilon_{n+h}^* + o_{P^*}(1) \leq u) + \\ & \quad - P(d_{1,h}(\underline{\phi}_{h-1}) \epsilon_{n+1} + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{n+h-1} + \epsilon_{n+h} + o_P(1) \leq u) \quad (1.50) \\ & = P^*(d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1^* + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1}^* + \epsilon_h^* + o_{P^*}(1) \leq u) + \\ & \quad - P(d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1} + \epsilon_h + o_P(1) \leq u), \end{aligned}$$

gdzie $\underline{\phi}_{h-1} = (\phi_1, \dots, \phi_{h-1})$. W ostatniej równości wykorzystaliśmy fakt, że ϵ_t jest ciągiem *i.i.d.* oraz ϵ_t^* jest warunkowo ciągiem *i.i.d.*

Zauważmy, że na mocy Lematu 5.4 (Bühlmann, (1997)) oraz wspomnianych własności ϵ_t^* i ϵ_t otrzymujemy, że:

$$\begin{aligned} & P^*(d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1^* + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1}^* + \epsilon_h^* \leq u) + \\ & - P(d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1} + \epsilon_h \leq u) \xrightarrow{P} 0, \end{aligned} \quad (1.51)$$

dla dowolnego u będącego punktem ciągłości dystrybuanty zmiennej losowej $d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1} + \epsilon_h$.

Wykorzystując (1.50) oraz Fakt (1.3.3) otrzymujemy ostatecznie, że:

$$P^*(X_{n+h}^* - \hat{X}_{n+h}^* \leq u) - P(X_{n+h} - \hat{X}_{n+h} \leq u) \xrightarrow{P} 0$$

co należało dowieść.

□

Wnioski z Tw. 1.3.4

(i) Zbieżność funkcji charakterystycznych

Ponieważ zbiór punktów, dla których zachodzi Tw.1.3.4 jest zbiorem gęstym w R i jego dopełnienie ma miarę Lebesgue'a 0 oraz ciąg $\Delta_n(h) = X_{n+h} - \hat{X}_{n+h}$ jest ciągiem ciasnym (Fakt 1.3.7), to na mocy twierdzenia o ciągłości (Tw. 1.3.3) otrzymujemy także zbieżność funkcji charakterystycznych, tzn:

$$\phi_{\Delta_n^*(h)}^*(t) - \phi_{\Delta_n(h)}(t) \xrightarrow{P} 0$$

dla dowolnego $t \in R$, gdzie $\phi_{\Delta_n^*(h)}^*$ i $\phi_{\Delta_n(h)}$ są funkcjami charakterystycznymi odpowiadającymi $\Delta_n^*(h)$ i $\Delta_n(h)$.

(ii) Zgodność w sensie Belyaeva (słaba aproksymacja)

Wykorzystując ciasność ciągu $\Delta_n(h)$ oraz odpowiednią wersję twierdzenia o ciągłości (Tw. 1.3.1) otrzymujemy dalej, że zbieżność funkcji charakterystycznych (Wniosek (i)) jest równoważna słabej aproksymacji ciągu rozkładów, tzn. że:

$$\mathcal{L}^*(X_{n+h}^* - \hat{X}_{n+h}^*) \xleftrightarrow{wa(P)} \mathcal{L}(X_{n+h} - \hat{X}_{n+h}) \quad \text{gdy } n \rightarrow \infty.$$

(iii) Zgodność w metryce Lévy'ego

Wykorzystując Lemat 1.3.5 oraz jak poprzednio ciasność $\Delta_n(h)$ otrzymujemy równoważność słabej aproksymacji ciągu rozkładów (Wniosek (ii)) ze zbieżnością dystrybuant w sensie metryki Lévy'ego, tzn.:

$$\rho_L(F_{\Delta_n^*(h)}^*, F_{\Delta_n(h)}) \xrightarrow{P} 0 \quad \text{gdy } n \rightarrow \infty,$$

gdzie ρ_L oznacza metrykę Lévy'ego natomiast $F_{\Delta_n^*(h)}^*$ i $F_{\Delta_n(h)}$ oznaczają dystrybuantę rozkładu warunkowego $\Delta_n^*(h)$ oraz odpowiednio dystrybuantę rozkładu $\Delta_n(h)$.

1.3.4 Zgodność metody *bootstrap-t*

Sformułujmy na wstępie kilka podstawowych faktów związanych z postacią błędu predykcji (PE) oraz średniokwadratowego błędu predykcji (PMSE).

Jak poprzednio, niech X_t będzie procesem stacjonarnym, o którym zakładamy dodatkowo, że jest przyczynowy i odwracalny, tzn.:

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad \psi_0 = 1, \quad t \in Z, \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty, \quad (1.52)$$

oraz

$$\sum_{j=0}^{\infty} \phi_j X_{t-j} = \epsilon_t, \quad \phi_0 = 1, \quad t \in Z, \quad \sum_{j=0}^{\infty} \phi_j^2 < \infty, \quad (1.53)$$

gdzie $\{\epsilon_t\}_{t \in Z}$ jest białym szumem o $E[\epsilon_t] = 0$ i $E[\epsilon_t^2] = \sigma^2$.

Przyjmijmy następujące oznaczenia:

$\tilde{X}_{n+h}$ – najlepszy liniowy predyktor średniokwadratowy dla X_{n+h} skonstruowany pod warunkiem, że znamy całą historię $\{X_n, X_{n-1}, \dots\}$,

$\tilde{X}_{n+h}^n$ – najlepszy liniowy predyktor średniokwadratowy skonstruowany na bazie obserwacji $\{X_n, \dots, X_1\}$.

Fakt 1.3.11 (*Postać predyktora teoretycznego*)

$$\tilde{X}_{n+h} = \sum_{j=h}^{\infty} \psi_j \epsilon_{n+h-j}$$

DOWÓD FAKTU 1.3.11

Najlepszy predyktor średniokwadratowy dla X_{n+h} możemy przedstawić jako:

$$\tilde{X}_{n+h} = E(X_{n+h} | X_n, X_{n-1}, \dots).$$

Wykorzystując teraz fakt, że:

$$E(\epsilon_t | X_n, X_{n-1}, \dots) = \begin{cases} 0, & t > n \\ \epsilon_t, & t \leq n \end{cases}$$

otrzymujemy tezę Faktu 1.3.11.

□

Jako natychmiastowe wnioski z Faktu 1.3.11 otrzymujemy postać błędu predykcji (Fakt 1.3.12) oraz średniokwadratowego błędu predykcji (Fakt 1.3.13):

Fakt 1.3.12 (*Błąd predykcji (PE)*)

$$X_{n+h} - \tilde{X}_{n+h} = \sum_{j=0}^{h-1} \psi_j \epsilon_{n+h-j}$$

Fakt 1.3.13 (*Średniokwadratowy błąd predykcji (PMSE)*)

$$E(X_{n+h} - \tilde{X}_{n+h})^2 = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$$

DOWÓD FAKTU 1.3.13

Wykorzystując postać błędu predykcji (Fakt 1.3.12) otrzymujemy:

$$\begin{aligned} E(X_{n+h} - \tilde{X}_{n+h})^2 &= E(\sum_{j=0}^{h-1} \psi_j \epsilon_{n+h-j})^2 \\ &= E(\sum_{i=0}^{h-1} \sum_{j=0}^{h-1} \psi_i \psi_j \epsilon_{n+h-j} \epsilon_{n+h-j}) \\ &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2. \end{aligned}$$

W ostatniej równości wykorzystaliśmy fakt, że $\{\epsilon_t\}$ tworzą ciąg nieskorelowany o $E\epsilon_t = 0$ i $E\epsilon_t^2 = \sigma^2$.

□

Odnotujmy jeszcze równość obu predyktorów oraz średniokwadratowych błędów predykcji w szczególnym przypadku, kiedy X_t jest procesem $AR(p)$.

Fakt 1.3.14 (Predyktor i PMSE dla $AR(p)$)

Jeżeli X_t jest procesem $AR(p)$ oraz $n > p$, to:

$$\tilde{X}_{n+h}^n = \tilde{X}_{n+h},$$

a w następnym:

$$E(X_{n+h} - \tilde{X}_{n+h}^n)^2 = E(X_{n+h} - \tilde{X}_{n+h})^2.$$

Przypomnijmy, że dla prognoz opartych na aproksymacji procesem $AR(p(n))$ stosowane były następujące oznaczenia:

$\hat{X}_{n+h}$ – predyktor dla X_{n+h} wyznaczony na bazie modelu $AR(p(n))$ z estymowanymi współczynnikami $(\hat{\phi}_1, \dots, \hat{\phi}_{p(n)})$, na bazie obserwacji $\{X_1, \dots, X_n\}$,

$\hat{X}_{n+h}^*$ – predyktor dla obserwacji X_{n+h}^* wyznaczony na bazie modelu $AR(p(n))$ o współczynnikach $(\hat{\phi}_1^*, \dots, \hat{\phi}_{p(n)}^*)$ i na bazie obserwacji $\{X_1^*, \dots, X_n^*\}$.

Ustalmy również oznaczenia dla błędów predykcji (PE):

$$\begin{aligned} \Delta_n(h) &= X_{n+h} - \hat{X}_{n+h}, \\ \Delta_n^*(h) &= X_{n+h}^* - \hat{X}_{n+h}^*. \end{aligned}$$

Oznaczmy teraz średniokwadratowe błędy predykcji:

$\sigma^2(h)$ – średniokwadratowy błąd predykcji (PMSE) dla predyktora teoretycznego $\tilde{X}_{n+h}$ wyznaczonego na bazie modelu $AR(\infty)$,

$\tilde{\sigma}_n^2(h)$ – PMSE dla predyktora wyznaczonego na bazie modelu $AR(p(n))$ o znanych współczynnikach $\underline{\phi}_{p(n)} = (\phi_{1,n}, \dots, \phi_{p(n),n})$ spełniających teoretyczne równania Yule’a-Walkera, tzn.:

$$\Gamma_{p(n)} \underline{\phi}_{p(n)} = -\underline{\gamma}_{p(n)},$$

gdzie $\Gamma_{p(n)} = [\gamma(i-j)]_{i,j=1}^{p(n)}$, $\underline{\gamma}_{p(n)} = (\gamma(1), \dots, \gamma(p(n)))'$,

$\hat{\sigma}_n^{*2}(h)$ – PMSE dla predyktora bootstrapowego $\hat{X}_{n+h}^*$.

Ponadto, niech $\hat{\sigma}_n^2(h)$ oznacza oszacowanie średniokwadratowego błędu predykcji $\hat{\sigma}_n^2(h)$ uzyskane po zastąpieniu nieznanych współczynników modelu $AR(p(n))$ ich estymatorami $(\hat{\phi}_1, \dots, \hat{\phi}_{p(n)})$ otrzymanymi metodą Yule’a-Walkera.

Zauważmy, że Fakty 1.3.13 oraz 1.3.14 pozwalają nam zapisać:

$$\begin{aligned}\sigma^2(h) &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2, \\ \tilde{\sigma}_n^2(h) &= \sigma^2 \sum_{j=0}^{h-1} \psi_{j,n}^2, \\ \hat{\sigma}_n^2(h) &= \hat{\sigma}^2 \sum_{j=0}^{h-1} \hat{\psi}_j^2, \\ \hat{\sigma}_n^{*2}(h) &= \hat{\sigma}^{*2} \sum_{j=0}^{h-1} \hat{\psi}_j^{*2},\end{aligned}$$

gdzie:

$\{\psi_{j,n}\}_{j=0}^\infty$ – współczynniki rozwinięcia $MA(\infty)$ odpowiadające procesowi $AR(p(n))$ o współczynnikach $(\phi_{1,n}, \dots, \phi_{p(n),n})$,

$\{\hat{\psi}_j\}_{j=0}^\infty$ – współczynniki rozwinięcia $MA(\infty)$ dla procesu $AR(p(n))$ o estymowanych parametrach $(\hat{\phi}_1, \dots, \hat{\phi}_{p(n)})$,

$\{\hat{\psi}_j^*\}_{j=0}^\infty$ – współczynniki rozwinięcia $MA(\infty)$ dla procesu $AR(p(n))$ o replikowanych parametrach $(\hat{\phi}_1^*, \dots, \hat{\phi}_{p(n)}^*)$.

Sformułujmy teraz fakty pomocnicze, które następnie będą wykorzystane w dowodzie twierdzenia o zgodności metody studentyzowanej.

Lemat 1.3.9 *Niech spełnione będą założenia: A1 dla $s = 4$, A2 dla $r > 2$ ($r \in \mathbb{N}$) oraz B dla $p(n) = o((n/\log(n))^{\frac{1}{2r+2}})$. Wówczas:*

$$\max_{1 \leq u \leq p(n)} |\hat{\psi}_u^* - \psi_u| \xrightarrow{P^*} 0 \text{ wg. } p\text{-stwa.}$$

DOWÓD LEMATU 1.3.9.

Wykorzystując ideę dowodu zaproponowaną przez Bühlmann ((1995), Tw. 3.2) możemy napisać:

$$\sigma^2 \psi_u = \sum_{j=0}^{\infty} \phi_j \gamma(u+j) \implies \psi_u = \sigma^{-2} \sum_{j=0}^{\infty} \phi_j \gamma(u+j)$$

oraz

$$\hat{\sigma}^{*2} \hat{\psi}_u^* = \sum_{j=0}^{\infty} \hat{\phi}_j^* \hat{\gamma}^*(u+j) \implies \hat{\psi}_u^* = \hat{\sigma}^{*-2} \sum_{j=0}^{\infty} \hat{\phi}_j^* \hat{\gamma}^*(u+j).$$

Wówczas:

$$\begin{aligned}
|\hat{\psi}_u^* - \psi_u| &= \\
&= |\hat{\sigma}^{*-2} \sum_{j=0}^{\infty} \hat{\phi}_j^* \hat{\gamma}^*(u+j) - \sigma^{-2} \sum_{j=0}^{\infty} \phi_j \gamma(u+j)| \\
&= |\hat{\sigma}^{*-2} \sum_{j=0}^{p(n)} \hat{\phi}_j^* \hat{\gamma}^*(u+j) - \sigma^{-2} \sum_{j=0}^{\infty} \phi_j \gamma(u+j)| \\
&= |\hat{\sigma}^{*-2} \sum_{j=0}^{p(n)} (\hat{\phi}_j^* - \phi_j) \hat{\gamma}^*(u+j) + \\
&\quad + \hat{\sigma}^{*-2} \sum_{j=0}^{p(n)} \phi_j \hat{\gamma}^*(u+j) - \sigma^{-2} \sum_{j=0}^{\infty} \phi_j \gamma(u+j)| \\
&= |\hat{\sigma}^{*-2} \sum_{j=0}^{p(n)} (\hat{\phi}_j^* - \phi_j) \hat{\gamma}^*(u+j) + (\hat{\sigma}^{*-2} - \sigma^{-2}) \sum_{j=0}^{p(n)} \phi_j \hat{\gamma}^*(u+j) + \\
&\quad + \sigma^{-2} \sum_{j=0}^{p(n)} \phi_j (\hat{\gamma}^*(u+j) - \gamma(u+j)) - \sigma^{-2} \sum_{j=p(n)+1}^{\infty} \phi_j \gamma(u+j)| \\
&\leq |\hat{\sigma}^{*-2} \sum_{j=0}^{p(n)} (\hat{\phi}_j^* - \phi_j) \hat{\gamma}^*(u+j)| + |(\hat{\sigma}^{*-2} - \sigma^{-2}) \sum_{j=0}^{p(n)} \phi_j \hat{\gamma}^*(u+j)| + \\
&\quad + |\sigma^{-2} \sum_{j=0}^{p(n)} \phi_j (\hat{\gamma}^*(u+j) - \gamma(u+j))| + |\sigma^{-2} \sum_{j=p(n)+1}^{\infty} \phi_j \gamma(u+j)| \\
&= I_1 + I_2 + I_3 + I_4.
\end{aligned}$$

Oszacujemy teraz poszczególne składniki.

$$\begin{aligned}
(I_1) \quad I_1 &\leq \hat{\sigma}^{*-2} \sum_{j=0}^{p(n)} |\hat{\phi}_j^* - \phi_j| |\hat{\gamma}^*(u+j)| \\
&\leq \max_{0 \leq j \leq p(n)} |\hat{\phi}_j^* - \phi_j| \sum_{j=0}^{p(n)} |\hat{\gamma}^*(u+j)| \\
&\leq \max_{1 \leq j \leq p(n)} |\hat{\phi}_j^* - \phi_j| \cdot p(n)
\end{aligned}$$

Wykorzystując Fakt 1.3.4, tzn.

$$\max_{1 \leq j \leq p(n)} |\hat{\phi}_j^* - \phi_j| = O_{P^*}((n/\log n)^{-\frac{(r-1)}{2r+2}}) \quad \text{wg. p-stwa}$$

otrzymujemy:

$$\begin{aligned}
(I_1) \quad I_1 &= O_{P^*}((n/\log n)^{-\frac{(r-1)}{(2r+2)}} \cdot p(n)) = O_{P^*}((n/\log n)^{-\frac{(r-2)}{(2r+2)}}) \quad \text{wg. p-stwa} . \\
(I_2) \quad I_2 &\leq |\hat{\sigma}^{*-2} - \sigma^{-2}| \cdot |\sum_{j=0}^{p(n)} \phi_j \hat{\gamma}^*(u+j)| \\
&= |\hat{\sigma}^{*-2} - \sigma^{-2}| \cdot \hat{\sigma}^{*-2} \cdot |\sum_{j=0}^{p(n)} \phi_j \hat{\rho}^*(u+j)| \\
&\leq |\hat{\sigma}^{*-2} - \sigma^{-2}| \cdot \hat{\sigma}^{*-2} \cdot |\sum_{j=0}^{p(n)} \phi_j| \\
&\leq o_P(1) \cdot \sum_{j=0}^{\infty} |\phi_j|.
\end{aligned}$$

W powyższym oszacowaniu wykorzystaliśmy Lemat 5.3 (Bühlmann,(1997)).

Ponieważ $\sum_{j=0}^{\infty} |\phi_j| < \infty$ otrzymujemy ostatecznie, że:

$$\begin{aligned}
(I_2) \quad I_2 &= o_P(1). \\
(I_3) \quad I_3 &\leq \sigma^{-2} \sum_{j=0}^{p(n)} |\phi_j| \cdot |\hat{\gamma}^*(u+j) - \gamma(u+j)| \\
&\leq \sigma^{-2} \max_{0 \leq j \leq p(n)} |\hat{\gamma}^*(u+j) - \gamma(u+j)| \sum_{j=0}^{p(n)} |\phi_j| \\
&\leq \sigma^{-2} \max_{1 \leq u \leq p(n)} \max_{0 \leq j \leq p(n)} |\hat{\gamma}^*(u+j) - \gamma(u+j)| \sum_{j=0}^{\infty} |\phi_j| \\
&\leq \sigma^{-2} \max_{1 \leq k \leq 2p(n)} |\hat{\gamma}^*(k) - \gamma(k)| \sum_{j=0}^{\infty} |\phi_j|
\end{aligned}$$

Oznaczmy $\tilde{p}(n) = 2 \cdot p(n)$. Ponieważ $\sum_{j=0}^{\infty} |\phi_j| < \infty$, pozostaje do pokazania, że:

$$\max_{1 \leq k \leq \tilde{p}(n)} |\hat{\gamma}^*(k) - \gamma(k)| \xrightarrow{P^*} 0 \quad \text{wg. prawdopodobieństwa.}$$

Zauważmy, że dla $p(n) = o((n/\log n)^{\frac{1}{2r+2}})$ (założenie B) otrzymujemy, że również $\tilde{p}(n) = 2p(n) = o((n/\log n)^{\frac{1}{2r+2}})$. Dalej, wykorzystamy analogiczne oszacowanie jak w dowodzie Proposition 3.5 (Alonso, Peña i Romo, (2001)). W tym celu ustalmy następujące oznaczenia dla norm wektorowych:

$$\begin{aligned} \|\underline{x}\|_{\infty} &= \max_{1 \leq i \leq k} |x_i|, \\ \|\underline{x}\|_1 &= \sum_{i=1}^k |x_i|, \\ \|\underline{x}\|_2 &= \sqrt{\sum_{i=1}^k x_i^2}, \end{aligned}$$

gdzie $\underline{x} = (x_1, \dots, x_k)$.

Otrzymujemy wówczas:

$$\|\hat{\gamma}_{\tilde{p}(n)}^* - \gamma_{\tilde{p}(n)}\|_{\infty} \leq \|\hat{\gamma}_{\tilde{p}(n)}^* - \gamma_{\tilde{p}(n)}\|_1 \leq (\tilde{p}(n))^{1/2} \|\hat{\gamma}_{\tilde{p}(n)}^* - \gamma_{\tilde{p}(n)}\|_2,$$

gdzie $\hat{\gamma}_{\tilde{p}(n)}^* = (\hat{\gamma}^*(1), \dots, \hat{\gamma}^*(\tilde{p}(n)))$ i $\gamma_{\tilde{p}(n)} = (\gamma(1), \dots, \gamma(\tilde{p}(n)))$, a następnie:

$$(\tilde{p}(n))^{1/2} \|\hat{\gamma}_{\tilde{p}(n)}^* - \gamma_{\tilde{p}(n)}\|_2 = O_{P^*}((n/\log(n))^{-\frac{(r-1)}{(2r+2)}}) \quad \text{wg. p-stwa.}$$

Ostatecznie, pokazaliśmy więc, że:

$$I_3 = O_{P^*}((n/\log(n))^{-\frac{(r-1)}{(2r+2)}}) \quad \text{wg. p-stwa.}$$

$$\begin{aligned} (I_4) \quad I_4 &\leq \sigma^{-2} \sum_{j=p(n)+1}^{\infty} |\phi_j| |\gamma(u+j)| \\ &= \sum_{j=p(n)+1}^{\infty} |\phi_j| |\rho(u+j)| \\ &\leq \sum_{j=p(n)+1}^{\infty} |\phi_j| \end{aligned}$$

Założenie A2 implikuje dalej, że

$$I_4 = o(p(n)^{-r}).$$

□

Lemat 1.3.10 *Niech spełnione będą założenia: A2 dla $r \in \mathbb{N}$ ($r \geq 1$), B dla $p(n) = o((n/\log(n))^{\frac{1}{2r+2}})$. Wówczas:*

$$\hat{\sigma}_n^2(h) \xrightarrow{P} \sigma^2(h).$$

DOWÓD LEMATU 1.3.10.

Wiadomo, że średniokwadratowy błąd predykcji (PMSE) dla predyktora h -krokowego możemy przedstawić jako:

$$\sigma^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2,$$

gdzie σ^2 - wariancja białego szumu, natomiast $\{\psi_j\}_{j=0}^{\infty}$ - współczynniki rozwinięcia $MA(\infty)$ dla procesu X_t .

Podobnie możemy przedstawić:

$$\hat{\sigma}_n^2(h) = \hat{\sigma}^2 \sum_{j=0}^{h-1} \hat{\psi}_j^2,$$

gdzie $\hat{\sigma}^2$ - estymator wariancji białego szumu otrzymany metodą Yule'a-Walkera, $\{\hat{\psi}_j\}$ - współczynniki rozwinięcia $MA(\infty)$ dla modelu $AR(p(n))$ o estymowanych współczynnikach $(\hat{\phi}_1, \dots, \hat{\phi}_{p(n)})$.

Wykorzystaliśmy tutaj dobrze znany fakt (Brockwell, Davis (1987)), że model autoregresji o współczynnikach estymowanych metodą Yule'a-Walkera jest zawsze przyczynowy (ang. *causal*).

Z Twierdzenia 3.2 (Bühlmann, (1995)) wiadomo, że

$$\sup_{j \in N} |\hat{\psi}_j - \psi_j| = O((\log(n)/n)^{1/2}) + O(p(n)^{-r}) \quad p.w.$$

oraz

$$|\hat{\sigma}^2 - \sigma^2| = O((\log(n)/n)^{1/2}) + o(p(n)^{-r}) \quad p.w.$$

Stąd łatwo otrzymujemy już tezę Lematu 1.3.10.

□

Lemat 1.3.11 Przy założeniach jak w Lemacie 1.3.9

$$\hat{\sigma}_n^{*2}(h) \xrightarrow{P^*} \sigma^2(h) \quad wg. \ p\text{-stwa}.$$

DOWÓD LEMATU 1.3.11

Jak poprzednio możemy przedstawić

$$\sigma^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$$

oraz analogicznie:

$$\hat{\sigma}_n^{*2}(h) = \hat{\sigma}^{*2} \sum_{j=0}^{h-1} \hat{\psi}_j^{*2},$$

gdzie $\hat{\sigma}^{*2} = \text{Var}^* \epsilon_t^* = E^* \epsilon_t^{*2}$ oraz $\{\hat{\psi}_j^*\}_{j=0}^\infty$ - współczynniki rozwinięcia $MA(\infty)$ dla procesu $AR(p(n))$ o współczynnikach $\hat{\phi}_1^*, \dots, \hat{\phi}_{p(n)}^*$

Z Lematu 5.3 (Bühlmann, (1997)) otrzymujemy, że: $\hat{\sigma}^{*2} - \sigma^2 \xrightarrow{P} 0$. Z tego wynika, że również

$$\hat{\sigma}^{*2} - \sigma^2 \xrightarrow{P^*} 0 \quad \text{wg. p-stwa.}$$

Zastosowanie Lematu 1.3.9 kończy uzasadnienie Lematu 1.3.11.

□

Sformułujemy teraz podstawowy rezultat dotyczący zgodności metody studentyzowanej wykorzystując przedstawioną w rozdziale 1.3.1 koncepcję słabej aproksymacji ciągu rozkładów.

Twierdzenie 1.3.5 (Zgodność metody *bootstrap-t*)

Przypuśćmy, że spełnione są założenia $A1$ dla $s = 4$, $A2$ dla $r > 2$ ($r \in N$) oraz B dla $p(n) = o((n/\log(n))^{\frac{1}{2r+2}})$. Wówczas

$$\mathcal{L}^* \left(\frac{\Delta_n^*(h)}{\hat{\sigma}_n^*(h)} \right) \xleftrightarrow{wa(P)} \mathcal{L} \left(\frac{\Delta_n(h)}{\hat{\sigma}_n(h)} \right).$$

DOWÓD TWIERDZENIA 1.3.5.

W dowodzie wykorzystamy podciągową charakteryzację bliskości rozkładów (Lemat 1 (Belyaev i Sjöstedt-de Luna (2000))) i jego uogólnienie Lemat 1.3.4).

Na mocy Lematu 1.3.4

$$\mathcal{L}^* \left(\frac{\Delta_n^*(h)}{\hat{\sigma}_n^*(h)} \right) \xleftrightarrow{wa(P)} \mathcal{L} \left(\frac{\Delta_n(h)}{\hat{\sigma}_n(h)} \right) \iff$$

$$\iff \forall \text{podciągu } \{n_k\} \exists \text{podciąg } \{n_{k_l}\} \exists Y_0 - \text{zm. losowa, takie że :}$$

$$\mathcal{L}^* \left(\frac{\Delta_{n_{k_l}}^*(h)}{\hat{\sigma}_{n_{k_l}}^*(h)} \right) \xrightarrow{w} \mathcal{L}(Y_0) \quad \text{wg. p-stwa}$$

oraz

$$\mathcal{L} \left(\frac{\Delta_{n_{k_l}}(h)}{\hat{\sigma}_{n_{k_l}}(h)} \right) \xrightarrow{w} \mathcal{L}(Y_0).$$

Z twierdzenia o zgodności metody *hybrid bootstrap* (Tw. 1.3.4) wiemy, że

$$\mathcal{L}^*(\Delta_n^*(h)) \xleftrightarrow{wa(P)} \mathcal{L}(\Delta_n(h)).$$

Na mocy Lematu 1.3.4 oznacza to, że

$$\forall \{n_k\} \exists \{n_{k_l}\} \exists \Delta_0, \quad \text{takie że:}$$

$$\mathcal{L}^*(\Delta_{n_{k_l}}^*(h)) \xrightarrow{w} \mathcal{L}(\Delta_0) \quad \text{wg. p-stwa}$$

oraz

$$\mathcal{L}(\Delta_{n_{k_l}}(h)) \xrightarrow{w} \mathcal{L}(\Delta_0),$$

lub w innej notacji

$$\Delta_{n_{k_l}}^* \xrightarrow{d^*} \Delta_0 \quad \text{wg. p-stwa},$$

$$\Delta_{n_{k_l}} \xrightarrow{d} \Delta_0.$$

Z Lematu 1.3.10 i 1.3.11 otrzymujemy, że:

$$\hat{\sigma}_n^*(h) \xrightarrow{d^*} \sigma(h), \quad \text{wg. p-stwa}$$

oraz

$$\hat{\sigma}_n(h) \xrightarrow{d} \sigma(h).$$

Weźmy teraz dowolny podciąg $\{n_k\}$. Niech $\{n_{k_l}\}$ będzie podciągami ciągu $\{n_k\}$, dla którego:

$$\Delta_{n_{k_l}}^*(h) \xrightarrow{d^*} \Delta_0 \quad \text{wg. p-stwa}$$

oraz

$$\Delta_{n_{k_l}}(h) \xrightarrow{d} \Delta_0.$$

Wówczas otrzymujemy również, że:

$$\hat{\sigma}_{n_{k_l}}^*(h) \xrightarrow{d^*} \sigma(h), \quad \text{wg. p-stwa}$$

oraz

$$\hat{\sigma}_{n_{k_l}}(h) \xrightarrow{d} \sigma(h).$$

Ponieważ $\sigma(h) = \text{const.}$ (nie zależy od n), to korzystając z Lematu Śluckiego oraz jego uogólnionej wersji (Lemat 1.3.3) otrzymujemy:

$$\frac{\Delta_{n_{k_l}}^*(h)}{\hat{\sigma}_{n_{k_l}}^*(h)} \xrightarrow{d^*} \frac{\Delta_0}{\sigma(h)} \quad \text{wg. p-stwa.}$$

oraz

$$\frac{\Delta_{n_{k_l}}(h)}{\hat{\sigma}_{n_{k_l}}(h)} \xrightarrow{d} \frac{\Delta_0}{\sigma(h)},$$

a to na mocy Lematu 1.3.4 oznacza, że:

$$\mathcal{L}^*\left(\frac{\Delta_n^*(h)}{\hat{\sigma}_n^*(h)}\right) \xleftrightarrow{wa(P)} \mathcal{L}\left(\frac{\Delta_n(h)}{\hat{\sigma}_n(h)}\right).$$

□

1.3.5 Zgodność jednostajna

Na wstępie sformułujemy kilka faktów pomocniczych wykorzystanych następnie w dowodzie twierdzeń o jednostajnej zgodności metody *hybrid bootstrap* oraz *bootstrap*.

Definicja 1.3.9 (*Rodzina funkcji jednakowo ciągłych*)

Rodzinę $\mathcal{F}$ funkcji f określonych na R będziemy nazywali rodziną funkcji jednakowo ciągłych, jeżeli:

$$\forall \epsilon > 0, \quad \exists \delta > 0, \quad |x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon$$

dla dowolnej funkcji $f \in \mathcal{F}$.

Lemat 1.3.12 *Jeżeli ciąg dystrybuant $\{G_n\}$ jest ciągiem funkcji ciągłych, zbieżnym jednostajnie, to $\{G_n\}$ jest rodziną funkcji jednakowo ciągłych.*

DOWÓD LEMATU 1.3.12.

Mamy pokazać, że dla dowolnego $\epsilon > 0$ istnieje takie $\delta > 0$, że jeżeli $|x - y| < \delta$ to

$$|G_n(x) - G_n(y)| < \epsilon \quad \text{dla } n = 1, 2, \dots$$

Niech $\epsilon > 0$ będzie dowolne, ustalone. Ponieważ z założenia G_n jest ciągiem zbieżnym jednostajnie, to z kryterium Cauchy'ego wynika, że istnieje taka liczba naturalna N , że:

$$\sup_{x \in R} |G_n(x) - G_N(x)| < \epsilon/3 \quad \text{dla } n > N.$$

Zauważmy dalej, że dla dowolnego i , $G_i(\cdot)$ jako dystrybuantę ciągłą w każdym punkcie $x \in R$ jest również funkcją jednostajnie ciągłą. Możemy zatem dla $\{G_i, i = 1, \dots, N\}$ znaleźć takie $\delta > 0$, że jeżeli $|x - y| < \delta$ to

$$|G_i(x) - G_i(y)| < \epsilon/3 \quad \text{dla } 1 \leq i \leq N. \quad (1.54)$$

Z kolei dla $|x - y| < \delta$ i $n > N$ otrzymujemy:

$$\begin{aligned} |G_n(x) - G_n(y)| &\leq \\ &\leq |G_n(x) - G_N(x)| + |G_N(x) - G_N(y)| + |G_N(y) - G_n(y)| < \epsilon. \end{aligned} \quad (1.55)$$

Ostatecznie zatem (1.54) i (1.55) oznacza, że dla $|x - y| < \delta$

$$|G_n(x) - G_n(y)| < \epsilon \quad \text{dla } n = 1, 2, \dots$$

co kończy dowód Lematu 1.3.12.

□

Lemat 1.3.13 *Jeżeli $\{\epsilon_t\}$ jest ciągiem zmiennych losowych i.i.d. o rozkładzie ciągłym, to*

$$G_n(\cdot) = P(X_{n+h} - \hat{X}_{n+h} \leq \cdot)$$

jest ciągiem dystrybuant ciągłych, zbieżnym jednostajnie.

DOWÓD LEMATU 1.3.13.

Zauważmy najpierw, że przyjęte założenia o ciągu $\{\epsilon_t\}$ implikują ciągłość rozkładu X_t , a w konsekwencji również ciągłość rozkładu błędu predykcji $X_{n+h} - \hat{X}_{n+h}$.

Wykażemy teraz, że ciąg dystrybuant G_n odpowiadających błędom predykcji $X_{n+h} - \hat{X}_{n+h}$ jest jednostajnie zbieżny.

Na mocy Faktu 1.3.9 możemy dla dowolnego h przedstawić błąd predykcji jako:

$$X_{n+h} - \hat{X}_{n+h} = d_{1,h}(\underline{\phi}_{h-1}) \epsilon_{n+1} + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{n+h-1} + \epsilon_{n+h} + o_P(1),$$

gdzie $d_{1,h}, \dots, d_{h-1,h}$ – pewne funkcje ciągłe oraz $\underline{\phi}_{h-1} = (\phi_1, \dots, \phi_{h-1})$.

Z przyjętego założenia o niezależności i jednakowym rozkładzie ciągu $\{\epsilon_t\}$, $t = 1, 2, \dots$ otrzymujemy, że:

$$X_{n+h} - \hat{X}_{n+h} \xrightarrow{d} d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1} + \epsilon_h.$$

Założenie o ciągłości rozkładu ϵ_t implikuje również ciągłość rozkładu granicznego $d_{1,h}(\underline{\phi}_{h-1}) \epsilon_{t+1} + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{t+h-1} + \epsilon_{t+h}$, a to na mocy tw. Pólya’ego oznacza jednostajną zbieżność ciągu dystrybuant G_n .

□

Lemat 1.3.14 *Jeżeli $\{\epsilon_t\}$ jest ciągiem zmiennych losowych i.i.d. o rozkładzie ciągłym, to*

$$H_n(\cdot) = P\left(\frac{X_{n+h} - \hat{X}_{n+h}}{\hat{\sigma}_n(h)} \leq \cdot\right)$$

jest ciągiem dystrybuant ciągłych, zbieżnym jednostajnie.

DOWÓD LEMATU 1.3.14.

Dowód przebiega analogicznie jak dla Lematu 1.3.13. Wykorzystujemy dodatkowo Fakt 1.3.10, tzn. że:

$$\hat{\sigma}_n^2(h) \xrightarrow{P} \sigma^2(h).$$

□

Twierdzenie 1.3.6 *(Uogólnienie tw. Polya)*

Jeżeli $F_n^ \Rightarrow F$ według prawdopodobieństwa, oraz F jest funkcją ciągłą (na R), to:*

$$\sup_{t \in R} |F_n^*(t) - F(t)| \xrightarrow{P} 0$$

DOWÓD TWIERDZENIA 1.3.6

Niech $r_n := \rho_L(F_n^*, F)$ oznacza odległość Lévy'ego dla dystrybuant F_n^* i F .

Z założenia $F_n^* \Rightarrow F$ według prawdopodobieństwa, co implikuje, że $r_n \xrightarrow{P} 0$.

Na mocy definicji metryki Lévy'ego mamy:

$$F(x - r_n) - r_n \leq F_n^*(x) \leq F(x + r_n) + r_n \quad \forall x \in R.$$

Otrzymujemy dalej, że:

$$(F(x - r_n) - F(x)) - r_n \leq F_n^*(x) - F(x) \leq (F(x + r_n) - F(x)) + r_n \quad \forall x \in R.$$

Stąd wynika, że:

$$\sup_x (F(x - r_n) - F(x)) - r_n \leq \sup_x (F_n^*(x) - F(x)) \leq \sup_x (F(x + r_n) - F(x)) + r_n.$$

Ponieważ F jest z założenia funkcją ciągłą na R , to jako dystrybunata jest również funkcją jednostajnie ciągłą, co oznacza, że:

$$\forall \epsilon > 0 \quad \exists \delta > 0 \quad \sup_{x \in R} |F(x) - F(x - h)| < \epsilon \quad \text{dla } |h| < \delta.$$

Otrzymujemy stąd, że:

$$P(|r_n| < \delta) \leq P(\sup_x |F(x) - F(x - r_n)| < \epsilon),$$

a to, ponieważ ϵ jest dowolne i $r_n \xrightarrow{P} 0$ implikuje, że

$$\sup_x |F(x) - F(x - r_n)| \xrightarrow{P} 0.$$

Przeprowadzając analogiczne rozumowanie można pokazać, że również

$$\sup_x |F(x + r_n) - F(x)| \xrightarrow{P} 0.$$

Ostatecznie zatem otrzymujemy, że:

$$\sup_x (F_n^*(x) - F(x)) \xrightarrow{P} 0,$$

co kończy dowód Twierdzenia 1.3.6.

□

Belyaev i Sjöstedt-de Luna (2000) pokazali, że przy umiarkowanych założeniach otrzymujemy jednostajną zgodność dystrybuant dla ciągów zmiennych losowych słabo aproksymujących się według prawdopodobieństwa.

Lemat 1.3.15 (*Lemat 9, Belyaev i Sjöstedt-de Luna, (2000)*)

Niech $\{X_n, Z_n\}$ oraz Y_n będą dwoma ciągami zmiennych losowych (jak w Definicji 1.3.5). Niech $\{G_n\}$ będzie dystrybuantą rozkładu $\{Y_n\}$ natomiast $F_n(s|Z_n)$ dystrybuantą rozkładu warunkowego X_n pod warunkiem Z_n . Jeżeli:

- (i) $\{Y_n\}$ jest ciasny,
- (ii) $\{G_n\}$ jest ciągiem funkcji jednakowo ciągłych,
- (iii) $\mathcal{L}(X_n | Z_n) \xrightarrow{w_a(P)} \mathcal{L}(Y_n)$

to

$$\sup_{s \in R} |F_n(s | Z_n) - G_n(s)| \xrightarrow{P} 0 \quad \text{gdy } n \rightarrow \infty.$$

Sformułujmy teraz główne rezultaty dotyczące jednostajnej zgodności metody *hybrid bootstrap* oraz *bootstrap-t*.

Twierdzenie 1.3.7 (Jednostajna zgodność metody *hybrid bootstrap*)

Przypuśćmy, że spełnione są założenia: A1 dla $s = 4$, A2 dla $r \geq 2$ oraz B dla $p(n) = o((\frac{n}{\log n})^{\frac{1}{2r+2}})$. Dodatkowo założymy, że $\{\epsilon_t\}$ jest ciągiem zmiennych losowych i.i.d. o rozkładzie ciągłym. Wówczas

$$\sup_{x \in R} \left| P^* (X_{n+h}^* - \hat{X}_{n+h}^* \leq x) - P (X_{n+h} - \hat{X}_{n+h} \leq x) \right| \xrightarrow{P} 0.$$

DOWÓD TWIERDZENIA 1.3.7

Zauważmy najpierw, że przyjęte założenia implikują punktową zgodność metody *hybrid-bootstrap* (Tw. 1.3.4), a w konsekwencji także zgodność w sensie słabej aproksymacji rozkładów, tzn.:

$$\mathcal{L}^*(X_{n+h}^* - \hat{X}_{n+h}^*) \xrightarrow{w_a(P)} \mathcal{L}(X_{n+h} - \hat{X}_{n+h}).$$

Ponieważ $\{\epsilon_t\}$ jest z założenia ciągiem i.i.d. o rozkładzie ciągłym, to na mocy Lematu 1.3.13 otrzymujemy, że ciąg dystrybuant $\{G_n\}$ odpowiadających błędom predykcji $\Delta_n(h) = X_{n+h} - \hat{X}_{n+h}$ jest ciągiem dystrybuant ciągłych, zbieżnym jednostajnie. Dalej, korzystając z Lematu 1.3.12 otrzymujemy, że $\{G_n\}$ jest rodziną funkcji jednakowo ciągłych.

W rozdziale 1.3.3 pokazaliśmy, że ciąg rozkładów błędów predykcji $\Delta_n(h)$ jest ciągiem ciasnym.

Ostatecznie, stosując Lemat 1.3.15 otrzymujemy już jednostajną zgodność metody *hybrid bootstrap*.

□

Analogicznie możemy sformułować i udowodnić twierdzenie o jednostajnej zgodności metody *bootstrap-t*.

Twierdzenie 1.3.8 (Jednostajna zgodność metody *bootstrap-t*)

Przypuśćmy, że spełnione są założenia A1 dla $s = 4$, A2 dla $r > 2$ ($r \in N$) oraz B dla $p(n) = o((n/\log(n))^{\frac{1}{2r+2}})$. Dodatkowo założymy, że $\{\epsilon_t\}$ jest ciągiem zmiennych losowych i.i.d. o rozkładzie ciągłym. Wówczas

$$\sup_{x \in R} \left| P^* \left(\frac{X_{n+h}^* - \hat{X}_{n+h}^*}{\hat{\sigma}_n^*(h)} \leq x \right) - P \left(\frac{X_{n+h} - \hat{X}_{n+h}}{\hat{\sigma}_n(h)} \leq x \right) \right| \xrightarrow{P} 0.$$

DOWÓD TWIERDZENIA 1.3.8

Dowód przebiega analogicznie jak w przypadku Tw. 1.3.7. Wykorzystujemy kolejno: Tw. 1.3.5 (Zgodność metody *bootstrap-t*), Lemat 1.3.14, Lemat 1.3.12 oraz Lemat 1.3.15.

□

Uwaga 1.3.3

Zauważmy, że zasadniczą ideą dowodu Tw. 1.3.7 oraz Tw. 1.3.8 było wykorzystanie Lematu 1.3.15 (Lemat 9, Belyaev i Sjöstedt-de Luna (2000)) podającego warunki, których spełnienie gwarantuje, że ciągi rozkładów zgodne w sensie słabej aproksymacji są również zgodne jednostajnie.

Alternatywnie można uzasadnić powyższe twierdzenia wykorzystując fakt, że błąd predykcji oraz bootstrapowy błąd predykcji są zbieżne według rozkładu (i odpowiednio według rozkładu warunkowego) do tej samej granicy, tzn.:

$$X_{n+h} - \hat{X}_{n+h} \xrightarrow{d} d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1} + \epsilon_h$$

oraz

$$X_{n+h}^* - \hat{X}_{n+h}^* \xrightarrow{d^*} d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1} + \epsilon_h$$

według prawdopodobieństwa, gdzie $\underline{\phi}_{h-1} = (\phi_1, \dots, \phi_{h-1})$. W tym przypadku tezę o zbieżności jednostajnej ciągu dystrybuant otrzymamy stosując tw. Póly'a oraz jego uogólnioną wersję (Tw. 1.3.6).

1.3.6 Zgodność przedziałów predykcyjnych

W Podrozdziale 1.3.1 omówiliśmy ogólną koncepcję zgodności metody bootstrap. Przedstawione zostały także najczęściej wykorzystywane odległości (takie jak: metryka supremum, odległość Lévy'ego, etc.) będące miarami bliskości rozkładu bootstrapowego i rozkładu aproksymowanego.

Badając asymptotyczne zachowanie bootstrapowych przedziałów ufności najbardziej interesuje nas jednak czy i ewentualnie jak szybko, prawdopodobieństwo pokrycia dla skonstruowanego przedziału zbiega do nominalnego poziomu ufności, gdy rozmiar próby $n \rightarrow \infty$. W ten sposób definiowana jest zgodność oraz dokładność (ang. *accuracy*) bootstrapowych przedziałów ufności (Definicja 4.1 i 4.2, Shao i Tu (1995)).

Aby lepiej porównać właściwości skonstruowanych przedziałów ufności niekiedy rozważa się także inne miary dokładności, takie jak np. długość przedziału. Zagadnieniom tym poświęcimy więcej miejsca w Podrozdziale 1.5.

Warto zauważyć, że dla danych i.i.d. zgodność rozkładu bootstrapowego zazwyczaj implikuje zgodność bootstrapowych przedziałów ufności (Tw. 4.1, Shao i Tu (1995)), podczas gdy wynikanie odwrotne może nie zachodzić (Przykład 3.8, Shao i Tu (1995)).

Podamy teraz formalną definicję zgodnego przedziału predykcyjnego.

Definicja 1.3.10 (Zgodny przedział predykcyjny)

Przedział predykcyjny $\hat{I}(h)$ o nominalnym poziomie ufności $(1 - 2\alpha)$, skonstruowany dla przyszłej (nieznanej) wartości X_{n+h} , ($h \geq 1$) będziemy nazywali zgodnym jeżeli

$$P(X_{n+h} \in \hat{I}(h)) \longrightarrow 1 - 2\alpha, \quad \text{gdy } n \longrightarrow \infty.$$

W uzasadnieniu zgodności przedziałów predykcyjnych zasadniczą rolę odgrywać będą rezultaty dotyczące zbieżności kwantyli dla słabo zbieżnego ciągu dystrybuant. Dlatego też sformułujemy teraz odpowiednie fakty pomocnicze.

Definicja 1.3.11 (Kwantyl rzędu α)

Kwantylem rzędu α rozkładu o dystrybuancie G będziemy nazywali:

$$G^{-1}(\alpha) = \inf\{x : G(x) \geq \alpha\}.$$

Lemat 1.3.16 (Lemat 1.2.1, Politis i Romano (1999))

Jeżeli $\{G_n\}$ jest ciągiem dystrybuant zbieżnych słabo do dystrybuanty G (ozn. $G_n \Rightarrow G$) oraz G jest ciągła i ściśle rosnąca w $y = G^{-1}(\alpha)$, to

$$G_n^{-1}(\alpha) \rightarrow G^{-1}(\alpha) \quad \text{gdy } n \longrightarrow \infty.$$

Lemat 1.3.17 stanowi uogólnienie powyższego wyniku na przypadek rozkładów warunkowych.

Lemat 1.3.17 Jeżeli $\{F_n^*\}$ jest ciągiem dystrybuant zbieżnych słabo do dystrybuanty F według prawdopodobieństwa (ozn. $F_n^* \Rightarrow F$ wg. p -stwa) oraz F jest ciągła i ściśle rosnąca w $y = F^{-1}(\alpha)$, to

$$F_n^{*-1}(\alpha) \rightarrow F^{-1}(\alpha) \quad \text{wg. } p\text{-stwa}.$$

DOWÓD LEMATU 1.3.17.

Niech $\delta > 0$ będzie dowolne, ustalone. Oznaczmy przez y kwantyl rzędu α odpowiadający rozkładowi o dystrybuancie F , tzn. $y = F^{-1}(\alpha)$. Niech $y - \epsilon$ oraz $y + \epsilon$ będą punktami ciągłości dystrybuanty F , gdzie $0 < \epsilon \leq \delta$.

Wówczas z założenia otrzymujemy, że:

$$F_n^*(y - \epsilon) \xrightarrow{P} F(y - \epsilon)$$

oraz

$$F_n^*(y + \epsilon) \xrightarrow{P} F(y + \epsilon).$$

Zauważmy, że $F(y - \epsilon) < \alpha$ oraz $F(y + \epsilon) > \alpha$.

Określmy teraz kwantyl rzędu α dla rozkładu o dystrybuancie F_n^* ,

$$F_n^{*-1}(\alpha) := \inf\{x : F_n^*(x) \geq \alpha\}.$$

Zauważmy, że:

$$\begin{aligned} P(F_n^{\star-1}(\alpha) \leq y + \epsilon) &= P(\alpha \leq F_n^{\star}(y + \epsilon)) \\ &= P(\alpha - F(y + \epsilon) \leq 0) \\ &\leq P(F_n^{\star}(y + \epsilon) - F(y + \epsilon) \leq 0). \end{aligned}$$

Ponieważ z założenia $F_n^{\star}(y + \epsilon) - F(y + \epsilon) \xrightarrow{P} 0$ oraz $\alpha - F(y + \epsilon) < 0$ otrzymujemy, że:

$$P(F_n^{\star-1}(\alpha) \leq y + \epsilon) \rightarrow 1 \quad \text{gdy } n \rightarrow \infty. \quad (1.56)$$

Analogicznie można pokazać, że:

$$P(F_n^{\star-1}(\alpha) \geq y - \epsilon) \rightarrow 1 \quad \text{gdy } n \rightarrow \infty. \quad (1.57)$$

Z (1.56) i (1.57) wynika natomiast, że

$$P(y - \epsilon \leq F_n^{\star-1}(\alpha) \leq y + \epsilon) \rightarrow 1 \quad \text{gdy } n \rightarrow \infty. \quad (1.58)$$

Pokazaliśmy więc, że dla dowolnego, ustalonego $\epsilon > 0$:

$$P(|F_n^{\star-1}(\alpha) - y| \leq \epsilon) \rightarrow 1 \quad \text{gdy } n \rightarrow \infty, \quad (1.59)$$

to oznacza, że:

$$F_n^{\star-1}(\alpha) \xrightarrow{P} y = F^{-1}(\alpha), \quad (1.60)$$

co kończy dowód Lematu 1.3.17.

□

Odnótujmy jeszcze prosty wniosek będący natychmiastową konsekwencją Lematu 1.3.17 oraz faktu, że zbiór punktów nieciągłości dla dowolnej dystrybucyjności F jest co najwyżej przeliczalny.

Wniosek 1.3.1

Jeżeli $F_n^{\star} \Rightarrow F$ wg. p -stwa, to zbiór

$$\{t : F_n^{\star-1}(t) \not\rightarrow F^{-1}(t) \text{ wg. } p\text{-stwa} \}$$

jest co najwyżej przeliczalny.

Zauważmy, że Wniosek 1.3.1 stanowi uogólnienie analogicznego rezultatu (Lemat 1.5.6, Serfling (1991)) udowodnionego dla przypadku słabej zbieżności ciągu dystrybucyjności (tzn. $G_n \Rightarrow G$).

Przejdźmy teraz do sformułowania i udowodnienia zasadniczych rezultatów związanych ze zgodnością przedziałów predykcyjnych.

W Podrozdziale 1.2 zdefiniowaliśmy bootstrapowy przedział predykcyjny skonstruowany z wykorzystaniem metody *hybrid bootstrap* na nominalnym poziomie ufności $(1 - 2\alpha)$:

$$I_B(h) = [\hat{X}_{n+h} + q_\alpha^*, \hat{X}_{n+h} + q_{1-\alpha}^*], \quad (1.61)$$

gdzie q_α^* oraz $q_{1-\alpha}^*$ są kwantylami rozkładu bootstrapowego błędu predykcji $\Delta_n^*(h) = X_{n+h}^* - \hat{X}_{n+h}^*$.

Analogicznie określony był studentyzowany przedział predykcyjny:

$$I_{B-t}(h) = [\hat{X}_{n+h} + t_\alpha^* \hat{\sigma}_n(h), \hat{X}_{n+h} + t_{1-\alpha}^* \hat{\sigma}_n(h)], \quad (1.62)$$

gdzie t_α^* oraz $t_{1-\alpha}^*$ są kwantylami bootstrapowego rozkładu unormowanego błędu predykcji $T_n^*(h) = \frac{X_{n+h}^* - \hat{X}_{n+h}^*}{\hat{\sigma}_n^*(h)}$.

Zastępując nieznane kwantyle $q_\alpha^*, q_{1-\alpha}^*$ oraz $t_\alpha^*, t_{1-\alpha}^*$ ich estymatorami uzyskanymi za pomocą metody Monte Carlo, na bazie B bootstrapowych replikacji statystyk $\Delta_n^*(h)$ i $T_n^*(h)$, otrzymujemy odpowiednio:

$$\hat{I}_B(h) = [\hat{X}_{n+h} + \hat{q}_\alpha^*, \hat{X}_{n+h} + \hat{q}_{1-\alpha}^*], \quad (1.63)$$

oraz

$$\hat{I}_{B-t}(h) = [\hat{X}_{n+h} + \hat{t}_\alpha^* \hat{\sigma}_n(h), \hat{X}_{n+h} + \hat{t}_{1-\alpha}^* \hat{\sigma}_n(h)]. \quad (1.64)$$

Twierdzenie 1.3.9 (Zgodność przedziałów hybrydowych)

Niech spełnione będą założenia: A1 dla $s = 4$, A2 dla $r > 1$ oraz B dla $p(n) = o((n/\log(n))^{1/(2r+2)})$. Ponadto, niech $c_\alpha, c_{1-\alpha}$ – kwantyle rozkładu zmiennej losowej $A_{\epsilon,h}$ (określonej jak w dowodzie) będą punktami ciągłości dystrybucyjnego tego rozkładu.

Wówczas

$$P(X_{n+h} \in \hat{I}_B(h)) \longrightarrow 1 - 2\alpha, \quad \text{gdy } n \longrightarrow \infty.$$

DOWÓD TWIERDZENIA 1.3.9

Zauważmy, że:

$$\begin{aligned} P(X_{n+h} \in \hat{I}_B(h)) &= P(\hat{X}_{n+h} + \hat{q}_\alpha^* \leq X_{n+h} \leq \hat{X}_{n+h} + \hat{q}_{1-\alpha}^*) \\ &= P(\hat{q}_\alpha^* \leq X_{n+h} - \hat{X}_{n+h} \leq \hat{q}_{1-\alpha}^*) \\ &= P(X_{n+h} - \hat{X}_{n+h} \leq \hat{q}_{1-\alpha}^*) - P(X_{n+h} - \hat{X}_{n+h} < \hat{q}_\alpha^*). \end{aligned}$$

Aby uzasadnić zgodność przedziałów $\hat{I}_B(h)$ wystarczy zatem pokazać, że przy n zmierzającym do ∞ :

$$P(X_{n+h} - \hat{X}_{n+h} \leq \hat{q}_{1-\alpha}^*) \longrightarrow 1 - \alpha$$

oraz

$$P(X_{n+h} - \hat{X}_{n+h} < \hat{q}_\alpha^*) \longrightarrow \alpha.$$

Przypomnijmy, że na mocy Faktów 1.3.9 oraz 1.3.10 możemy przedstawić błąd predykcji i odpowiednio bootstrapowy błąd predykcji jako:

$$\begin{aligned} X_{n+h} - \hat{X}_{n+h} &= \\ &= d_{1,h}(\underline{\phi}_{h-1}) \epsilon_{n+1} + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{n+h-1} + \epsilon_{n+h} + o_P(1) \end{aligned}$$

oraz

$$\begin{aligned} X_{n+h}^* - \hat{X}_{n+h}^* &= \\ &= d_{1,h}(\underline{\phi}_{h-1}) \epsilon_{n+1}^* + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{n+h-1}^* + \epsilon_{n+h}^* + o_{P^*}(1), \end{aligned}$$

gdzie $\underline{\phi}_{h-1} = (\phi_1, \dots, \phi_{h-1})$.

Oznaczmy

$$A_{\epsilon,h} := d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1} + \epsilon_h.$$

Wówczas

$$\Delta_n(h) = X_{n+h} - \hat{X}_{n+h} \xrightarrow{d} A_{\epsilon,h}. \quad (1.65)$$

Na mocy Lematu 5.4 (Bühlmann, (1997)) otrzymujemy również:

$$\Delta_n^*(h) = X_{n+h}^* - \hat{X}_{n+h}^* \xrightarrow{d^*} A_{\epsilon,h} \quad \text{według prawdopodobieństwa.} \quad (1.66)$$

Niech $q_{1-\alpha}$ oznacza kwantyl rzędu $(1-\alpha)$ rozkładu błędu predykcji $\Delta_n(h) = X_{n+h} - \hat{X}_{n+h}$. Oznaczmy również przez $c_{1-\alpha}$ kwantyl rzędu $(1-\alpha)$ rozkładu zmiennej losowej $A_{\epsilon,h}$. Ponieważ z założenia $c_{1-\alpha}$ jest punktem ciągłości dystrybuanty rozkładu zmiennej losowej $A_{\epsilon,h}$, to na mocy Lematu 1.3.16, Lematu 1.3.17 oraz zgodności kwantyli próbkowych mamy:

$$\begin{aligned} q_{1-\alpha}^* - q_{1-\alpha} &= o_P(1), \\ q_{1-\alpha} - c_{1-\alpha} &= o(1), \\ \hat{q}_{1-\alpha}^* - q_{1-\alpha}^* &= o_P(1). \end{aligned}$$

Możemy zatem napisać:

$$\begin{aligned} P(X_{n+h} - \hat{X}_{n+h} \leq \hat{q}_{1-\alpha}^*) &= \\ &= P(X_{n+h} - \hat{X}_{n+h} \leq (\hat{q}_{1-\alpha}^* - q_{1-\alpha}^*) + (q_{1-\alpha}^* - q_{1-\alpha}) + (q_{1-\alpha} - c_{1-\alpha}) + c_{1-\alpha}) \\ &= P(X_{n+h} - \hat{X}_{n+h} + o_P(1) \leq c_{1-\alpha}). \end{aligned}$$

Na mocy Lematu Śluckiego

$$X_{n+h} - \hat{X}_{n+h} + o_P(1) \xrightarrow{d} A_{\epsilon,h},$$

a to oznacza, że:

$$P(X_{n+h} - \hat{X}_{n+h} + o_P(1) \leq c_{1-\alpha}) \longrightarrow P(A_{\epsilon,h} \leq c_{1-\alpha}) = 1 - \alpha$$

jeżeli tylko $c_{1-\alpha}$ jest punktem ciągłości dystrybuanty rozkładu zmiennej $A_{\epsilon,h}$.

W analogiczny sposób możemy uzasadnić, że:

$$P(X_{n+h} - \hat{X}_{n+h} < \hat{q}_{\alpha}^*) \longrightarrow \alpha$$

jeżeli c_{α} jest punktem ciągłości dystrybuanty zmiennej $A_{\epsilon,h}$, co kończy dowód Tw.1.3.9.

□

Podobny rezultat otrzymujemy dla przedziałów predykcyjnych skonstruowanych w oparciu o metodę *bootstrap-t*.

Twierdzenie 1.3.10 (Zgodność przedziałów studentyzowanych)

Niech spełnione będą założenia: A1 dla $s = 4$, A2 dla $r > 2$ oraz B dla $p(n) = o((n/\log(n))^{1/(2r+2)})$. Ponadto, niech u_α , $u_{1-\alpha}$ – kwantyle rozkładu zmiennej losowej $A_{\epsilon,h}/\sigma(h)$ (określonej jak w dowodzie) będą punktami ciągłości dystrybuanty tego rozkładu.

Wówczas

$$P(X_{n+h} \in \hat{I}_{B-t}(h)) \longrightarrow 1 - 2\alpha, \quad \text{gdy } n \longrightarrow \infty.$$

DOWÓD TWIERDZENIA 1.3.10.

Oznaczmy przez $t_{1-\alpha}$ kwantyl rzędu $(1 - \alpha)$ rozkładu unormowanego błędu predykcji $T_n(h) = \Delta_n(h)/\hat{\sigma}_n(h)$. Ponadto, niech $u_{1-\alpha}$ oznacza kwantyl rozkładu zmiennej losowej $A_{\epsilon,h}/\sigma(h)$, gdzie $\sigma(h)$ określone jak w Podrozdziale 1.3.4, zmienna losowa $A_{\epsilon,h}$ jest określona jak w dowodzie Tw.1.3.9.

Wykorzystując Fakt 1.3.10 otrzymujemy, że $\hat{\sigma}_n(h) \xrightarrow{P} \sigma(h)$, a to na mocy (1.65) oraz Lematu Śluckiego implikuje:

$$T_n(h) = \frac{\Delta_n(h)}{\hat{\sigma}_n(h)} \xrightarrow{d} \frac{A_{\epsilon,h}}{\sigma(h)}. \quad (1.67)$$

Analogicznie, wykorzystując Fakt 1.3.11, (1.66) oraz uogólniony Lemat Śluckiego (Lemat 1.3.3) otrzymujemy:

$$T_n^*(h) = \frac{\Delta_n^*(h)}{\hat{\sigma}_n^*(h)} \xrightarrow{d^*} \frac{A_{\epsilon,h}}{\sigma(h)} \quad \text{według prawdopodobieństwa.} \quad (1.68)$$

Ponieważ, z założenia $u_{1-\alpha}$ jest punktem ciągłości dystrybuanty rozkładu zmiennej losowej $A_{\epsilon,h}/\sigma(h)$ wykorzystując Lemat 1.3.16, Lemat 1.3.17 oraz zgodność kwantyli próbkowych otrzymujemy:

$$\begin{aligned} t_{1-\alpha}^* - t_{1-\alpha} &= o_P(1), \\ t_{1-\alpha} - u_{1-\alpha} &= o(1), \\ \hat{t}_{1-\alpha}^* - t_{1-\alpha}^* &= o_P(1). \end{aligned} \quad (1.69)$$

Zbieżność kwantyli (1.69) oraz Lemat Śluckiego implikują, że:

$$\begin{aligned} P\left(\frac{X_{n+h} - \hat{X}_{n+h}}{\hat{\sigma}_n(h)} \leq \hat{t}_{1-\alpha}^*\right) &= \\ &= P(T_n(h) \leq (\hat{t}_{1-\alpha}^* - t_{1-\alpha}^*) + (t_{1-\alpha}^* - t_{1-\alpha}) + (t_{1-\alpha} - u_{1-\alpha}) + u_{1-\alpha}) \\ &= P(T_n(h) + o_P(1) \leq u_{1-\alpha}) \longrightarrow P\left(\frac{A_{\epsilon,h}}{\sigma(h)} \leq u_{1-\alpha}\right) = 1 - \alpha, \end{aligned} \quad (1.70)$$

o ile $u_{1-\alpha}$ jest punktem ciągłości dystrybuanty rozkładu $A_{\epsilon,h}/\sigma(h)$.

W podobny sposób możemy uzasadnić, że:

$$P\left(\frac{X_{n+h} - \hat{X}_{n+h}}{\hat{\sigma}_n(h)} < \hat{t}_\alpha^*\right) \rightarrow \alpha. \quad (1.71)$$

Na mocy (1.70) i (1.71) otrzymujemy już tezę Twierdzenia 1.3.10.

□

Uwaga 1.3.4

Zauważmy, że zgodność przedziałów hybrydowych udowodniliśmy dla poziomu ufności $(1 - 2\alpha)$, takiego że $c_\alpha, c_{1-\alpha}$ – kwantyle rozkładu zmiennej losowej $A_{\epsilon, h} = d_{1, h}(\phi_{h-1}) \epsilon_1 + \dots + d_{h-1, h}(\phi_{h-1}) \epsilon_{h-1} + \epsilon_h$ są punktami ciągłości dystrybucyjności tego rozkładu. Analogiczny warunek sformułowany był dla przedziałów predykcyjnych skonstruowanych na bazie metody bootstrap-t.

Jeżeli założymy dodatkowo, że dystrybucyjność rozkładu ϵ_t jest funkcją ciągłą otrzymamy zgodność bootstrapowych przedziałów predykcyjnych (hybrydowych i studentyzowanych) na dowolnym poziomie ufności $(1 - 2\alpha)$.

1.4 Bootstrap wygładzony

W standardowej metodzie bootstrap, do generowania replikacji residuów, tzn. ϵ_t^* wykorzystywana była dystrybucyjność empiryczna $\hat{F}_{\epsilon, n}$ określona na zbiorze estymowanych i następnie scentrowanych residuów $\{\tilde{\epsilon}_t\}_{t=p(n)+1}^n$.

Niekiedy rozważa się modyfikację standardowej metody bootstrap określaną w literaturze jako bootstrap wygładzony (ang. *smoothed bootstrap*). Główną ideą bootstrapu wygładzonego jest generowanie obserwacji z rozkładu określonego przez wygładzony estymator dystrybucyjności, tzn.:

$$\tilde{F}_{\epsilon, h}(x) = \frac{1}{n - p(n)} \sum_{t=1}^{n-p(n)} \tilde{K}\left(\frac{x - \tilde{\epsilon}_t}{h}\right),$$

gdzie $\tilde{K}(z) = \int_{-\infty}^z K(u) du$, natomiast funkcja jądrowa K jest symetryczną funkcją gęstości rozkładu.

Alternatywnie możemy przedstawić estymator $\tilde{F}_{\epsilon, h}$ jako:

$$\tilde{F}_{\epsilon, h}(x) = \int_{-\infty}^x \hat{f}_{\epsilon, h}(t) dt,$$

gdzie $\hat{f}_{\epsilon, h}(t)$ -jądrowy estymator gęstości rozkładu, tzn.

$$\hat{f}_{\epsilon, h}(t) = \int \frac{1}{h} K\left(\frac{t - y}{h}\right) \hat{F}_{\epsilon, n}(dy) = \frac{1}{(n - p(n))h} \sum_{i=1}^{n-p(n)} K\left(\frac{\tilde{\epsilon}_i - t}{h}\right).$$

W dalszej części przez $\{\tilde{\epsilon}_{t, h}^*\}$ będziemy oznaczali zmienną losową o dystrybucyjności $\tilde{F}_{\epsilon, h}$.

Zauważmy, że próba bootstrapowa generowana z rozkładu o dystrybuancie $\tilde{F}_{\epsilon,h}$ nie będzie zawierała powtarzających się obserwacji. Sytuacja taka ma często miejsce w przypadku standardowej metody bootstrap i wydaje się być nie do zaakceptowania gdy mamy do czynienia ze zmiennymi losowymi o rozkładzie ciągłym. Co więcej, powtarzające się obserwacje w replikacji mogą prowadzić np. do niestabilności studentyzowanych przedziałów ufności będącej konsekwencją faktu, że estymator błędu standardowego może być w tym przypadku bliski zeru. Bardziej szczegółowo problem ten będzie omówiony w rozdziale 1.4.4.

1.4.1 Przegląd wyników

Nadaraya (1964) udowodnił (przy umiarkowanych założeniach), że $\tilde{F}_{\epsilon,h}$ ma tą samą asymptotyczną średnią i wariancję jak dystrybuanta empiryczna $\hat{F}_n$.

Azzalini (1981) otrzymał natomiast rezultaty dotyczące błędu średniokwadratowego (MSE), pokazujące asymptotyczną poprawę w estymacji dystrybuanty F za pomocą wygładzonego estymatora $\tilde{F}_{\epsilon,h}$ zamiast dystrybuanty empirycznej $\hat{F}_n$ jeżeli F spełnia pewne warunki regularności.

Wielu autorów, poczynawszy od pomysłodawcy metody bootstrap B. Efrona, analizowało możliwe zalety związane z generowaniem replikacji obserwacji z wykorzystaniem gładkiego estymatora $\tilde{F}_{\epsilon,h}$.

Efron (1982) wykorzystując symulacje komputerowe porównał zachowanie bootstrapu wygładzonego i standardowej metody bootstrap wykorzystując jako funkcje jądrowe dwuwymiarowy rozkład normalny (N-smoothed) oraz rozkład jednostajny na rombie (U-smoothed). Rozważanymi przez niego statystykami były próbkowy współczynnik korelacji $\hat{\rho}_n$ i jego transformacja stabilizująca wariancję, tzn. $\hat{\phi}_n = \frac{1}{2} \log[(1 + \hat{\rho}_n)/(1 - \hat{\rho}_n)]$.

Analizując współczynnik zmienności oraz błąd średniokwadratowy otrzymano, że bootstrap wygładzony (dla obu użytych jąder) przynosi istotną poprawę w stosunku do bootstrapu niewygładzonego w przypadku $T_n = \hat{\phi}_n$ ale jedynie nieznacznie gdy $T_n = \hat{\rho}_n$.

Wyniki symulacji nie dają jednak w ogólnej sytuacji odpowiedzi na pytanie czy powinniśmy stosować wygładzenie.

Silverman i Young (1987) badali zachowanie estymatorów dla parametrów postaci: $A(F) = \int a(x) dF(x)$. Analizując MSE pokazali, że w pewnych przypadkach jest możliwa poprawa zachowania estymatorów poprzez wygładzanie. Otrzymane warunki nie mają jednak jasnej interpretacji statystycznej, a ponadto w większości rozważanych problemów wygładzanie może mieć tylko drugorzędny efekt na zachowanie estymatorów, w tym sensie że wariancja optymalnie wygładzonego estymatora jest tego samego rzędu co wariancja estymatora niewygładzonego i oba estymatory są asymptotycznie równoważne.

Wang (1989) badał wpływ wygładzania na estymację gęstości rozkładu statystyki $\sqrt{n}(\bar{X} - \mu)$ (w ustalonym punkcie x). Pokazał, że przy określonych założeniach estymator oparty na bootstrapie wygładzonym ma mniejszy MSE niż estymator niewygładzony. Nie daje to jednak definitywnego rozstrzygnięcia ponieważ w niektó-

rych przypadkach bootstrap wygładzony nie koniecznie jest lepszy.

Cytowane powyżej rezultaty pokazują, że w przypadku gdy n jest odpowiednio duże i niewygładzony estymator bootstrapowy ma rząd zbieżności $n^{-1/2}$ nie możemy oczekiwać aby wygładzenie przyniosło istotną poprawę. Z odmienną sytuacją mamy do czynienia gdy statystyka T_n nie jest wystarczająco gładka, w tym sensie, że niewygładzony estymator ma wolniejszy rząd zbieżności (np. zbiega dużo wolniej niż $n^{-1/2}$) co występuje np. dla kwantyli próbkowych.

Falk i Reiss (1989) analizując zachowanie bootstrapowych estymatorów dla kwantyli próbkowych pokazali, że

$$H_{sboot}(x) = P^*\{\sqrt{n}[\tilde{F}_{n,h}^{*-1}(t) - \tilde{F}_{n,h}^{-1}(t)] \leq x\}$$

ma szybszy rząd zbieżności niż

$$H_{boot}(x) = P^*\{\sqrt{n}[\hat{F}_n^{*-1}(t) - \hat{F}_n^{-1}(t)] \leq x\},$$

przy czym zbieżność $H_{boot}(x)$ do

$$H_n(x) = P\{\sqrt{n}[\hat{F}_n^{-1}(t) - F^{-1}(t)] \leq x\}$$

jest rzędu $n^{-1/4}$.

Hall, DiCiccio i Romano (1989) badali wpływ wygładzania na estymację wariancji kwantyli próbkowych. Oznaczmy odpowiednio przez V_{boot} oraz V_{sboot} wariancję uzyskaną z wykorzystaniem standardowej metody bootstrap oraz metody wygładzonej. V_{boot} ma rząd zbieżności $n^{-1/4}$, który nie jest zadowalający. Przyjmując określone warunki regularności (dotyczące m.in. gęstości rozkładu oraz funkcji jądrowej K) oraz przy odpowiednim doborze szerokości okna h_n otrzymano, że V_{sboot} ma rząd zbieżności $n^{-r/(2r+1)}$, gdzie r jest rzędem jądra tzn.

$$\begin{aligned} \int |x^r K(x)| dx &< \infty \\ \int x^j K(x) dx &= 1 \quad \text{dla } j = 0, \\ \int x^j K(x) dx &= 0 \quad \text{dla } 1 \leq j \leq r-1 \\ \int x^r K(x) dx &\neq 0. \end{aligned}$$

Wybierając jądro rzędu r możemy otrzymać rząd zbieżności dla V_{sboot} : $n^{-1/2+\epsilon_0}$ dla dowolnego $\epsilon_0 > 0$.

Uwaga 1.4.1

Zauważmy, że jedynie w przypadku $r = 2$ jądro K może być rozkładem prawdopodobieństwa, ale $V_{sboot} = V_n(\tilde{F}_{n,h})$ jest dalej dobrze określone nawet gdy $\tilde{F}_{n,h}$ nie jest dystrybuantą rozkładu. Lee i Young (1994) zaproponowali zastosowanie tzw. „poprawki ujemności” jądra K , tak aby $\tilde{F}_{n,h}$ była dystrybuantą rozkładu.

W omówionych powyżej pracach wpływ wygładzania na poprawę dokładności estymacji mierzony był z wykorzystaniem błędu średniokwadratowego. Otrzymane wyniki nie wykazały jednak znaczącej poprawy MSE związanej z wykorzystaniem

$\tilde{F}_{\epsilon,h}$. Z istotną poprawą mamy jednak do czynienia przyjmując jako kryterium porównawcze prawdopodobieństwo pokrycia dla bootstrapowych przedziałów ufności.

Polansky i Schucany (1997) rozważając stosunkowo ogólny tzw. gładki model (*smooth function model*) (obejmujący np. takie parametry jak średnia i wariancja) pokazali, że wybierając szerokość okna rzędu $n^{-1/4}$ (a nie $n^{-1/5}$ jak w przypadku estymacji gęstości) można otrzymać redukcję pierwszego rzędu dla błędu pokrycia w przypadku jednostronnych przedziałów kwantylowych. Metoda kwantylowa z wykorzystaniem bootstrapu wygładzonego staje się więc tak dokładna jak bootstrap-t albo BCa.

Analogiczną redukcję można przeprowadzić także dla przedziałów ufności, które mają już dokładność drugiego rzędu (np. dwustronny przedział kwantylowy lub jednostronny przedział studentyzowany). W tym przypadku optymalna szerokość okna jest rzędu $n^{-1/2}$.

Cao (1997) zastosował ideę bootstrapu wygładzonego do konstrukcji warunkowych, bootstrapowych przedziałów predykcyjnych dla procesów $AR(p)$. Otrzymał obiecujące wyniki symulacyjne dotyczące prawdopodobieństw pokrycia, które wykazują przewagę metody wygładzonej nad standardową metodą bootstrap opartą na dystrybuancie empirycznej.

1.4.2 Zgodność wygładzonej metody bootstrap

Wykorzystując strukturę splotową estymatora $\tilde{F}_{\epsilon,h}$ możemy przedstawić:

$$\tilde{\epsilon}_i^* = \epsilon_i^* + h \cdot u_i, \quad (1.72)$$

gdzie $\epsilon_i^* \sim i.i.d.$ o dystrybuancie empirycznej $\hat{F}_{\epsilon,n}$, oraz $u_i \sim i.i.d.$ o dystrybuancie $\tilde{K}$. Reprezentacja (1.72) pozwala m.in. łatwo generować liczby losowe z rozkładu $\tilde{F}_{\epsilon,h}$.

Lemat 1.4.1 *Założmy, że spełnione są warunki: A1 dla $s = 4$, A2 dla $r = 1$ oraz B dla $p(n) = o((n/\log(n))^{1/2})$. Ponadto, niech $h = h(n) \rightarrow 0$ gdy $n \rightarrow \infty$. Wówczas:*

$$\tilde{\epsilon}_{t,h}^* \xrightarrow{d^*} \epsilon_t \text{ wg. prawdopodobieństwa.}$$

DOWÓD LEMATU 1.4.1.

Niech $\hat{\phi}_n^*$, $\hat{\phi}_{n,h}^*$ oraz $\phi_{\tilde{K}}$ oznaczają funkcje charakterystyczne odpowiadające dystrybuantom $\hat{F}_{\epsilon,n}$, $\tilde{F}_{\epsilon,h}$ oraz $\tilde{K}$. Wówczas możemy zapisać:

$$\hat{\phi}_{n,h}^*(z) = \hat{\phi}_n^*(z) \phi_{\tilde{K}}(hz).$$

Z Lematu 5.4 (Bühlmann (1997)) otrzymujemy, że $\epsilon_t^* \xrightarrow{d^*} \epsilon_t$ wg. prawdopodobieństwa. To oznacza, że

$$\hat{\phi}_n^*(z) \rightarrow \phi_\epsilon(z) \text{ wg. prawdopodobieństwa,}$$

gdzie $\phi_\epsilon(\cdot)$ jest funkcją charakterystyczną odpowiadającą ϵ_t .

Oczywiście ponieważ $h = h(n) \rightarrow 0$ gdy $n \rightarrow \infty$, to $\phi_{\hat{K}}(hz) \rightarrow 1$.
Ostatecznie zatem, z Lematu Słuckiego otrzymujemy, że:

$$\hat{\phi}_{n,h}^*(z) \rightarrow \phi_\epsilon(z) \quad \text{wg. prawdopodobieństwa}$$

co jest równoważne zbieżności według rozkładu, tzn.:

$$\tilde{\epsilon}_{t,h}^* \xrightarrow{d^*} \epsilon_t \quad \text{wg. prawdopodobieństwa.}$$

□

1.4.3 Dobór szerokości okna h

Istotnym, a zarazem niełatwym zagadnieniem związanym ze stosowaniem bootstrapu wygładzonego jest dobór optymalnej szerokości okna h . W tym celu możemy wykorzystać algorytm zaproponowany przez Cao (1997).

Wiadomo (Azzalini (1981)), że wielkością asymptotycznie minimalizującą MISE jest

$$h_{AMISE} = \left[\frac{1 - \int_{-1}^1 (\int_{-1}^v K(z) dz)^2 dv}{(\int_{-1}^1 z^2 K(z) dz)(n - p(n)) \int (f'(x))^2 dx} \right]^{1/3},$$

gdzie K - jądro o nośniku $[-1, 1]$. (W symulacjach przyjęto jądro Bartletta).

Estymujemy $I = \int (f'(x))^2 dx$. (Hall i Marron (1987))

$$\hat{I} = -\frac{1}{(n - p(n))^2 g^3} \sum_{i \neq j} K^{(2)} \left(\frac{\tilde{\epsilon}_i - \tilde{\epsilon}_j}{g} \right),$$

gdzie g - tzw. „pilot bandwidth”.

Asymptotycznie optymalnym wyborem szerokości g jest

$$g_0 = \left[\frac{5 \int K^{(2)2} \int f^2}{(n - p(n))^2 (\int K^2)^2 (\int f^{(2)2})^2} \right]^{1/9}.$$

Dla jądra gaussowskiego i przy założeniu normalności rozkładu mamy

$$g_0 = \left(\frac{80\pi}{3(n - p(n))^2} \right)^{1/9} \sigma,$$

którą możemy estymować

$$\hat{g}_0 = \left(\frac{80\pi}{3(n - p(n))^2} \right)^{1/9} \hat{\sigma}.$$

Ostatecznie schemat doboru szerokości okna wygląda następująco:

$$\hat{g}_0 \rightarrow \hat{I} \rightarrow \hat{h}_{AMISE}.$$

1.4.4 Stabilizacja studentyzowanych (bootstrap-t) przedziałów ufności

Badania symulacyjne potwierdziły, że metoda bootstrap-t zachowuje się bardzo dobrze analizując błąd pokrycia, nawet dla prób o małym rozmiarze. Z drugiej jednak strony te same badania wykazują, że otrzymane przedziały mogą być niestabilne i zbyt szerokie (np. Polansky i Schucany (1997)) szczególnie dla prób o małej i średniej długości.

Przyczyny takiej sytuacji mogą być następujące:

- (i) Metoda bootstrap-t jest wrażliwa na obserwacje odstające (Efron i Tibshirani (1993)),
- (ii) Dla wielu funkcjonałów estymator błędu standardowego jest niestabilny i wymagana jest transformacja stabilizująca wariancję (np. dla współczynnika korelacji),
- (iii) Dla prób o małym rozmiarze jest wysoce prawdopodobne, że estymator błędu standardowego $\hat{\sigma}^*$ może być bliski (lub nawet równy) zeru powodując tym samym, względnie duże wartości T_n^* . Sytuacja taka jest następstwem dyskretnej natury estymatora $\hat{F}_n$, która powoduje występowanie powtarzających się obserwacji w replikacji.

Efron (1992) próbował rozwiązać opisany wyżej problem dodając ad hoc stałą do estymatora błędu standardowego co stabilizowało rozkład bootstrapowy. Nie zostało jednak przedstawione odpowiednie uzasadnienie teoretyczne takiej poprawy.

Polansky (2001) zaproponował ostatnio dwie metody stabilizujące końce przedziałów studentyzowanych w przypadku małych prób i jednocześnie zachowujące na odpowiednim poziomie prawdopodobieństwo pokrycia:

- metodę addytywną polegającą na dodaniu stałej do replikacji estymatora $\hat{\sigma}^*$,
- metodę wygładzoną opartą na replikowaniu z ciągłego, nieparametrycznego estymatora $\tilde{F}_{n,h}$.

Rezultaty teoretyczne w pracy Polansky'ego pokazują, że przy odpowiednim doborze szerokości okna zachowana jest dokładność drugiego rzędu dla przedziałów studentyzowanych skonstruowanych z wykorzystaniem powyższych metod. Badania symulacyjne porównujące średnią długości przedziałów i odchylenie standardowe długości wykazały natomiast skuteczność obu zaproponowanych metod w redukowaniu niestabilności z jednoczesnym zachowaniem na odpowiednim poziomie prawdopodobieństwa pokrycia.

Warto zauważyć, że procedury zaproponowane przez Polansky'ego stanowią konkurencję dla metody stabilizacji wariancji wprowadzonej przez **Tibshiraniego (1988)**, która może okazać się zawodna w przypadku małych prób i prowadzi do redukcji średniej długości przedziałów za cenę wzrostu błędu pokrycia.

W rozdziale 1.5 poświęconym symulacjom komputerowym będziemy badali czy zastosowanie bootstrapu wygładzonego ma wpływ na poprawę stabilności studentyzowanych przedziałów predykcyjnych skonstruowanych z wykorzystaniem metody *sieve bootstrap*.

1.5 Symulacje komputerowe

Aby zilustrować zachowanie zaproponowanych metod konstrukcji przedziałów predykcyjnych przeprowadzono symulacje komputerowe. W badaniach rozważono następujące modele:

$$(M1) \text{ } ARMA(1, 1), \quad X_t = 0.8X_{t-1} - 0.6\varepsilon_{t-1} + \varepsilon_t,$$

$$(M2) \text{ } AR(48), \quad X_t = \sum_{j=1}^{48} \phi_j X_{t-j} + \varepsilon_t, \quad \phi_j = (-1)^{j+1} 7.5/(j+1)^3 \quad (j = 1, \dots, 48),$$

$$(M3) \text{ } ARFIMA(0, d, 0), \quad (1 - B)^d X_t = \varepsilon_t, \quad \text{dla } d = 0.3,$$

gdzie operator $\nabla^d = (1 - B)^d$ jest określony jak np. u Hoskinga (1981). Za-uważmy, że dla $d \in (-1/2, 1/2)$ tak zdefiniowany X_t jest procesem stacjonarnym, który możemy przedstawić jako proces $AR(\infty)$, Hosking (1981).

Rozważyliśmy następujące rozkłady dla szeregu zakłóceń ε_t :

$$(N) \text{ standardowy rozkład normalny: } N(0, 1),$$

$$(t) \text{ unormowany rozkład Studenta: } t(3)/\sqrt{3},$$

$$(\log N) \text{ unormowany rozkład logarytmiczno-normalny: } (\log N(0, 1) - \sqrt{e})/\sqrt{e(e-1)},$$

$$(M) \text{ mieszanka rozkładów normalnych: } 0.9 N(-1, 1) + 0.1 N(9, 1).$$

Przedziały predykcyjne były konstruowane wykorzystując klasyczną metodę Boxa-Jenkinsa (opartą na aproksymacji gaussowskiej) oraz dwie metody wykorzystujące algorytm *sieve bootstrap*, tzn.: hybrydowe przedziały predykcyjne oraz studentyzowane (*bootstrap-t*) przedziały predykcyjne.

Przedziały gaussowskie były wyznaczone przy założeniu, że znamy postać modelu, z którego wygenerowano dane.

W przypadku bootstrapowych przedziałów predykcyjnych do wyznaczenia predyktorów oraz odpowiednio błędów średniokwadratowych prognoz wykorzystano natomiast aproksymację procesem $AR(p(n))$.

W symulacjach przyjęto następujące parametry:

- rozmiary próby: $n = 25, 50, 100, 500$,
- ilość replikacji bootstrapowych: $B = 1000$,
- liczba powtórzeń Monte Carlo: 1000.

Aproksymujący rząd autoregresyjny $p(n)$ wybrany był na podstawie minimalizacji kryterium AIC (Akaike Information Criteria) dla zakresu $p(n) \in [0, 10 \log_{10}(n)]$ (który jest domyślnym zakresem np. w pakiecie S-PLUS). Wyniki wyboru $\hat{p}_{AIC}(n)$ dla modeli M1-M3 zawiera tabela 1.1.

Rysunki 1.1-1.6 przedstawiają przykładowe przedziały predykcyjne skonstruowane dla wszystkich rozważanych modeli (M1-M3) na nominalnym poziomie ufności 95%.

Dokładność skonstruowanych przedziałów predykcyjnych była badana poprzez wyznaczenie empirycznych prawdopodobieństw pokrycia oraz średnich długości przedziałów ($E(length)$) na bazie 1000 powtórzeń Monte Carlo. Porównania przeprowadzono dla nominalnych poziomów ufności 80% i 95% oraz dla horyzontów $h = 1, \dots, 5$. Oprócz empirycznych prawdopodobieństw pokrycia wyznaczono również (w nawiasach) błędy standardowe. Poniżej przedstawiono tylko wybrane wyniki.

Możemy zaobserwować, że we wszystkich analizowanych przypadkach przedziały studentyzowane charakteryzują się lepszymi wynikami pokrycia empirycznego niż przedziały hybrydowe.

W przypadku prób o małym rozmiarze ($n=25$) wyniki pokrycia dla przedziałów bootstrapowych nie zawsze są jednak zadowalające, co może być związane z obciążeniem obecnym przy estymacji parametrów modelu.

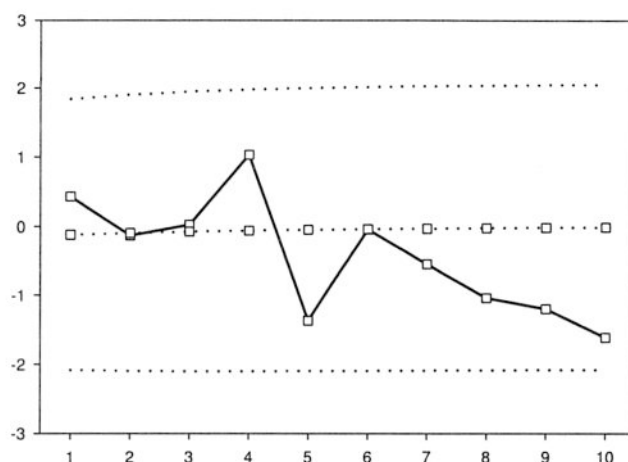
Ponadto, studentyzowane przedziały predykcyjne charakteryzują się znacznie lepszymi wynikami pokrycia niż gaussowskie przedziały Boxa-Jenkinsa w przypadku dwumodalnego (M) rozkładu błędów, co jest szczególnie widocznie dla modelu M1.

Dla nominalnego poziomu ufności 80% możemy zaobserwować (np. Tabela 1.7, podobne wyniki otrzymano dla pozostałych modeli), że dla szeregów innych niż gaussowskie, przedziały Boxa-Jenkinsa są zbyt konserwatywne z czym związana jest również ich największa średnia długość.

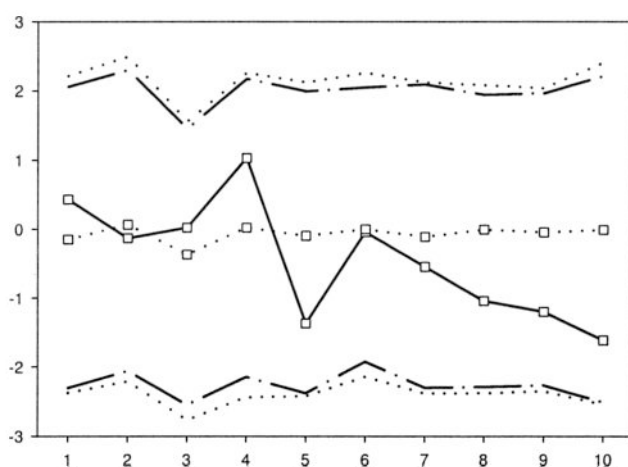
Dodatkowo, w porównaniach uwzględniono również zastosowanie bootstrapu wygładzonego (Podrozdział 1.4). Wyniki doboru optymalnej szerokości okna zawarto w Tabeli 1.6.

Przeanalizowano m.in. wpływ bootstrapu wygładzonego na poprawę stabilności studentyzowanych przedziałów predykcyjnych dla przypadku małych prób ($n=25$). Podobnie jak u Polansky'ego (2000) analiza oparta była na porównaniu średnich długości przedziałów (Tabela 1.4) oraz odchyłeń standardowych długości (Tabela 1.5). Możemy zaobserwować, że w tym przypadku zastosowanie bootstrapu wygładzonego przynosi redukcję odchylenia standardowego długości oraz dla rozkładów błędów innych niż normalne również zmniejszenie średniej długości przedziałów. Ponadto, zaobserwowanej poprawie towarzyszy zachowanie empirycznych prawdopodobieństw pokrycia na podobnym poziomie jak w przypadku standardowej metody bootstrap (Tabela 1.2).

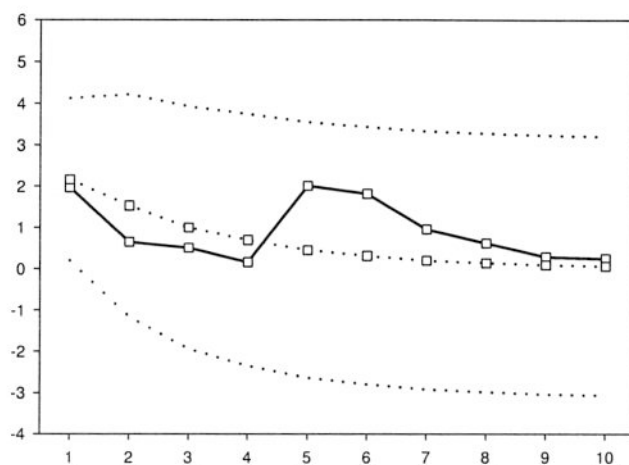
Dodatkowo, możemy zaobserwować, że zastosowanie bootstrapu wygładzonego przynosi poprawę wyników pokrycia empirycznego w przypadku metody *hybrid bootstrap* (Tabela 1.2 i 1.3). Ceną za tą poprawę jest jednak wzrost średniej długości przedziałów (Tabela 1.4).



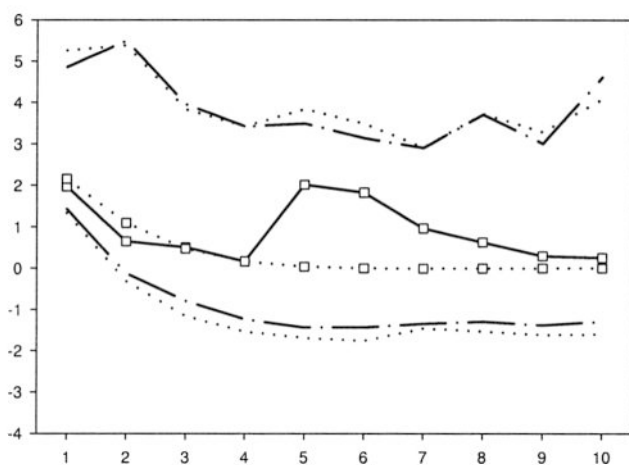
Rys. 1.1: Gaussowskie przedziały predykcyjne dla modelu M1 i rozkładu $N(0,1)$: prawdziwe przyszłe wartości (linia ciągła z kwadratami), prognozy (linia kropkowana z kwadratami), przedziały gaussowskie (linia kropkowana).



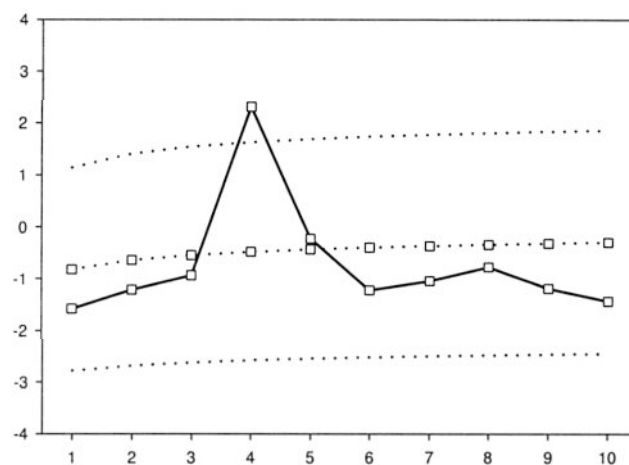
Rys. 1.2: Bootstrapowe przedziały predykcyjne dla M1 i rozkładu $N(0,1)$: prawdziwe przyszłe wartości (linia ciągła z kwadratami), prognozy (linia kropkowana z kwadratami), hybrid sieve bootstrap (linia kreska-kropka), sieve bootstrap-t (linia kropkowana).



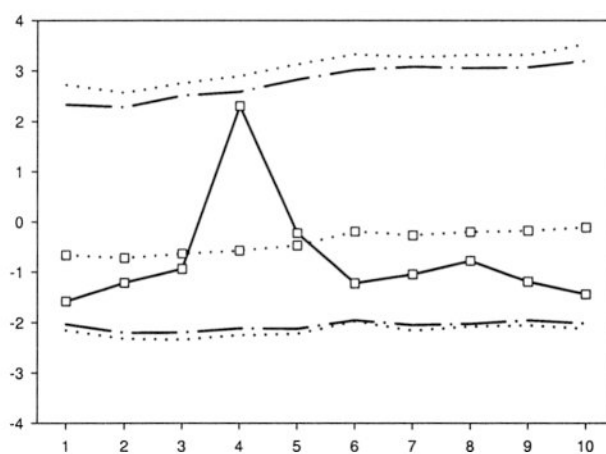
Rys. 1.3: Gaussowskie przedziały predykcyjne dla modelu M2 i rozkładu logarytmiczno-normalnego: prawdziwe przyszłe wartości (linia ciągła z kwadratami), prognozy (linia kropkowana z kwadratami), przedziały gaussowskie (linia kropkowana).



Rys. 1.4: Bootstrapowe przedziały predykcyjne dla M2 i rozkładu logarytmiczno-normalnego: prawdziwe przyszłe wartości (linia ciągła z kwadratami), prognozy (linia kropkowana z kwadratami), hybrid sieve bootstrap (linia kreska-kropka), sieve bootstrap-t (linia kropka-kropka).



Rys. 1.5: Gaussowskie przedziały predykcyjne dla modelu M3 i rozkładu mieszanki: prawdziwe przyszłe wartości (linia ciągła z kwadratami), prognozy (linia kropkowana z kwadratami), przedziały gaussowskie (linia kropkowana).



Rys. 1.6: Bootstrapowe przedziały predykcyjne dla M3 i rozkładu mieszanki: prawdziwe przyszłe wartości (linia ciągła z kwadratami), prognozy (linia kropkowana z kwadratami), hybrid sieve bootstrap (linia kreska-kropka), sieve bootstrap-t (linia kropkowana).

Model	Rozkład	n	estymacja $p(n)$ za pomocą AIC			
			$E(\hat{p}_{AIC})$	$std(\hat{p}_{AIC})$	$min(\hat{p}_{AIC})$	$max(\hat{p}_{AIC})$
M1	N	25	0.88	1.24	0	8
		50	1.52	1.89	0	16
		100	2.35	2.32	0	20
	t	25	0.70	1.06	0	7
		50	1.45	1.67	0	14
		100	2.22	2.23	0	20
	logN	25	0.79	1.21	0	8
		50	1.45	1.78	0	13
		100	2.27	2.16	0	18
	M	25	0.84	1.30	0	10
		50	1.48	1.85	0	12
		100	2.36	2.28	0	18
M2	N	25	1.56	1.03	0	8
		50	2.05	1.54	1	15
		100	2.77	2.12	1	19
	t	25	1.48	0.89	0	7
		50	1.96	1.36	1	12
		100	2.52	1.66	1	16
	logN	25	1.46	0.85	0	8
		50	1.93	1.50	1	14
		100	2.52	1.73	1	14
	M	25	1.54	1.03	0	10
		50	2.09	1.63	1	14
		100	2.65	1.99	1	19
M3	N	100	2.26	2.13	0	20
		500	5.40	3.48	1	26
	t	100	2.23	2.24	0	18
		500	5.47	3.63	1	26
	logN	100	2.06	1.79	0	16
		500	5.57	4.11	1	26
	M	100	2.25	2.26	0	17
		500	5.40	3.60	1	25

Tabela 1.1: Tabela estymacji rzędu $p(n)$ z wykorzystaniem kryterium AIC.

Model	Rozkład	h	Box-Jenkins pokrycie	hybrid bootstrap		bootstrap-t	
				standard	smoothed	standard	smoothed
M1	N	1	94.8%(0.702)	89.1%(0.985)	92.2%(0.848)	91.4%(0.887)	92.6%(0.828)
		2	94.3%(0.733)	90.2%(0.940)	92.3%(0.843)	92.3%(0.843)	92.1%(0.853)
		3	95.6%(0.649)	90.2%(0.940)	92.7%(0.823)	92.2%(0.848)	92.8%(0.817)
		4	94.6%(0.715)	88.5%(1.009)	92.3%(0.843)	92.6%(0.828)	92.9%(0.812)
		5	94.3%(0.733)	89.5%(0.969)	92.5%(0.833)	92.4%(0.838)	92.9%(0.812)
	t	1	95.9%(0.627)	93.1%(0.801)	94.4%(0.727)	94.9%(0.696)	95.0%(0.689)
		2	95.3%(0.669)	92.6%(0.828)	92.8%(0.817)	94.0%(0.751)	93.9%(0.757)
		3	95.0%(0.689)	91.5%(0.882)	91.5%(0.882)	91.9%(0.863)	92.1%(0.853)
		4	96.3%(0.597)	91.8%(0.868)	92.9%(0.812)	93.2%(0.796)	93.3%(0.791)
		5	96.2%(0.605)	92.1%(0.853)	92.9%(0.812)	92.9%(0.812)	93.0%(0.807)
	logN	1	96.6%(0.573)	85.3%(1.120)	89.0%(0.989)	92.5%(0.833)	93.4%(0.785)
		2	95.8%(0.634)	84.3%(1.150)	88.2%(1.020)	91.5%(0.882)	92.3%(0.843)
		3	97.5%(0.494)	85.6%(1.110)	89.1%(0.985)	91.3%(0.891)	92.6%(0.828)
		4	96.2%(0.605)	85.0%(1.129)	89.3%(0.977)	91.9%(0.863)	93.0%(0.807)
		5	95.4%(0.662)	84.6%(1.141)	88.0%(1.028)	90.4%(0.932)	92.2%(0.848)
	M	1	89.7%(0.961)	84.6%(1.141)	87.6%(1.042)	93.5%(0.780)	93.2%(0.796)
		2	90.7%(0.918)	87.1%(1.060)	89.5%(0.969)	95.3%(0.669)	94.9%(0.696)
		3	90.0%(0.949)	86.9%(1.067)	89.4%(0.973)	95.5%(0.656)	94.5%(0.721)
		4	91.6%(0.877)	85.6%(1.110)	88.2%(1.020)	95.1%(0.683)	95.3%(0.669)
		5	89.3%(0.978)	85.0%(1.129)	87.1%(1.060)	94.9%(0.696)	94.9%(0.696)
M2	N	1	94.5%(0.721)	89.3%(0.978)	92.0%(0.858)	92.4%(0.838)	92.9%(0.812)
		2	95.5%(0.656)	85.2%(1.123)	88.3%(1.016)	90.6%(0.923)	90.3%(0.936)
		3	95.0%(0.689)	81.3%(1.233)	85.9%(1.101)	90.2%(0.940)	90.1%(0.944)
		4	93.8%(0.763)	81.7%(1.223)	85.3%(1.120)	89.3%(0.978)	89.4%(0.973)
		5	95.8%(0.634)	82.6%(1.199)	87.3%(1.053)	90.4%(0.932)	90.2%(0.940)
	t	1	95.7%(0.641)	91.6%(0.877)	92.4%(0.838)	94.1%(0.745)	93.7%(0.768)
		2	95.8%(0.634)	85.8%(1.104)	87.0%(1.064)	89.7%(0.961)	90.0%(0.949)
		3	95.5%(0.656)	84.8%(1.135)	86.9%(1.067)	90.8%(0.914)	90.8%(0.914)
		4	94.8%(0.702)	82.7%(1.196)	85.8%(1.104)	90.4%(0.932)	90.7%(0.918)
		5	95.0%(0.689)	84.5%(1.144)	86.1%(1.094)	90.8%(0.914)	90.2%(0.940)
	logN	1	95.9%(0.627)	83.9%(1.162)	88.8%(0.997)	91.4%(0.887)	92.6%(0.828)
		2	95.8%(0.634)	77.0%(1.331)	80.8%(1.246)	87.7%(1.039)	88.4%(1.013)
		3	96.0%(0.620)	76.3%(1.345)	80.0%(1.265)	87.2%(1.057)	87.5%(1.046)
		4	96.2%(0.605)	75.8%(1.354)	78.7%(1.295)	87.1%(1.060)	87.8%(1.035)
		5	97.0%(0.539)	77.3%(1.325)	79.5%(1.277)	87.0%(1.064)	87.4%(1.049)
	M	1	91.7%(0.872)	87.1%(1.060)	89.3%(0.978)	94.7%(0.708)	94.6%(0.715)
		2	93.5%(0.780)	79.1%(1.286)	82.4%(1.204)	93.5%(0.780)	93.4%(0.785)
		3	96.0%(0.620)	76.3%(1.345)	79.1%(1.286)	92.3%(0.843)	91.8%(0.868)
		4	96.3%(0.597)	78.1%(1.308)	80.6%(1.251)	92.3%(0.843)	92.5%(0.833)
		5	96.7%(0.565)	79.1%(1.286)	81.8%(1.220)	92.9%(0.812)	93.0%(0.807)

Tabela 1.2: Tabela pokrycia empirycznego dla n=25, poziom ufności 95%.

Model	Rozkład	h	Box-Jenkins	hybrid bootstrap		bootstrap-t	
				standard	smoothed	standard	smoothed
M1	N	1	94.7%(0.708)	90.8%(0.914)	92.5%(0.833)	93.0%(0.807)	92.4%(0.838)
		2	95.7%(0.641)	92.8%(0.817)	94.0%(0.751)	94.2%(0.739)	93.7%(0.768)
		3	94.4%(0.727)	92.6%(0.828)	93.7%(0.768)	93.9%(0.757)	93.8%(0.763)
		4	94.3%(0.733)	92.5%(0.833)	93.7%(0.768)	93.3%(0.791)	94.0%(0.751)
		5	95.3%(0.669)	92.9%(0.812)	93.5%(0.780)	93.8%(0.763)	93.4%(0.785)
	t	1	96.0%(0.620)	93.0%(0.807)	93.6%(0.774)	95.3%(0.669)	95.0%(0.689)
		2	95.7%(0.641)	93.3%(0.791)	93.6%(0.774)	94.8%(0.702)	94.2%(0.739)
		3	95.8%(0.634)	92.6%(0.828)	93.0%(0.807)	93.8%(0.763)	93.7%(0.768)
		4	95.7%(0.641)	93.7%(0.768)	93.8%(0.763)	95.7%(0.641)	94.7%(0.708)
		5	95.6%(0.649)	93.2%(0.796)	93.8%(0.763)	94.5%(0.721)	94.5%(0.721)
	logN	1	96.7%(0.565)	87.1%(1.060)	89.5%(0.969)	93.4%(0.785)	94.9%(0.696)
		2	95.7%(0.641)	86.7%(1.074)	89.8%(0.957)	93.2%(0.796)	93.6%(0.774)
		3	96.2%(0.605)	88.6%(1.005)	90.5%(0.927)	93.8%(0.763)	94.7%(0.708)
		4	95.4%(0.662)	87.6%(1.042)	89.7%(0.961)	92.8%(0.817)	94.0%(0.751)
		5	97.1%(0.531)	88.5%(1.009)	90.7%(0.918)	93.7%(0.768)	94.8%(0.702)
	M	1	91.3%(0.891)	90.4%(0.932)	92.3%(0.843)	94.8%(0.702)	95.0%(0.689)
		2	90.4%(0.932)	90.3%(0.936)	91.6%(0.877)	95.3%(0.669)	95.3%(0.669)
		3	89.6%(0.965)	89.8%(0.957)	90.8%(0.914)	94.1%(0.745)	94.4%(0.727)
		4	89.4%(0.973)	90.0%(0.949)	91.5%(0.882)	95.6%(0.649)	95.7%(0.641)
		5	89.4%(0.973)	89.4%(0.973)	90.8%(0.914)	94.8%(0.702)	95.3%(0.669)
M2	N	1	94.9%(0.696)	89.8%(0.957)	91.8%(0.868)	92.6%(0.828)	92.8%(0.817)
		2	95.3%(0.669)	88.2%(1.020)	90.3%(0.936)	91.7%(0.872)	91.6%(0.877)
		3	95.0%(0.689)	89.6%(0.965)	90.8%(0.914)	92.5%(0.833)	93.1%(0.801)
		4	94.9%(0.696)	88.2%(1.020)	89.7%(0.961)	92.0%(0.858)	92.0%(0.858)
		5	95.0%(0.689)	88.2%(1.020)	90.5%(0.927)	91.9%(0.863)	92.1%(0.853)
	t	1	96.7%(0.565)	92.3%(0.843)	93.6%(0.774)	94.8%(0.702)	94.9%(0.696)
		2	96.4%(0.589)	91.7%(0.872)	92.2%(0.848)	94.4%(0.727)	94.0%(0.751)
		3	96.1%(0.612)	88.7%(1.001)	90.3%(0.936)	92.5%(0.833)	91.8%(0.868)
		4	95.6%(0.649)	88.1%(1.024)	89.0%(0.989)	91.2%(0.896)	90.8%(0.914)
		5	95.4%(0.662)	89.0%(0.989)	89.5%(0.969)	92.1%(0.853)	92.1%(0.853)
	logN	1	96.2%(0.605)	88.4%(1.013)	91.4%(0.887)	94.6%(0.715)	95.0%(0.689)
		2	95.1%(0.683)	86.2%(1.091)	88.4%(1.013)	93.0%(0.807)	92.7%(0.823)
		3	95.3%(0.669)	85.4%(1.117)	86.0%(1.097)	91.7%(0.872)	92.3%(0.843)
		4	95.2%(0.676)	85.8%(1.104)	86.4%(1.084)	90.6%(0.923)	91.0%(0.905)
		5	96.0%(0.620)	85.2%(1.123)	86.1%(1.094)	90.8%(0.914)	90.5%(0.927)
	M	1	90.0%(0.949)	90.4%(0.932)	92.1%(0.853)	95.3%(0.669)	95.5%(0.656)
		2	93.8%(0.763)	88.3%(1.016)	89.7%(0.961)	95.8%(0.634)	95.6%(0.649)
		3	95.2%(0.676)	84.9%(1.132)	86.5%(1.081)	93.1%(0.801)	93.6%(0.774)
		4	95.8%(0.634)	85.1%(1.126)	86.5%(1.081)	92.3%(0.843)	92.6%(0.828)
		5	94.8%(0.702)	84.1%(1.156)	84.9%(1.132)	92.1%(0.853)	91.7%(0.872)

Tabela 1.3: Tabela pokrycia empirycznego dla n=50, poziom ufności 95%

Model	Rozkład	h	Box-Jenkins	hybrid bootstrap		bootstrap-t	
				standard	smoothed	standard	smoothed
M1	N	1	3.920	3.693	3.910	3.907	3.925
		2	3.998	3.733	3.984	3.993	4.015
		3	4.047	3.768	4.029	4.044	4.067
		4	4.077	3.784	4.043	4.071	4.081
		5	4.097	3.793	4.054	4.083	4.097
	t	1	3.920	4.129	4.154	4.390	4.328
		2	3.998	4.137	4.192	4.466	4.388
		3	4.047	4.129	4.210	4.482	4.415
		4	4.077	4.155	4.220	4.486	4.417
		5	4.097	4.161	4.224	4.498	4.428
	logN	1	3.920	3.743	3.820	4.288	4.274
		2	3.998	3.767	3.859	4.380	4.358
		3	4.047	3.775	3.883	4.419	4.394
		4	4.077	3.790	3.884	4.427	4.412
		5	4.097	3.797	3.903	4.422	4.424
	M	1	3.920	3.728	3.806	4.443	4.411
		2	3.998	3.766	3.853	4.550	4.505
		3	4.047	3.793	3.873	4.595	4.551
		4	4.077	3.804	3.896	4.622	4.570
		5	4.097	3.809	3.907	4.626	4.602
M2	N	1	3.920	3.724	3.969	4.096	4.118
		2	5.373	4.233	4.621	5.051	5.069
		3	5.867	4.420	4.828	5.387	5.399
		4	6.094	4.504	4.924	5.527	5.542
		5	6.190	4.547	4.979	5.597	5.610
	t	1	3.920	4.109	4.164	4.535	4.471
		2	5.373	4.375	4.564	5.467	5.388
		3	5.867	4.520	4.725	5.842	5.728
		4	6.094	4.591	4.810	5.976	5.899
		5	6.190	4.629	4.852	6.071	5.955
	logN	1	3.920	3.494	3.604	4.164	4.152
		2	5.373	3.777	3.938	5.204	5.158
		3	5.867	3.933	4.102	5.637	5.511
		4	6.094	4.020	4.181	5.789	5.642
		5	6.190	4.073	4.231	5.851	5.735
	M	1	3.920	3.700	3.802	4.491	4.490
		2	5.373	4.085	4.273	5.565	5.541
		3	5.867	4.308	4.477	5.960	5.913
		4	6.094	4.419	4.581	6.134	6.087
		5	6.190	4.478	4.648	6.235	6.179

Tabela 1.4: Porównanie średniej długości dla bootstrapu klasycznego i wygładzonego dla $n=25$ i poziomu ufności 95%.

Model	Rozkład	h	hybrid bootstrap		bootstrap-t	
			standard	smoothed	standard	smoothed
M1	N	1	0.755	0.722	0.748	0.697
		2	0.730	0.722	0.748	0.703
		3	0.713	0.715	0.736	0.706
		4	0.710	0.713	0.754	0.708
		5	0.706	0.715	0.754	0.706
	t	1	2.196	2.150	2.481	2.458
		2	2.188	2.133	2.505	2.457
		3	2.090	2.126	2.440	2.425
		4	2.175	2.131	2.495	2.376
		5	2.175	2.124	2.511	2.423
	logN	1	2.704	2.721	3.188	3.125
		2	2.703	2.748	3.332	3.336
		3	2.698	2.756	3.351	3.369
		4	2.698	2.744	3.386	3.369
		5	2.694	2.738	3.340	3.328
	M	1	0.844	0.833	1.013	0.989
		2	0.847	0.850	1.043	1.023
		3	0.855	0.876	1.070	1.058
		4	0.860	0.876	1.095	1.067
		5	0.868	0.876	1.087	1.079
M2	N	1	0.732	0.701	0.792	0.752
		2	0.778	0.848	0.985	0.976
		3	0.842	0.919	1.137	1.127
		4	0.868	0.959	1.203	1.195
		5	0.893	0.972	1.248	1.224
	t	1	2.269	2.212	2.561	2.523
		2	2.221	2.205	3.200	3.053
		3	2.232	2.226	3.529	3.291
		4	2.256	2.270	3.589	3.379
		5	2.266	2.268	3.728	3.490
	logN	1	2.204	2.226	2.874	2.705
		2	2.197	2.246	4.190	3.755
		3	2.235	2.300	4.778	4.126
		4	2.281	2.319	4.917	4.164
		5	2.319	2.356	4.833	4.277
	M	1	0.858	0.872	1.014	1.030
		2	1.084	1.133	1.389	1.397
		3	1.212	1.256	1.582	1.572
		4	1.268	1.306	1.678	1.665
		5	1.303	1.342	1.737	1.719

Tabela 1.5: Porównanie odchylenia standardowego długości przedziałów dla bootstrapu klasycznego i wygładzonego dla $n=25$ i poziomu ufności 95 %.

Model	Rozkład	n	dobór szerokości okna $\hat{h}_{opt}$			
			$E(\hat{h}_{opt})$	$std(\hat{h}_{opt})$	$min(\hat{h}_{opt})$	$max(\hat{h}_{opt})$
M1	N	25	0.8363	0.2678	0.3737	4.4116
		50	0.6238	0.1090	0.3478	1.0458
		100	0.4759	0.0603	0.3020	0.6754
	t	25	0.6115	0.2346	0.2478	5.6813
		50	0.4466	0.0915	0.2120	1.4692
		100	0.3337	0.0439	0.2295	0.6493
	logN	25	0.4308	0.1901	0.1390	3.1441
		50	0.3128	0.0910	0.1509	0.9732
		100	0.2252	0.0490	0.1288	0.5795
	M	25	0.4219	0.1036	0.1503	1.0913
		50	0.3060	0.0540	0.1495	0.6175
		100	0.2231	0.0273	0.1561	0.3725
M2	N	25	0.8664	0.2684	0.2795	3.7220
		50	0.6161	0.1100	0.3349	1.0859
		100	0.4726	0.0601	0.2977	0.6970
	t	25	0.6327	0.1835	0.2507	2.1583
		50	0.4431	0.0829	0.2321	0.9076
		100	0.3331	0.0461	0.2237	0.8995
	logN	25	0.4433	0.1569	0.1421	1.3011
		50	0.3209	0.1004	0.1543	1.2801
		100	0.2249	0.0470	0.1381	0.5201
	M	25	0.4613	0.1412	0.1731	1.4685
		50	0.3130	0.0534	0.1825	0.5334
		100	0.2214	0.0282	0.1414	0.3683

Tabela 1.6: Wyniki doboru optymalnej szerokości okna.

Model	Rozkład	h	Box-Jenkins		hybrid bootstrap		bootstrap-t	
			pokrycie	E(length)	pokrycie	E(length)	pokrycie	E(length)
M3	N	1	80.7%(1.248)	2.564	77.9%(1.312)	2.489	79.2%(1.284)	2.546
		2	81.2%(1.236)	2.678	78.3%(1.304)	2.596	80.2%(1.260)	2.661
		3	82.3%(1.207)	2.725	79.9%(1.267)	2.646	81.2%(1.236)	2.714
		4	79.2%(1.284)	2.752	76.4%(1.343)	2.663	77.7%(1.316)	2.734
		5	80.8%(1.246)	2.770	78.4%(1.301)	2.671	80.1%(1.263)	2.741
	t	1	88.0%(1.028)	2.564	76.4%(1.343)	1.907	78.1%(1.308)	2.012
		2	88.8%(0.997)	2.678	78.2%(1.306)	2.042	79.8%(1.270)	2.162
		3	88.5%(1.009)	2.725	78.2%(1.306)	2.087	80.6%(1.251)	2.213
		4	87.8%(1.035)	2.752	78.9%(1.290)	2.115	80.2%(1.260)	2.241
		5	87.9%(1.031)	2.770	77.2%(1.327)	2.127	78.9%(1.290)	2.257
	logN	1	93.2%(0.796)	2.564	76.6%(1.339)	1.601	81.6%(1.225)	1.741
		2	92.4%(0.838)	2.678	74.1%(1.385)	1.752	79.4%(1.279)	1.913
		3	92.0%(0.858)	2.725	75.1%(1.368)	1.807	78.3%(1.304)	1.970
		4	93.1%(0.801)	2.752	74.2%(1.384)	1.841	79.0%(1.288)	2.010
		5	93.4%(0.785)	2.770	77.0%(1.331)	1.851	80.8%(1.246)	2.022
	M	1	90.7%(0.918)	2.564	79.7%(1.272)	2.099	81.2%(1.236)	2.134
		2	89.4%(0.973)	2.678	77.5%(1.321)	2.309	79.6%(1.274)	2.358
		3	90.4%(0.932)	2.725	77.2%(1.327)	2.372	78.4%(1.301)	2.427
		4	89.7%(0.961)	2.752	76.9%(1.333)	2.403	78.9%(1.290)	2.459
		5	91.9%(0.863)	2.770	78.6%(1.297)	2.404	80.0%(1.265)	2.464

Tabela 1.7: Wyniki pokrycia empirycznego oraz średniej długości przedziałów dla modelu M3, $n=100$ i nominalnego poziomu ufności 80%.

Model	Rozkład	h	Box-Jenkins		hybrid bootstrap		bootstrap-t	
			pokrycie	E(length)	pokrycie	E(length)	pokrycie	E(length)
M3	N	1	95.3%(0.669)	3.922	93.4%(0.785)	3.808	93.6%(0.774)	3.912
		2	95.3%(0.669)	4.095	93.1%(0.801)	3.976	94.0%(0.751)	4.091
		3	95.0%(0.689)	4.167	94.5%(0.721)	4.056	94.8%(0.702)	4.173
		4	95.2%(0.676)	4.209	93.2%(0.796)	4.071	94.1%(0.745)	4.192
		5	95.4%(0.662)	4.237	93.8%(0.763)	4.088	94.3%(0.733)	4.213
	t	1	95.7%(0.641)	3.922	92.6%(0.828)	3.672	94.1%(0.745)	3.937
		2	96.0%(0.620)	4.095	93.3%(0.791)	3.841	94.2%(0.739)	4.158
		3	95.6%(0.649)	4.167	93.9%(0.757)	3.921	95.1%(0.683)	4.232
		4	95.5%(0.656)	4.209	93.6%(0.774)	3.961	94.3%(0.733)	4.280
		5	96.3%(0.597)	4.237	93.7%(0.768)	3.968	94.6%(0.715)	4.293
	logN	1	95.8%(0.634)	3.922	92.3%(0.843)	3.349	95.8%(0.634)	3.799
		2	96.5%(0.581)	4.095	92.5%(0.833)	3.573	96.2%(0.605)	4.034
		3	95.3%(0.669)	4.167	91.5%(0.882)	3.647	95.0%(0.689)	4.117
		4	96.5%(0.581)	4.209	92.7%(0.823)	3.700	95.6%(0.649)	4.178
		5	96.2%(0.605)	4.237	91.3%(0.891)	3.682	95.0%(0.689)	4.184
	M	1	90.7%(0.918)	3.922	94.1%(0.745)	3.949	95.6%(0.649)	4.173
		2	90.1%(0.944)	4.095	93.4%(0.785)	4.065	95.1%(0.683)	4.317
		3	90.9%(0.910)	4.167	93.5%(0.780)	4.122	96.1%(0.612)	4.384
		4	90.7%(0.918)	4.209	93.6%(0.774)	4.157	95.5%(0.656)	4.418
		5	93.2%(0.796)	4.237	93.2%(0.796)	4.169	94.8%(0.702)	4.436

Tabela 1.8: Wyniki pokrycia empirycznego oraz średniej długości przedziałów dla modelu M3, $n=100$ i nominalnego poziomu ufności 95%.

Rozdział 2

Zastosowanie metody sita do estymacji intensywności multiplikatywnych procesów punktowych

2.1 Metoda sita

Wprowadzona przez U. Grenandera w 1981 r. metoda sita stała się od tego czasu popularnym narzędziem estymacji nieparametrycznej.

Różne sita były wykorzystywane przez autorów do estymacji parametrów funkcyjnych procesów stochastycznych, takich jak procesy punktowe należące do multiplikatywnego modelu intensywności, procesy dyfuzyjne czy też modele semimartyn-gałowe. (np. Grenander (1981), Geman i Hwang (1982), Prakasa Rao (1999)).

Przypomnijmy, że rodzina $S(m_n)$ podzbiorów przestrzeni parametrów A jest nazywana sitem jeżeli wymiar m_n rośnie wraz z n oraz $\bigcup_n S(n)$ jest zbiorem gęstym w A .

W naszym przypadku będziemy zakładali, że rodzina $S(m_n)$ jest **sitem histogramowym**, to znaczy:

$$S(n) = \left\{ \alpha \in A : \alpha(s) = \sum_{l=1}^{m(n)} x_l 1_{B_{l,m(n)}}(s) \right\}, \quad (2.1)$$

gdzie $1_{B_{l,m(n)}}$ oznacza indykator zbioru $B_{l,m(n)}$,

$$B_{l,m(n)} = \left(\frac{l-1}{m(n)}, \frac{l}{m(n)} \right] \quad \text{dla } l = 2, \dots, m(n)$$

oraz

$$B_{1,m(n)} = \left[0, \frac{1}{m(n)} \right),$$

gdzie $x_l \geq 0$, $\sum_l x_l^2 > 0$, $s \in [0, 1]$.

Zauważmy, że ciąg $\{B_{l,m(n)}\}$ definiuje partycję odcinka $[0, 1]$ na podprzedziały o długości $1/m(n)$. Ciąg $\{m(n)\}$ odpowiada natomiast za szybkość wzrostu sita $S(m_n)$.

Jak już wspominaliśmy we Wstępie zasadniczą ideą metody sita jest optymalizacja określonego kryterium (np. maksymalizacja funkcji wiarygodności) na pewnym podzbiórze przestrzeni parametrów $S(m_n)$, a następnie pozwolenie aby podzbiór ten „rósł” wraz ze wzrostem rozmiaru próby.

W naszym wypadku estymator największej wiarygodności $\hat{\alpha}_n$ skonstruowany dla sita $S(m_n)$ jest określony następująco:

$$L_n(\hat{\alpha}_n) = \max_{\alpha \in S(m_n)} L_n(\alpha), \quad (2.2)$$

gdzie L_n jest funkcją wiarygodności.

W przypadku sita histogramowego możemy stosunkowo łatwo wyznaczyć dokładną postać estymatora $\hat{\alpha}_n$.

2.2 Przegląd wyników

Karr (1987) zakładając, że obserwujemy ciąg niezależnych procesów $\{N_i\}$, $i = 1, \dots, n$ o multiplikatywnym modelu intensywności:

$$\lambda_i(t) = \alpha(t) Y_i(t),$$

otrzymał zgodny w L^1 estymator funkcji $\alpha(t)$ wykorzystując metodę największej wiarygodności dla sita postaci:

$$S(m_n) = \{\alpha : (1/m_n) \leq \alpha(t) \leq m_n; |\alpha'(t)| \leq m_n \alpha(t) \text{ dla } t \in [0, 1]\}.$$

Leśkow i Różański (1989a) stosując sito histogramowe skonstruowali zgodny i asymptotycznie normalny estymator największej wiarygodności $\hat{\alpha}(t)$ zakładając, że dysponujemy ciągiem n kopii procesów $N_i(t)$, $i = 1, \dots, n$ o multiplikatywnej strukturze intensywności. Założenia, dla których otrzymano wyniki asymptotyczne, w tym m.in. warunki mieszania i oddzielenie od zera wartości oczekiwanej procesu $Y_i(t)$, mogą jednak w praktyce okazać się trudne do zweryfikowania.

Leśkow (1989) zaproponował konstrukcję wygładzonego metodą jądrową estymatora histogramowego

$$\hat{\eta}_n(t) = (1/b_n) \int_0^1 K((t-s)/b_n) \hat{\alpha}(s) ds.$$

przy założeniu, że α jest funkcją okresową o znanym okresie. Przy podobnych założeniach jak w pracy Leśkowskiego i Różańskiego (1989a) otrzymano twierdzenia o zgodności i asymptotycznym rozkładzie skonstruowanego estymatora.

Leśkow (1993) rozważył także konstrukcję estymatora największej wiarygodności w oparciu o metodę sita dla prawie-okresowych modeli procesów stochastycznych.

2.3 Estymacja intensywności w modelu multiplikatywnym

Aalen (1978) wprowadził tak zwany multiplikatywny model intensywności, pokazując tym samym, w jaki sposób teoria wielowymiarowych procesów punktowych pozwala na określenie ogólnej struktury, dzięki której możemy analizować zarówno modele związane z klasyczną analizą przeżycia jak i bardziej złożone modele procesów Markowa. Przedstawimy teraz pokrótce założenia modelu Aalena.

Niech $(\Omega, \mathcal{F}, \mathcal{P})$ będzie przestrzenią probabilistyczną, na której określony jest ciąg procesów punktowych $\{N_n(t), t \in [0, 1], n \in N\}$. Będziemy zakładali, że procesy te są adaptowane do filtracji $\{\mathcal{F}_{n,t}, t \in [0, 1], n \in N\}$. Rozważamy takie modele, dla których intensywność stochastyczna $\alpha_n(t)$ procesu $N_n(t), t \in [0, 1]$ istnieje i może być zdefiniowana jako:

$$\alpha_n(t) = \lim_{h \rightarrow 0} \left(\frac{1}{h} \mathcal{P}(N_n(t+h) - N_n(t) \geq 1 | \mathcal{F}_{n,t-}) \right).$$

Zakładamy, że proces punktowy $N_n(t), t \in [0, 1]$, należy do klasy multiplikatywnych modeli intensywności, a zatem funkcja intensywności $\alpha_n(t)$ ma postać:

$$\alpha_n(t) = Y_n(t) \alpha_0(t), \quad (2.3)$$

gdzie $Y_n(t)$ jest nieujemnym, lewostronnie ciągłym procesem, o prawostronnych granicach (lub bardziej ogólnie procesem przewidywalnym), α_0 jest natomiast nieznaną funkcją nieujemną, którą będziemy estymowali. Procesy N_n są więc zdefiniowane jako:

$$N_n(t) = \int_0^t \alpha(s) Y_n(s) ds + M_n(t), \quad t \in [0, 1],$$

gdzie $M_n(t)$ jest martyngałem względem filtracji $\{\mathcal{F}_{n,t}\}$ (lub bardziej ogólnie lokalnym martyngałem).

Niech $\mathcal{P}^n = \{\mathcal{P}_\alpha^n, \alpha \in A\}$ będzie zbiorem miar probabilistycznych odpowiadających procesom punktowym $\{N_n(t), t \in [0, 1], n \in N\}$. Jako przestrzeń A możemy przyjąć podzbiór $L_+^p[0, 1]$. Wiadomo (Liptser i Shirayev (1981) lub Prakasa Rao (1999)), że gęstość miary $\mathcal{P}_\alpha^n$ w odniesieniu do $\mathcal{P}$ -miary generowanej przez proces Poissona o jednostkowej intensywności jest następującej postaci:

$$\frac{d\mathcal{P}_\alpha^n}{d\mathcal{P}} = \exp \left\{ \int_0^1 \log(\alpha(s) Y_n(s)) dN_n(s) + \int_0^1 (1 - \alpha(s) Y_n(s)) ds \right\}. \quad (2.4)$$

Logarytm prawej strony (2.4) będziemy oznaczali przez $L_n(\alpha)$.

Estymator największej wiarygodności $\hat{\alpha}_n$ skonstruowany dla sita $S(m_n)$ jest zdefiniowany przez następujące równanie:

$$L_n(\hat{\alpha}_n) = \max_{\alpha \in S(m_n)} L_n(\alpha). \quad (2.5)$$

Zauważmy, że funkcja wiarygodności $L_n(\alpha)$ dla sita histogramowego $S(m_n)$ będzie funkcją $m(n)$ zmiennych $L_n(x_1, \dots, x_{m(n)})$. Stosując standardowe obliczenia (podobnie jak Leśkow i Różański (1989a)) możemy wyznaczyć postać estymatora $\hat{\alpha}_n$.

Lemat 2.3.1 (Estymator histogramowy)

Estymator największej wiarygodności $\hat{\alpha}_n$ wyznaczony dla sita histogramowego $S(m_n)$ ma następującą postać:

$$\hat{\alpha}_n(s) = \sum_{l=1}^{m(n)} \frac{N_n(B_{l,m(n)})}{\int_{B_{l,m(n)}} Y_n(u) du} 1_{C_{l,m(n)}} 1_{B_{l,m(n)}}(s), \quad (2.6)$$

gdzie $C_{l,m(n)} = \{\int_{B_{l,m(n)}} Y_n(u) du > 0\}$, $l = 1, \dots, m(n)$.

Dla ustalonego $s \in [0, 1]$ oznaczmy przez $l(n, s) \in \{1, \dots, m(n)\}$ taką liczbę, że $s \in B_{l(n,s),m(n)}$. Kładąc $B_{m(n)} = B_{l(n,s),m(n)}$ i $C_{m(n)} = C_{l(n,s),m(n)}$ estymator (2.6) może być zapisany w następującej, prostszej postaci:

$$\hat{\alpha}_n(s) = \frac{N_n(B_{m(n)})}{\int_{B_{m(n)}} Y_n(u) du} 1_{C_{m(n)}}. \quad (2.7)$$

Uwaga 2.3.1 (Wygładzony estymator histogramowy)

Możemy również rozważyć wygładzenie otrzymanego estymatora histogramowego $\hat{\alpha}_n(t)$ wykorzystując metodę jądrową (podobnie jak Leśkow (1989)). Procedura ta prowadzi do estymatora postaci:

$$\hat{\eta}_n(t) = (1/b_n) \int_0^1 K((t-s)/b_n) \hat{\alpha}_n(s) ds, \quad (2.8)$$

gdzie $K(\cdot)$ - funkcja jądrowa, b_n - szerokość okna.

2.4 Zgodność i asymptotyczna normalność histogramowego estymatora sitowego.

Będziemy zakładali dalej, że spełnione są niektóre z następujących warunków:

(C.1) Istnieje $\delta > 0$ oraz funkcja $y : [0, 1] \rightarrow R_+$, dodatnia i ciągła w s , taka że:

$$\sup_{t \in [s-\delta, s+\delta]} \left| \frac{Y_n(t)}{n} - y(t) \right| \xrightarrow{P} 0.$$

($\xrightarrow{P}$ oznacza zbieżność według prawdopodobieństwa).

(C.2) Funkcja $\alpha_0(s)$ jest ciągła na $[0, 1]$.

(C.3) Szybkość wzrostu sita $m_n = n^{1/2}$.

(C.4) Istnieje $\delta > 0$, $\alpha > \frac{1}{2}$ oraz dodatnia stała $C(\alpha_0, s)$, taka że:
 $\forall_{t \in [s-\delta, s+\delta]} |\alpha_0(s) - \alpha_0(t)| \leq C(\alpha_0, s) |t - s|^\alpha$.

Twierdzenie 2.4.1 (Zgodność estymatora histogramowego)

Jeżeli spełnione są warunki (C.1) do (C.3) wówczas estymator największej wiarygodności $\hat{\alpha}_n(s)$ określony w (2.6) jest zgodny dla każdego $s \in [0, 1]$, to znaczy:

$$\hat{\alpha}_n(s) \xrightarrow{P} \alpha_0(s)$$

dla dowolnego $s \in [0, 1]$ gdy $n \rightarrow \infty$.

DOWÓD TWIERDZENIA 2.4.1

Wykorzystując założenie o multiplikatywnym modelu intensywności oraz dekompozycję Dooba-Meyera możemy przedstawić proces $N_n(t)$ jako:

$$N_n(t) = \int_0^t \alpha_0(u) Y_n(u) du + M_n(t), \quad t \in [0, 1], \quad n \in N,$$

gdzie $M_n(t)$ jest martyngałem względem filtracji $\mathcal{F}_{n,t}$.

Przedstawmy różnicę $\hat{\alpha}_n(s) - \alpha_0(s)$, $s \in [0, 1]$ w następującej postaci:

$$\begin{aligned} \hat{\alpha}_n(s) - \alpha_0(s) &= \\ &= \frac{\frac{1}{\sqrt{n}} M_n(B_{m(n)}) + \frac{1}{\sqrt{n}} \int_{B_{m(n)}} (\alpha_0(u) - \alpha_0(s)) Y_n(u) du}{\frac{1}{\sqrt{n}} \int_{B_{m(n)}} Y_n(u) du}, \end{aligned} \quad (2.9)$$

gdzie $M_n(B_{m(n)})$ jest przyrostem martyngału M_n na przedziale $B_{m(n)}$.

Sformułujemy i udowodnimy następnie lemat pomocniczy.

Lemat 2.4.1

$$\text{Ciąg } \frac{1}{\sqrt{n}} M_n(B_{m(n)}) \xrightarrow{P} 0 \quad \text{gdy } n \rightarrow \infty.$$

DOWÓD LEMATU 2.4.1

Rozważmy ciąg martyngałów $\frac{1}{\sqrt{n}} M_n(t)$, $t \in [0, 1]$. Ciąg ten jest również ciągiem elementów losowych przestrzeni $\mathcal{D}([0, 1])$ (przestrzeń funkcji prawostronnie ciągłych posiadających lewostronne granice z topologią Skorohoda).

Na mocy twierdzenia Rebolledo (Rebolledo (1980)) ciąg $\frac{1}{\sqrt{n}} M_n(\cdot)$ jest zbieżny według rozkładu w $\mathcal{D}([0, 1])$ do elementu $\tilde{M}(\cdot)$, który jest ciągłym gaussowskim martyngałem o niezależnych przyrostach, dla którego $E\tilde{M}^2(t) = \int_0^t \alpha(u) y(u) du$.

Wykorzystując dalej Tw.5.5 (Billingsley (1968)) o słabej zbieżności ciągu odwzorowań ciągłych elementów losowych otrzymujemy, że:

$$\frac{1}{\sqrt{n}} M_n(B_{m(n)}) = \frac{1}{\sqrt{n}} (M_n(t_{m(n)+1}(s)) - M_n(t_{m(n)}(s))) \xrightarrow{D} 0,$$

gdzie $B_{m(n)} = (t_{m(n)}(s), t_{m(n)+1}(s)]$, $t_{m(n)}(s) \nearrow s$, $t_{m(n)+1}(s) \searrow s$, co kończy dowód Lematu 2.4.1

□

Powróćmy teraz do dowodu Tw.2.4.1.

Z założeń (C.1), (C.2) oraz (C.3) wynika, że:

$$\frac{1}{\sqrt{n}} \int_{B_{m(n)}} Y_n(u) du \xrightarrow{P} y(s), \quad (2.10)$$

oraz

$$\frac{1}{\sqrt{n}} \int_{B_{m(n)}} (\alpha_0(u) - \alpha_0(s)) Y_n(u) du \xrightarrow{P} 0, \quad (2.11)$$

Stosując Lemat 2.4.1 oraz (2.10) i (2.11) otrzymujemy ostatecznie, że (2.9) zbiega do zera według prawdopodobieństwa, co kończy dowód Tw.2.4.1.

□

Przyjmując założenia C1–C4 możemy otrzymać rezultat związany z asymptotycznym rozkładem $\hat{\alpha}_n$.

Twierdzenie 2.4.2 (Asymptotyczna normalność)

Niech $\{s_1, \dots, s_p\}$ będzie dowolnym skończonym zbiorem punktów z odcinka $[0, 1]$. Jeżeli spełnione są warunki (C.1) do (C.4) wtedy ciąg wektorów losowych

$$n^{1/4}(\hat{\alpha}_n(s_1) - \alpha_0(s_1), \dots, \hat{\alpha}_n(s_p) - \alpha_0(s_p))$$

zbiega według rozkładu, gdy $n \rightarrow \infty$, do p -wymiarowego rozkładu normalnego o średniej zero i diagonalnej macierzy kowariancji Σ , o elementach na przekątnej $\sigma_i = \frac{\alpha_0(s_i)}{y(s_i)}$, $i = 1, \dots, p$.

DOWÓD TWIERDZENIA 2.4.2

Aby udowodnić Tw.2.4.2 pokażemy najpierw zbieżność rozkładów jednowymiarowych.

Możemy zapisać:

$$\begin{aligned} n^{1/4}(\hat{\alpha}_n(s) - \alpha_0(s)) &= \\ &= \frac{n^{1/4} \left(\frac{1}{\sqrt{n}} M_n(B_{m(n)}) \right) + n^{1/4} \left(\frac{1}{\sqrt{n}} \int_{B_{m(n)}} (\alpha_0(u) - \alpha_0(s)) Y_n(u) du \right)}{\frac{1}{\sqrt{n}} \int_{B_{m(n)}} Y_n(u) du}. \end{aligned} \quad (2.12)$$

Pokazaliśmy już (wzór 2.10), że mianownik (2.12) zbiega według prawdopodobieństwa do $y(s)$. Zauważmy, że założenie (C.4) implikuje, że drugi składnik w liczniku (2.12) jest zbieżny do zera według prawdopodobieństwa.

Na mocy wspomnianego już Tw. Rebolledo (Rebolledo (1980)) ciąg $\frac{1}{\sqrt{n}} M_n(\cdot)$ jest zbieżny według rozkładu w $D([0, 1])$ do elementu $\tilde{M}(\cdot)$, który jest ciągłym gaussowskim martyngałem o niezależnych przyrostach.

Oznaczmy odpowiednio przez Q_n i Q miary generowane przez $\frac{1}{\sqrt{n}} M_n(\cdot)$ oraz $\tilde{M}(\cdot)$ w $D([0, 1])$.

Niech $\mathcal{A}$ oznacza następującą klasę podzbiorów $D([0, 1])$:

$$\mathcal{A} = \{A \subset D([0, 1]) : A = \bigcap_k \{x : x(t_k(s)) \leq x_k\} \cap \bigcap_l \{x : x(t_l(s)) \leq y_l\}\},$$

gdzie $\{t_k(s)\}$ i $\{t_l(s)\}$ są skończonymi lub nieskończonym podzbiarami zbiorów $\{t_{m(n)}(s)\}$ i $\{t_{m(n)+1}(s)\}$, dla dowolnych liczb rzeczywistych $\{x_k\}$, $\{y_l\}$.

Na mocy Tw.3 (Topsøe (1967)) otrzymujemy, że $\mathcal{A}$ jest klasą Q -jednostajną.

Otrzymaliśmy zatem, iż:

$$\sup_{A \in \mathcal{A}} |Q_n(A) - Q(A)| \longrightarrow 0.$$

Łatwo zauważyć, że σ -algebra $\sigma(\mathcal{A})$ generowana przez klasę $\mathcal{A}$ jest również klasą Q -jednostajną.

Niech F_n i G_n oznaczają dystrybuanty rozkładów $\frac{1}{\sqrt{n}}M_n(B_{m(n)})$ oraz $\tilde{M}(B_{m(n)})$. Ponieważ $\sigma(\frac{1}{\sqrt{n}}M_n(t_{m(n)}(s)); \frac{1}{\sqrt{n}}M_n(t_{m(n)+1}(s)), n \geq 1) \subset \sigma(\mathcal{A})$ jest klasą Q -jednostajną, otrzymujemy że

$$\sup_x |F_n(x) - G_n(x)| \longrightarrow 0. \quad (2.13)$$

Oczywiście, $n^{\frac{1}{4}}(\tilde{M}(B_{m(n)})) \xrightarrow{D} U$, gdzie U jest zmienną losową o rozkładzie normalnym o wartości oczekiwanej zero i wariancji równej $\alpha_0(s)y(s)$.

A zatem z (2.13) wynika, że również: $n^{\frac{1}{4}}(\frac{1}{\sqrt{n}}M_n(B_{m(n)})) \xrightarrow{D} U$.

Pokazaliśmy więc, że licznik w (2.12) zbiega według rozkładu do zmiennej losowej U oraz, że mianownik w (2.12) zbiega według prawdopodobieństwa do $y(s)$, co oznacza asymptotyczną normalność o średniej zero i wariancji równej $\frac{\alpha_0(s)}{y(s)}$.

Stosując Lemat Śluckiego oraz Twierdzenia Craméra-Wolda otrzymujemy tezę Tw.2.4.2.

□

Uwaga 2.4.1 (Wybór ciągu $m(n)$)

Zauważmy, że założenie (C.3) może być zmienione. Ciąg $m_n = n^{1/2}$ możemy zastąpić ciągiem m_n zmierzającym do nieskończoności wolniej niż $n^{1/2}$. Dla takiego wyboru $m(n)$ otrzymujemy zgodność estymatora $\hat{\alpha}_n(s)$.

Aby otrzymać twierdzenie o rozkładzie asymptotycznym estymator $\hat{\alpha}_n(s)$ musimy przyjąć dodatkowe założenia o gładkości estymowanej funkcji α_0 oraz zastąpić ciąg $n^{1/4}$ (w Tw.2.4.2) ciągiem normalizacyjnym $(\frac{n}{m_n})^{1/2}$.

Zakładając na przykład, że funkcja α_0 jest różniczkowalna w s oraz m_n jest takie, że $\frac{n^{1/2}}{m_n} \rightarrow \infty$, $\frac{n}{m_n^3} \rightarrow 0$ otrzymujemy, że $\hat{\alpha}_n(s)$ jest asymptotycznie normalny $AN(\alpha_0(s), \frac{m_n}{n} \frac{\alpha_0(s)}{y(s)})$.

2.5 Przypadek wielowymiarowy

Przedstawiona w poprzednich rozdziałach metoda estymacji może być uogólniona na przypadek wielowymiarowy. Wprowadźmy teraz niezbędne oznaczenia i definicje.

Definicja 2.5.1 (Wielowymiarowy proces punktowy)

Proces stochastyczny k -wymiarowy $\mathcal{N} = [\{N_1(t), N_2(t), \dots, N_k(t)\}, t \in [0, 1]]$, który zlicza występowanie k różnych rodzajów zdarzeń będziemy nazywali wielowymiarowym procesem punktowym jeżeli:

- wszystkie składowe $N_j, j = 1, \dots, k$ są procesami punktowymi,
- procesy składowe nie mają skoków w tym samym momencie.

Niech $\mathcal{N}_n = (N_{n,1}, \dots, N_{n,k})$ będzie ciągiem k -wymiarowych procesów punktowych o procesie intensywności:

$$\lambda_n(t) = \{\lambda_{n,1}(t), \dots, \lambda_{n,k}(t)\}, \quad t \in [0, 1].$$

Założmy ponadto, że λ_n należy do klasy multiplikatywnych modeli intensywności, tzn.

$$\lambda_{n,h}(t) = \alpha_h Y_{n,h}(t), \quad h = 1, \dots, k, \quad t \in [0, 1],$$

gdzie $\alpha_{0,h}(t)$ jest nieujemną funkcją deterministyczną a $Y_{n,h}$ jest procesem przewidywalnym jak opisano w rozdziale 2.3.

Określając proces intensywności stochastycznej dla ciągu wielowymiarowych procesów punktowych $\tilde{N}_n = (N_{n,1}, \dots, N_{n,k})$ jako:

$$\lambda_n(t) = \{\lambda_{n,1}(t), \dots, \lambda_{n,k}(t)\}, \quad t \in [0, 1]$$

będziemy zakładali, że spełnione są założenia modelu multiplikatywnego, tzn.

$$\lambda_{n,h}(t) = \alpha_h Y_{n,h}(t), \quad \text{dla } h = 1, \dots, k \quad t \in [0, 1].$$

Wiadomo (Andersen (1993)), że logarytm funkcji wiarygodności możemy przedstawić następująco:

$$L_n(\alpha_1, \dots, \alpha_k) = \sum_{h=1}^k \left\{ \int_0^1 \log(\alpha_h(s) Y_{n,h}(s)) dN_{n,h}(s) - \int_0^1 \alpha_h(s) Y_{n,h}(s) ds \right\}.$$

Możemy teraz wyznaczyć estymator największej wiarygodności $\hat{\alpha}_n$.

Lemat 2.5.1

Estymator największej wiarygodności $\hat{\alpha}_n = (\hat{\alpha}_{n,1}, \dots, \hat{\alpha}_{n,k})$ skonstruowany dla sita histogramowego $S(n)$ jest następującej postaci:

$$\hat{\alpha}_{n,h}(s) = \sum_{l=1}^{m(n)} \frac{N_{n,h}(B_{l,m(n)})}{\int_{B_{l,m(n)}} Y_{n,h}(u) du} 1_{C_{l,m(n)}} 1_{B_{l,m(n)}}(s), \quad h = 1, \dots, k \quad (2.14)$$

gdzie $C_{l,m(n)} = \{\int_{B_{l,m(n)}} Y_{n,h}(u) du > 0\}$, $l = 1, \dots, m(n)$.

Dla ustalonego $s \in [0, 1]$ możemy przedstawić estymator (2.14) w prostszej postaci. Oznaczając $B_{m(n)} = B_{l(n,s),m(n)}$ i $C_{m(n)} = C_{l(n,s),m(n)}$, gdzie $l(n, s) \in \{1, \dots, m(n)\}$ jest takie, że $s \in B_{l(n,s),m(n)}$ estymator $\hat{\alpha}_{n,h}(s)$ możemy zapisać jako:

$$\hat{\alpha}_{n,h}(s) = \frac{N_{n,h}(B_{m(n)})}{\int_{B_{m(n)}} Y_{n,h}(u) du} 1_{C_{m(n)}}. \quad (2.15)$$

Przyjmując założenia analogiczne do tych z przypadku jednowymiarowego (rozdział 2.3) możemy otrzymać twierdzenia związane ze zgodnością i asymptotyczną normalnością histogramowego estymatora sitowego (2.14).

(D.1) Istnieje $\delta_h > 0$ oraz funkcja $y_h : [0, 1] \rightarrow R_+$, dodatnia i ciągła w s , taka że:

$$\sup_{t \in [s-\delta_h, s+\delta_h]} \left| \frac{Y_{n,h}(t)}{n} - y_h(t) \right| \xrightarrow{P} 0.$$

(D.2) Funkcja $\alpha_{0,h}(s)$ jest ciągła na $[0, 1]$.

(D.3) Szybkość wzrostu sita histogramowego $m_n = n^{1/2}$.

(D.4) Istnieje $\delta_h > 0$, $\beta > \frac{1}{2}$ oraz stała dodatnia $C(\alpha_{0,h}, s)$, taka że:

$$\forall t \in [s-\delta_h, s+\delta_h] \quad |\alpha_{0,h}(s) - \alpha_{0,h}(t)| \leq C(\alpha_{0,h}, s) |t - s|^\beta.$$

Twierdzenie 2.5.1 (Zgodność)

Jeżeli spełnione są założenia (D.1) – (D.3) wówczas estymator największej wiarygodności $\hat{\alpha}_n(s) = (\hat{\alpha}_{n,1}(s), \dots, \hat{\alpha}_{n,k}(s))$ zdefiniowany jak w (2.14) jest zgodny dla dowolnego $s \in [0, 1]$.

Twierdzenie 2.5.2 (Asymptotyczna normalność)

Niech $s \in [0, 1]$. Jeżeli zachodzą warunki (D.1) do (D.4) wówczas ciąg wektorów losowych

$$n^{1/4}(\hat{\alpha}_{n,1}(s) - \alpha_{0,1}(s), \dots, \hat{\alpha}_{n,k}(s) - \alpha_{0,k}(s))$$

zbiega według rozkładu, gdy $n \rightarrow \infty$, do k -wymiarowego rozkładu normalnego o średniej zero i diagonalnej macierzy kowariancji Σ , o elementach $\sigma_i = \frac{\alpha_{0,i}(s)}{y_i(s)}$, $i = 1, \dots, k$.

2.6 Przykłady

- (i) **Estymacja funkcji hazardu w modelu bez cenzurowania obserwacji.**
Niech $\{X_1, \dots, X_n\}$, $n \in N$ będzie ciągiem nieujemnych zmiennych losowych o jednakowym rozkładzie, określonym przez dystrybuantę F , funkcję gęstości

f oraz funkcję hazardu $\alpha_0(t) = \frac{f(t)}{1-F(t)}$. Możemy założyć, że X_i są ciągiem i.i.d. (podobnie jak zakładali Ramlau-Hansen (1983) i Aalen (1978)). Niech

$$N_n(t) = \sum_{k=1}^n 1(X_k \leq t).$$

Wtedy intensywność stochastyczna procesu $N_n(t)$ jest postaci:

$$\alpha_n(t) = \alpha_0(t) \sum_{k=1}^n 1(X_k \geq t).$$

W tym przypadku wszystkie założenia naszych twierdzeń są spełnione i możemy konstruować zgodny i asymptotycznie normalny estymator histogramowy α_0 .

(ii) **Cenzurowanie.**

Założmy, że zmienne losowe $X_1, X_2, \dots, X_n$, $n \in N$ są określone jak powyżej oraz dodatkowo niech $\{T_i\}$ będzie ciągiem zmiennych losowych obcinających, to znaczy: albo obserwujemy X_i albo obserwujemy jedynie, że X_i jest większe niż T_i . W tym modelu cenzurowania, zakłada się, że momenty obcinające $\{T_i\}_{i=1}^n$ są ciągiem i.i.d. oraz dodatkowo, że są niezależne od $\{X_i\}$. Wprowadzając zmienną losową D_i (nazywaną indykatorem obciążenia), która przyjmuje wartości 0-1 dysponujemy obserwacjami zmiennych losowych $(\tilde{X}_i, D_i)$, gdzie $\tilde{X}_i = X_i$ jeżeli $D_i = 1$ oraz $\tilde{X}_i < X_i$ jeżeli $D_i = 0$. Niech

$$Y_n(t) = \sum_{i=1}^n 1(\tilde{X}_i \geq t)$$

oraz

$$N_n(t) = \sum_{i=1}^n 1(\tilde{X}_i \leq t, D_i = 1).$$

Funkcja intensywności $\alpha_n(t)$ procesu $N_n(t)$ jest postaci:

$$\alpha_n(t) = Y_n(t) \alpha_0(t)$$

i możemy konstruować estymator histogramowy $\hat{\alpha}_n$ dla α_0 .

(iii) **Cenzurowanie II-go rodzaju.** (Przykład III.2.2, Andersen (1993)).

$X_1, X_2, \dots, X_n$ – i.i.d. jak w przykładzie (i). W tym przypadku, przyjmujemy jako zmienne losowe obcinające $T_i = X_{r(n)}$, gdzie $X_{r(n)}$ oznacza moment $r(n)$ -tej awarii. Tak określone cenzurowane zmienne losowe $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ nie są niezależne, niemniej jednak dalej pozostajemy w klasie multiplikatywnych modeli intensywności.

(iv) **Wielostanowy model Markowa.**

Założmy, że obserwujemy n ubezpieczonych (n polis) reprezentujących grupę jednorodną i że każda polisa opisana jest skończonym zbiorem stanów $J = \{1, 2, \dots, j\}$. Wówczas przyjmujemy $X_k(t)$, $k = 1, \dots, n$ jako niezależne łańcuchy Markowa określające stan k -tej polisy.

Definiujemy ciąg procesów punktowych:

N_{ijk} = ilość bezpośrednich przejść ze stanu i do j ($i \mapsto j, i \neq j$) obserwowanych dla k -tej polisy w przedziale $[0, t]$.

Przyjmujemy $Y_{ik} = 1$ jeżeli k -ta polisa jest w stanie i w chwili t^- (tzn. bezpośrednio przed t), w przeciwnym wypadku $Y_{ik} = 0$:

$$Y_{ik}(t) = 1_{\{X_k(t^-)=i\}}.$$

Wówczas proces $N_{ijk}(t)$ ma funkcję intensywności:

$$\lambda_{ijk} = \alpha_{ij}(t) Y_{ik}(t),$$

gdzie α_{ij} jest intensywnością przejścia ze stanu i do stanu j ($i \neq j$). Analogicznie jak poprzednio definiujemy proces:

$$N_{ij\cdot} = \sum_{k=1}^n N_{ijk}(t), \quad i, j \in J$$

określający w tym wypadku całkowitą liczbę przejść $i \mapsto j$ w przedziale $[0, t]$ oraz proces

$$Y_{i\cdot} = \sum_{k=1}^n Y_{ik}(t)$$

interpretowany jako łączna liczba polis znajdujących się w stanie i w chwili t^- . Wówczas z uwagi na to, że żadne dwa przejścia (zmiany stanów) nie występują równocześnie oraz ze względu na przyjętą jednorodność populacji otrzymujemy, że proces $N_{ij\cdot}(t)$ ma funkcję intensywności w postaci

$$\lambda_{ij}(t) = \alpha_{ij}(t) Y_{i\cdot}(t).$$

Warto zauważyć, że wiele modeli spotykanych w teorii niezawodności czy też analizie przeżycia jest szczególnym przypadkiem przedstawionego powyżej wielostanowego modelu Markowa. Najprostszym przykładem jest przedstawiony w (ii) klasyczny model przeżycia, w którym występują tylko dwa stany (życie – śmierć). Inne modele to np. model ryzyk konkurujących (*competing risks*) lub tzw. model choroby - śmierci (*illness-death*). Szczegółowy przegląd różnych modeli można znaleźć np. u Andersena i Borgana (1985) lub w monografii Andersena i innych (1993).

- (v) **Estymacja funkcji hazardu dla wymienialnych zmiennych losowych.** Niech $\{X_1, \dots, X_n\}$ będzie ciągiem zmiennych losowych o jednakowym rozkładzie, warunkowo niezależnych pod warunkiem pewnej σ -algebry $\mathcal{B}$. Przy tych założeniach $\{X_i\}$ tworzy ciąg wymienialnych zmiennych losowych.

Określając teraz

$$\tilde{N}_i(t) = 1(X_i \leq t)$$

oraz

$$N_n(t) = \sum_{i=1}^n \tilde{N}_i(t)$$

i wykorzystując dalej rozważania związane z łączeniem składowych warunkowo niezależnych (Rozdział II.4.3, Andersen (1993)) otrzymujemy, że $N_n(t)$ ma multiplikatywną postać intensywności.

2.7 Symulacje komputerowe

Aby zilustrować zachowanie estymatora histogramowego intensywności przeprowadziliśmy symulacje komputerowe. Wygenerowano $n = 500$ niezależnych czasów życia $X_1, \dots, X_n$ z rozkładów Weibulla, logarytmiczno-normalnego, gamma oraz Gompertza. Rozkłady te są szczególnie popularne w analizie przeżycia (np. Klein i Moeschberger (1998)).

Funkcje hazardu i parametry wykorzystywane w symulacjach są przedstawione w poniższej tabeli. I oznacza niekompletną funkcję gamma.

Rozkład	Funkcja hazardu	Parametry
Weibulla	$\gamma \rho t^{\rho-1}$	$\gamma = 2, \rho = 0.5$
Log-normalny	$\frac{\exp[-\frac{1}{2}(\frac{\ln t - \mu}{\sigma})^2]}{\sqrt{2\pi}\sigma t(1 - \Phi(\frac{\ln t - \mu}{\sigma}))}$	$\mu = 0, \sigma = 1$
Gompertza	$\theta e^{\alpha t}$	$\theta = 0.2, \alpha = 2$
Gamma	$\frac{\lambda^\beta t^{\beta-1} \exp(-\lambda t)}{\Gamma(\beta)(1 - I(\lambda t, \beta))}$	$\lambda = 1, \beta = 2$

Dodatkowo, symulowane czasy życia były obcinane przez n niezależnych zmiennych losowych $T_1, \dots, T_n$ (momenty obcięć) generowanych z rozkładu wykładniczego o średniej 1.

Przyjmujemy $\tilde{X}_i = \min(X_i, T_i)$ oraz indykator obciążenia $D_i = 1(\tilde{X}_i = X_i)$ dla $i = 1, \dots, n$. Wówczas proces punktowy $N_n(\cdot)$ jest określony przez

$$N_n(t) = \sum_{i=1}^n 1(\tilde{X}_i \leq t, D_i = 1).$$

Proces ten możemy interpretować jako całkowitą liczbę obserwowanych śmierci (awarii) na przedziale $[0, t]$, a jego intensywność α_n jest postaci:

$$\alpha_n(t) = \alpha_0(t)Y_n(t),$$

gdzie

$$Y_n(t) = \sum_{i=1}^n 1(\tilde{X}_i \geq t)$$

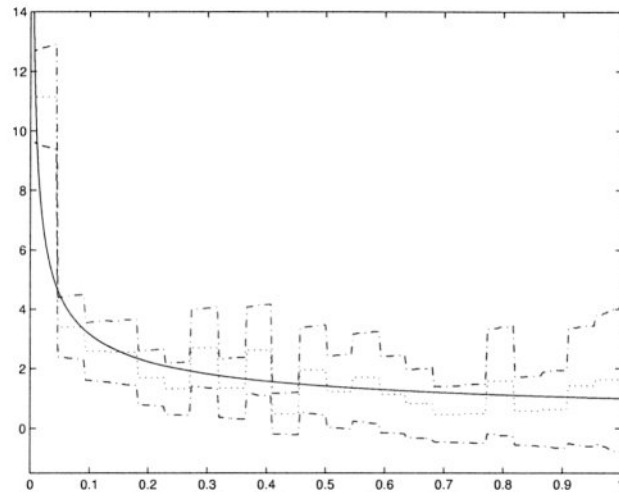
jest liczbą obiektów w stanie ryzyka bezpośrednio przed chwilą t .

Rysunki 2.1–2.4 przedstawiają estymator histogramowy funkcji hazardu skonstruowany dla wszystkich rozważanych rozkładów na przedziale $[0, 1]$ oraz dla $m(n) = \sqrt{n}$. Wyznaczono również 95% punktowe (*pointwise*) przedziały ufności w oparciu o asymptotyczny rozkład normalny oraz estymator wariancji.

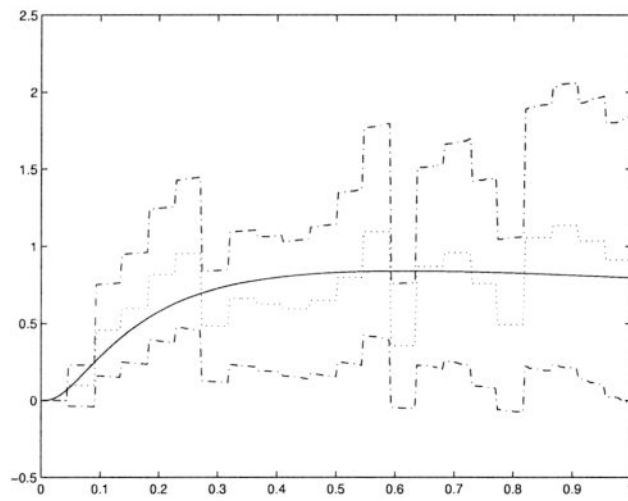
Dla każdego rysunku: teoretyczną funkcję hazardu oznaczono linią ciągłą (—), estymator histogramowy linią kropkowaną ($\cdot \cdot \cdot$), przedziały ufności oznaczono jako linia kreska-kropka ($-\cdot -$).

Dokładność skonstruowanych przedziałów ufności była badana poprzez wyznaczenie empirycznych prawdopodobieństw pokrycia na bazie 1000 powtórzeń Monte Carlo. Tabela 2.1 zawiera wyniki pokrycia empirycznego dla wybranych punktów oraz różnych wyborów ciągu $m(n)$. Oprócz ciągu $m(n) = \sqrt{n}$ wykorzystaliśmy w symulacjach również $m(n) = n^{4/9}$ oraz $m(n) = n^{2/5}$, które spełniają warunki określone w Uwadze 2.4.1.

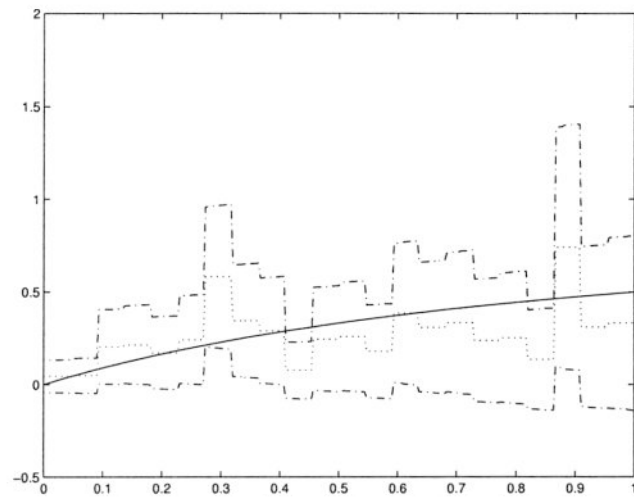
Możemy zauważyć, że wybierając ciąg $m(n) = \sqrt{n}$ nie w każdym przypadku otrzymujemy zadowalające wyniki pokrycia empirycznego, co jest szczególnie widoczne dla punktów brzegowych odcinka $[0, 1]$ i np. rozkładów Weibulla, logarytmiczno-normalnego i gamma.



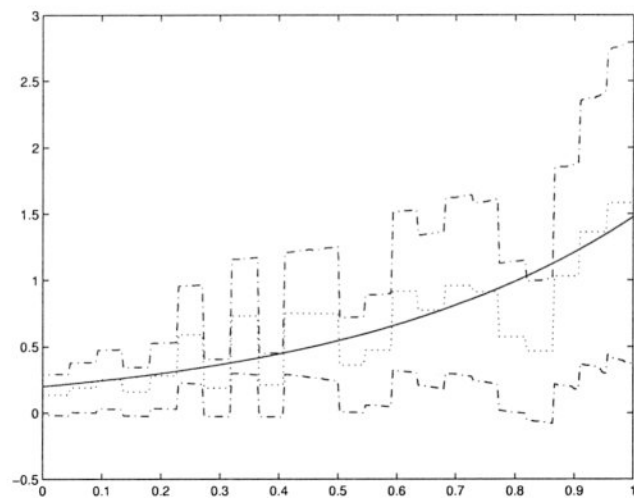
Rys. 2.1: Estymator histogramowy dla rozkładu Weibulla.



Rys. 2.2: Estymator histogramowy dla rozkładu log-normalnego.



Rys. 2.3: Estymator histogramowy dla rozkładu gamma.



Rys. 2.4: Estymator histogramowy dla rozkładu Gompertza.

Rozkład	m(n)	0.01	0.12	0.23	0.33	0.44	0.55	0.65	0.76	0.87	0.97
Weibull	22	49%	95%	96%	94%	93%	90%	88%	90%	79%	71%
	16	91%	89%	95%	93%	91%	92%	91%	86%	90%	82%
	12	52%	95%	96%	91%	93%	92%	93%	86%	91%	88%
LogN	22	39%	89%	91%	95%	94%	93%	93%	90%	91%	87%
	16	71%	72%	92%	95%	93%	93%	92%	91%	93%	92%
	12	59%	93%	92%	95%	93%	94%	93%	92%	92%	92%
Gamma	22	39%	87%	82%	90%	92%	92%	91%	89%	86%	84%
	16	59%	71%	94%	90%	90%	91%	92%	90%	90%	92%
	12	75%	91%	92%	92%	93%	92%	93%	91%	91%	91%
Gompertz	22	95%	86%	90%	92%	92%	93%	91%	92%	93%	92%
	16	95%	90%	93%	92%	94%	92%	94%	92%	93%	93%
	12	93%	95%	92%	94%	94%	93%	92%	94%	94%	93%

Tabela 2.1: Wyniki pokrycia empirycznego. Nominalne prawdopodobieństwo pokrycia=95%.

Rozdział 3

Zastosowanie metody sita do estymacji funkcji dryfu dla procesów Itô

3.1 Przegląd wyników

Jak wspominaliśmy już w rozdziale 2 metoda sita była z powodzeniem stosowana do estymacji parametrów funkcyjnych procesów stochastycznych, takich jak moltiplikatywne procesy punktowe czy procesy dyfuzyjne. Przedstawimy teraz przegląd wyników związanych z estymacją funkcji dryfu dla procesów dyfuzyjnych i modeli semimartynałowych.

3.1.1 Estymacja dla procesów dyfuzyjnych

Geman i Hwang (1982), wykorzystując sito oparte na układzie ortonormalnym otrzymali zgodny (w normie L^2) estymator nieznanej funkcji $\alpha(\cdot)$ dla modelu:

$$dX(t) = \alpha(t) dt + dW(t) \quad t \in [-1/2, 1/2],$$

gdzie $W(t)$ jest procesem Wienera.

Nguyen i Pham (1982) uogólnili powyższy rezultat na przypadek modelu

$$dX(t) = \alpha(t)X(t) dt + dW(t), \quad X(0) = x_0.$$

Leśkow i Różański (1989) wykorzystując sito histogramowe skonstruowali zgodny i asymptotycznie normalny estymator nieznanej funkcji $\alpha(\cdot)$ w bardziej ogólnym modelu dyfuzyjnym

$$dX(t) = \alpha(t)a(t, X(t))dt + dW(t).$$

3.1.2 Modele semimartynałowe

McKeague (1986) otrzymał zgodny (w normie L^2) estymator stosując sito oparte na układzie ortonormalnym dla modelu semimartynałowego:

$$dX(t) = Y(t)\alpha(t) dt + dM(t), \quad t \in [0, 1], \quad (3.1)$$

gdzie: $X = (X_1, \dots, X_n)$, $\alpha = (\alpha_1, \dots, \alpha_p)'$ jest wektorem nieznanych funkcji deterministycznych, $Y = ((Y_{ij}))_{n \times p}$ jest macierzą procesów czynnikowych, $M = (M_1, \dots, M_n)'$, gdzie M_i jest kwadratowo całkwalnym martynałem.

Warto zauważyć (np. Prakasa Rao (1999)), że szczególnymi przypadkami modelu (3.1) są:

(a) Proces dyfuzyjny

$$dX(t) = \alpha(t)b(X_t) dt + \sigma(t, X_t) dW_t, \quad t \in [0, 1], \quad X_0 = \eta.$$

W tym przypadku $Y(t) = b(X_t)$ i $M_t = \int_0^t \sigma(s, X_s) dW_s$.

(b) Proces punktowy – model Aalena

$N = (N(t), \mathcal{F}_t)$ – proces punktowy o intensywności

$$\lambda(t) = \sum_{j=1}^p \alpha_j(t) Y_j(t),$$

gdzie α_j jest nieznaną, nieujemną funkcją ciągłą oraz $Y_j(t)$ jest nieujemnym procesem przewidywalnym.

(c) Proces dyfuzyjny z komponentą skokową

$$dX(t) = \beta(t)b(X_t) dt + \sigma(t, X_t) dW_t + \varepsilon dN_t,$$

gdzie $X_0 = \eta$, $N_0 = 0$, $t \geq 0$.

Estymatory zaproponowane w wymienionych w 3.1.1 i 3.1.2 publikacjach były konstruowane na bazie n niezależnych (lub spełniających pewne warunki mieszania) kopii procesów, co może w pewnych wypadkach stanowić ograniczenie ich stosowności.

3.2 Estymacja funkcji dryfu

Niech $(\Omega, \mathcal{F}, P)$ oznacza przestrzeń probabilistyczną z ciągiem filtracji $\{\mathcal{F}_{n,t}\}$. Rozważmy proces Itô $X_n(t)$, $t \in [0, 1]$ zdefiniowany jako:

$$dX_n(t) = \alpha(t)D_n(t)dt + V_n(X_n, t)dW_n(t), \quad X_n(0) = 0. \quad (3.2)$$

Będziemy przyjmowali dalej następujące założenia:

- (A.1) $W_n(t)$ jest standardowym procesem Wienera, adaptowanym do filtracji $\{\mathcal{F}_{n,t}\}$.
- (A.2) $D_n(t)$ i $X_n(t)$ są uzgodnione z filtracją $\{\mathcal{F}_{n,t}\}$ oraz $V_n(X_n(\cdot), t)$ jest adaptowany do filtracji generowanej przez X_n na $[0, T]$, ponadto prawie wszystkie realizacje $D_n(\cdot)$ i $V_n(\cdot)$ są funkcjami ciągłymi zmiennej t .
- (A.3) $|V_n(t)| > 0 \quad P \text{ p.w.}$

Z Tw.7.18 (Liptser i Shirayev (1981)), otrzymujemy że miara μ_α^n generowana przez proces X_n jest absolutnie ciągła względem miary μ_{Y_n} odpowiadającej procesowi Y_n zdefiniowanemu jako $dY_n(t) = V_n(Y_n, t)dW_n(t)$. Gęstość miary μ_α^n w odniesieniu do μ_{Y_n} ma postać:

$$\frac{d\mu_\alpha^n}{d\mu_{Y_n}} = \exp \left(\int_0^1 \alpha(t) \frac{D_n(t)}{V_n^2(t)} dX_n(t) - \frac{1}{2} \int_0^1 \alpha^2(t) \frac{D_n^2(t)}{V_n^2(t)} dt \right). \quad (3.3)$$

Będziemy oznaczali dalej logarytm powyższej gęstości jako $L_n(\alpha)$.

Standardowe obliczenia (przeprowadzone analogicznie jak w pracy Leśkowska i Różańskiego (1989b)) prowadzą do wyznaczenia estymatora największej wiarygodności.

Lemat 3.2.1 *Estymator największej wiarygodności $\hat{\alpha}_n$ wyznaczony dla sita histogramowego $S(n)$ ma następującą postać:*

$$\hat{\alpha}_n(s) = \sum_{l=1}^{m(n)} \frac{\int_{B_{l,m(n)}} \frac{D_n(t)}{V_n^2(t)} dX_n(t)}{\int_{B_{l,m(n)}} \frac{D_n^2(t)}{V_n^2(t)} dt} 1_{B_{l,m(n)}}(s) 1_{C_{l,m(n)}}(s), \quad (3.4)$$

gdzie $C_{l,m(n)} = \{ \int_{B_{l,m(n)}} \frac{D_n^2(t)}{V_n^2(t)} dt > 0 \}$, $l = 1, \dots, m(n)$.

Niech $s \in [0, 1]$ będzie ustalone, wybierzmy $l(n, s) \in \{1, \dots, m(n)\}$, takie że $s \in B_{l(n,s), m(n)}$. Oznaczając $B_{m(n)}(s) = B_{l(n,s), m(n)}$ możemy zapisać estymator $\hat{\alpha}_n(s)$ w prostszej postaci:

$$\hat{\alpha}_n(s) = \frac{\int_{B_{m(n)}(s)} \frac{D_n(t)}{V_n^2(t)} dX_n(t)}{\int_{B_{m(n)}(s)} \frac{D_n^2(t)}{V_n^2(t)} dt} 1_{C_{m(n)}(s)}. \quad (3.5)$$

3.3 Zgodność i asymptotyczna normalność histogramowego estymatora sitowego.

Sformułujemy teraz założenia, dla których możemy otrzymać rezultaty związane ze zgodnością oraz rozkładem asymptotycznym estymatora (3.4).

(B.1) Istnieje $\delta > 0$ oraz funkcja $y : [0, 1] \rightarrow R_+$, dodatnia i ciągła w s , taka że:

$$\sup_{t \in [s-\delta, s+\delta]} \left| \frac{D_n^2(t)}{V_n^2(t)} \frac{1}{n} - y(t) \right| \xrightarrow{P} 0.$$

(B.2) Funkcja $\alpha(s)$ jest ciągła na $[0, 1]$.

(B.3) Szybkość wzrostu sita $m_n = n^{1/2}$.

(B.4) α spełnia lokalny warunek Höldera, tzn. istnieje $\delta > 0$, $\beta > \frac{1}{2}$ oraz stała dodatnia $C(\alpha, s)$, taka że:

$$\forall_{t \in [s-\delta, s+\delta]} |\alpha(s) - \alpha(t)| \leq C(\alpha, s) |t - s|^\beta.$$

Twierdzenie 3.3.1 (Zgodność estymatora histogramowego)

Jeżeli spełnione są założenia (B.1)–(B.3) wówczas estymator największej wiarygodności $\hat{\alpha}_n(s)$ określony jak we wzorze (3.4) lub (3.5) jest estymatorem punktowo zgodnym dla każdego $s \in [0, 1]$, tzn.

$$\hat{\alpha}_n(s) \xrightarrow{P} \alpha(s)$$

dla dowolnego $s \in [0, 1]$ gdy $n \rightarrow \infty$.

DOWÓD TWIERDZENIA 3.3.1

Aby udowodnić zgodność estymatora $\hat{\alpha}_n(s)$ przedstawmy różnicę $\hat{\alpha}_n(s) - \alpha(s)$, $s \in [0, 1]$ w następującej postaci:

$$\begin{aligned} \hat{\alpha}_n(s) - \alpha(s) &= \\ &= \frac{\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s)} \frac{D_n^2(t)}{V_n^2(t)} (\alpha(t) - \alpha(s)) dt + \frac{1}{\sqrt{n}} \int_{B_{m(n)}(s)} \frac{D_n(t)}{V_n(t)} dW_n(t)}{\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s)} \frac{D_n^2(t)}{V_n^2(t)} dt}. \end{aligned} \quad (3.6)$$

Rozważmy teraz następujący ciąg martyngałów

$$M_n(t) = \frac{1}{\sqrt{n}} \int_0^t \frac{D_n(u)}{V_n(u)} dW_n(u),$$

który jest także ciągiem elementów losowych przestrzeni $\mathcal{D}([0, 1])$. Z Twierdzenia Rebolledo (Rebolledo (1980)) ciąg $M_n(\cdot)$ jest zbieżny według rozkładu w $D([0, 1])$ do elementu $\tilde{M}(\cdot)$, który jest ciągłym, gaussowskim martyngałem o niezależnych przyrostach i dla którego $E\tilde{M}^2(t) = \int_0^t y(u) du$.

Wykorzystując Tw.5.5 (Billingsley (1968)) o słabej zbieżności ciągu odwzorowań ciągłych elementów losowych, otrzymujemy, że:

$$M_n(B_{m(n)}(s)) = M_n(t_{m(n)+1}(s)) - M_n(t_{m(n)}(s)) \xrightarrow{D} 0,$$

gdzie $B_{m(n)}(s) = (t_{m(n)}(s), t_{m(n)+1}(s)]$, $t_{m(n)}(s) \nearrow s$, $t_{m(n)+1}(s) \searrow s$.
Na mocy założeń (B.1), (B.2) oraz (B.3) mamy:

$$\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s)} \frac{D_n^2(t)}{V_n^2(t)} dt \xrightarrow{P} y(s), \quad (3.7)$$

oraz

$$\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s)} \frac{D_n^2(t)}{V_n^2(t)} (\alpha(t) - \alpha(s)) dt \xrightarrow{P} 0, \quad (3.8)$$

co implikuje, że (3.6) zbiega do zera według prawdopodobieństwa.

□

Twierdzenie 3.3.2 (Asymptotyczna normalność)

Niech $\{s_1, \dots, s_p\}$ będzie dowolnym skończonym zbiorem punktów z przedziału $[0, 1]$.
Jeżeli spełnione są warunki (B.1) – (B.4) wówczas ciąg wektorów losowych

$$n^{1/4}(\hat{\alpha}_n(s_1) - \alpha(s_1), \dots, \hat{\alpha}_n(s_p) - \alpha(s_p))$$

zbiega według rozkładu, gdy $n \rightarrow \infty$, do p -wymiarowego rozkładu normalnego o średniej zero i diagonalnej macierzy kowariancji Σ z elementami na przekątnej: $\sigma_i = \frac{1}{y(s_i)}$, $i = 1, \dots, p$.

DOWÓD TWIERDZENIA 3.3.2

Aby udowodnić Twierdzenie 3.3.2 pokażemy najpierw zbieżność rozkładów jednowymiarowych.

Możemy zapisać:

$$\begin{aligned} n^{1/4}(\hat{\alpha}_n(s) - \alpha(s)) &= \\ &= \frac{n^{1/4}(\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s)} \frac{D_n^2(t)}{V_n^2(t)} (\alpha(t) - \alpha(s)) dt) + n^{1/4}(\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s)} \frac{D_n(t)}{V_n(t)} dW_n(t))}{\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s)} \frac{D_n^2(t)}{V_n^2(t)} dt}, \end{aligned} \quad (3.9)$$

W dowodzie Tw.3.3.1 pokazaliśmy (wzór 3.7), że mianownik we wzorze (3.9) zbiega według prawdopodobieństwa do $y(s)$.

Zauważmy, że na mocy założenia (B.4) pierwszy składnik w liczniku (3.9) zbiega do zera według prawdopodobieństwa.

Na mocy wspomnianego już Twierdzenia Rebolledo (Rebolledo (1980)) ciąg $M_n(\cdot)$ jest zbieżny według rozkładu w $D([0, 1])$ do elementu $\tilde{M}(\cdot)$, który jest ciągłym, gausowskim martyngałem o niezależnych przyrostach.

Oznaczmy przez Q_n i Q miary generowane odpowiednio przez $M_n(\cdot)$ i $\tilde{M}(\cdot)$ w $D([0, 1])$.

Niech $\mathcal{A}$ oznacza następującą klasę podzbiorów $D([0, 1])$:

$$\mathcal{A} = \{A \subset D([0, 1]) : A = \bigcap_k \{x : x(t_k(s)) \leq x_k\} \cap \bigcap_l \{x : x(t_l(s)) \leq y_l\}\},$$

gdzie $\{t_k(s)\}$ i $\{t_l(s)\}$ są skończonymi lub nieskończonymi podzbiarami zbiorów $\{t_{m(n)}(s)\}$ i $\{t_{m(n)+1}(s)\}$, a $\{x_k\}$, $\{y_l\}$ są dowolnymi liczbami rzeczywistymi. Z Twierdzenia 3 (Topsøe (1967)) wynika, że $\mathcal{A}$ jest klasą Q -jednostajną. Otrzymujemy więc:

$$\sup_{A \in \mathcal{A}} |Q_n(A) - Q(A)| \longrightarrow 0.$$

Łatwo zauważyć, że σ -algebra $\sigma(\mathcal{A})$ generowana przez klasę $\mathcal{A}$ jest również klasą Q -jednostajną.

Niech F_n i G_n oznaczają dystrybuanty rozkładów odpowiednio $M_n(B_{m(n)}(s))$ i $\tilde{M}(B_{m(n)}(s))$. Ponieważ $\sigma(M_n(t_{m(n)}(s)); M_n(t_{m(n)+1}(s)), n \geq 1) \subset \sigma(\mathcal{A})$, gdzie $\sigma(\mathcal{A})$ klasa Q -jednostajna, otrzymujemy że:

$$\sup_x |F_n(x) - G_n(x)| \longrightarrow 0. \quad (3.10)$$

Oczywiście, $n^{\frac{1}{4}}(\tilde{M}(B_{m(n)}(s))) \xrightarrow{D} U$, gdzie U jest zmienną losową o rozkładzie normalnym o wartości oczekiwanej zero i wariancji równej $y(s)$. Z (3.10) wynika, że również $n^{\frac{1}{4}}(M_n(B_{m(n)}(s))) \xrightarrow{D} U$.

Udowodniliśmy więc, że licznik (3.9) zbiega według rozkładu do zmiennej losowej U oraz mianownik w (3.9) jest zbieżny według prawdopodobieństwa do $y(s)$. W ten sposób uzasadniliśmy, że (3.9) jest asymptotycznie normalny o średniej zero i wariancji $\frac{1}{y(s)}$.

Pokażemy teraz zbieżność rozkładów dwuwymiarowych. Możemy zapisać

$$\begin{aligned} & \left[n^{\frac{1}{4}}(\hat{\alpha}_n(s_1) - \alpha(s_1)), n^{\frac{1}{4}}(\hat{\alpha}_n(s_2) - \alpha(s_2)) \right] = \\ & \left[\frac{n^{\frac{1}{4}} \left(\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s_1)} \frac{D_n^2(t)}{V_n^2(t)} (\alpha(t) - \alpha(s_1)) dt \right)}{\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s_1)} \frac{D_n^2(t)}{V_n^2(t)} dt}, \frac{n^{\frac{1}{4}} \left(\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s_2)} \frac{D_n^2(t)}{V_n^2(t)} (\alpha(t) - \alpha(s_2)) dt \right)}{\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s_2)} \frac{D_n^2(t)}{V_n^2(t)} dt} \right] + \\ & + \left[\frac{n^{\frac{1}{4}} \left(\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s_1)} \frac{D_n(t)}{V_n(t)} dW_n(t) \right)}{\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s_1)} \frac{D_n^2(t)}{V_n^2(t)} dt}, \frac{n^{\frac{1}{4}} \left(\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s_2)} \frac{D_n(t)}{V_n(t)} dW_n(t) \right)}{\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s_2)} \frac{D_n^2(t)}{V_n^2(t)} dt} \right]. \end{aligned} \quad (3.11)$$

Na mocy wcześniejszych rozważań zauważmy, że pierwszy składnik w (3.11) jest zbieżny do 0 według prawdopodobieństwa.

Aby pokazać asymptotyczną normalność drugiego składnika w (3.11) wprowadzimy następujące oznaczenia:

$$\begin{aligned} Y_{n,i} &= n^{\frac{1}{4}} \left(\frac{1}{\sqrt{n}} \int_{B_{m(n)}(s_i)} \frac{D_n(t)}{V_n(t)} dW_n(t) \right), \\ Z_{n,i} &= \frac{1}{\sqrt{n}} \int_{B_{m(n)}(s_i)} \frac{D_n^2(t)}{V_n^2(t)} dt, \end{aligned} \quad i = 1, 2.$$

Wykorzystując podobną argumentację jak w przypadku jednowymiarowym można pokazać, że dla dowolnych $\underline{t} = (t_1, t_2)$:

$$\langle \underline{t}, \underline{Y}_n \rangle = t_1 Y_{n,1} + t_2 Y_{n,2} \xrightarrow{D} t_1 Y_{0,1} + t_2 Y_{0,2},$$

gdzie $Y_{0,1}$ i $Y_{0,2}$ są niezależnymi zmiennymi losowymi, o rozkładzie normalnym o średniej zero i wariancji odpowiednio $y(s_1)$ oraz $y(s_2)$. Na mocy Twierdzenia Craméra-Wolda otrzymujemy asymptotyczną normalność wektora $[Y_{n,1}, Y_{n,2}]$.

Oznaczmy $R_{n,i} = t_i/Z_{n,i}$, $i = 1, 2$. Możemy zapisać

$$\left\langle \underline{t}, \begin{bmatrix} Y_{n,1} \\ Z_{n,1} \end{bmatrix}, \begin{bmatrix} Y_{n,2} \\ Z_{n,2} \end{bmatrix} \right\rangle = R_{n,1} Y_{n,1} + R_{n,2} Y_{n,2} = \langle \underline{R}_n, \underline{Y}_n \rangle.$$

Stosując Lemat Śluckiego otrzymujemy:

$$\langle \underline{R}_n, \underline{Y}_n \rangle \xrightarrow{D} \frac{t_1}{y(s_1)} Y_{0,1} + \frac{t_2}{y(s_2)} Y_{0,2}.$$

Wykorzystując ponownie Twierdzenie Craméra-Wolda dostajemy zbieżność dla przypadku dwuwymiarowego. Dowód dla przypadku p -wymiarowego przebiega analogicznie. □

Uwaga 3.3.1

Założenie (B.3) określające szybkość wzrostu $sita$, może być zmienione. Ciąg $m(n) = \sqrt{n}$ możemy zastąpić dowolnym ciągiem $m(n)$ zbieżnym do nieskończoności wolniej niż $\sqrt{n}$. Warunek ten gwarantuje zgodność estymatora $\hat{\alpha}(t)$. Aby otrzymać wyniki związane z rozkładem asymptotycznym musimy przyjąć dodatkowe założenia dotyczące gładkości estymowanej funkcji $\alpha(t)$ oraz przyjąć jako ciąg normalizacyjny $(\frac{n}{m(n)})^{1/2}$. Dla przykładu, zakładając że funkcja α jest różniczkowalna w s oraz $m(n)$ jest takie że $\frac{n^{1/2}}{m(n)} \rightarrow \infty$, $\frac{n}{m(n)^3} \rightarrow 0$ otrzymujemy że $\hat{\alpha}_n(s)$ jest asymptotycznie normalny $AN(\alpha(s), \frac{m(n)}{n} \frac{1}{y(s)})$.

Przyjmując dodatkowe założenia (B.5)–(B.7), możemy otrzymać wyniki związane ze zgodnością w L^1 estymatora histogramowego $\hat{\alpha}(t)$.

(B.5) Funkcja $\left| \frac{D_n^2(t)}{V_n^2(t)} \frac{1}{n} \right|$ jest oddzielona od zera, tzn. Istnieje $\sigma > 0$, taka że:

$$\inf_{t \in [0,1]} \left| \frac{D_n^2(t)}{V_n^2(t)} \frac{1}{n} \right| > \sigma \quad P - \text{prawie wszędzie.}$$

(B.6) Funkcja $y(t)$ określona w zał. (B.1) spełnia:

$$E \left| \frac{D_n^2(t)}{V_n^2(t)} \frac{1}{n} - y(t) \right| \rightarrow 0 \quad \forall t \in [0,1],$$

ponadto $\forall_n \ E \frac{D_n^2(t)}{V_n^2(t)}$ jest ciągłą funkcją argumentu t .

(B.7) Dla $\alpha(t)$ zachodzi globalny warunek Hölder’a, tzn. istnieją stałe dodatnie $C(\alpha)$ i β , takie że: $|\alpha(t) - \alpha(s)| \leq C(\alpha)|t - s|^\beta$.

Twierdzenie 3.3.3 (Zgodność w L^1)

Założmy, że warunki (B.1) – (B.3) pozostają prawdziwe. Przyjmijmy dodatkowo, że spełnione są założenia (B.5) – (B.7). Wówczas estymator histogramowy $\hat{\alpha}_n(s)$ jest estymatorem zgodnym w L^1 , tzn.

$$\int_0^1 |\hat{\alpha}_n(s) - \alpha(s)| ds \xrightarrow{P} 0.$$

DOWÓD TWIERDZENIA 3.3.3

Założenie (B.5) pozwala nam zapisać:

$$E |\hat{\alpha}_n(s) - \alpha(s)| \leq \sigma^{-1} n^{-1/2} E \left| \int_{B_{m(n)}(s)} \frac{D_n^2(t)}{V_n^2(t)} (\alpha(t) - \alpha(s)) dt + \int_{B_{m(n)}(s)} \frac{D_n(t)}{V_n(t)} dW_n(t) \right|.$$

Z założenia (B.6) wynika, że istnieje stała dodatnia C_1 taka, że dla dowolnego $t \in [0, 1]$ oraz $n \in N$:

$$E \frac{D_n^2(t)}{V_n^2(t)} \leq C_1 n.$$

Stosując teraz warunek Höldera (założenie B.7) oraz nierówność Cauchy’ego otrzymujemy, że: $E|\hat{\alpha}_n(t) - \alpha(t)| \leq \sigma^{-1}(C(\alpha)C_1 n^{-\beta/2} + \sqrt{C_1} n^{-1/4})$.

Pokazaliśmy więc:

$$\lim_{n \rightarrow \infty} E |\hat{\alpha}_n(t) - \alpha(t)| = 0.$$

Stosując twierdzenie o zbieżności zmajoryzowanej oraz twierdzenie Fubiniego otrzymujemy ostatecznie, że:

$$\lim_{n \rightarrow \infty} E \int_0^1 |\hat{\alpha}_n(s) - \alpha(s)| ds = 0,$$

a to na mocy nierówności Czebyszewa daje już tezę Twierdzenia 3.3.3.

□

3.4 Przykłady

(I) Niech $Z_n(t)$ oznacza ciąg procesów dyfuzyjnych określonych jako:

$$dZ_n(t) = \alpha(t) dt + \frac{\sigma}{\sqrt{n}} dW_n(t), \quad Z_n(0) = 0. \quad t \in [0, 1]$$

W tym przypadku, estymacja nieznanej funkcji dryfu $\alpha(t)$ jest asymptotycznie równoważna estymacji funkcji regresji w modelu :

$$Y_i = \alpha(t_i) + \sigma \varepsilon_i, \quad i = 1, 2, \dots, n,$$

gdzie $t_i = \frac{i}{n}$ natomiast ε_i jest ciągiem i.i.d. z $E\varepsilon_i = 0$ oraz $Var(\varepsilon_i) = 1$.

(II) Niech $Y_k(t)$, $k = 1, 2, \dots, n$ będzie ciągiem niezależnych kopii procesu dyfuzyjnego, określonym jako:

$$dY_k(t) = \alpha(t)a(t, Y_k)dt + dW_k(t),$$

gdzie $Y_k(0) = \tilde{Y}$ jest zmienną losową, natomiast $W_k(t)$ jest procesem Wienera. Określmy ciąg procesów $X_n(t)$:

$$\begin{aligned} dX_n(t) &= \sum_{k=1}^n a(t, Y_k) dY_k(t) \\ &= \alpha(t) \left(\sum_{k=1}^n a^2(t, Y_k) \right) dt + \sum_{k=1}^n a(t, Y_k) dW_k(t). \end{aligned}$$

Otrzymaliśmy więc:

$$dX_n(t) = \alpha(t) \left(\sum_{k=1}^n a^2(t, Y_k) \right) dt + d\tilde{M}_n(t),$$

gdzie

$$d < \tilde{M}_n(t) > = \left(\sum_{k=1}^n a^2(t, Y_k) \right) dt.$$

Określając dalej:

$$dM_n(t) = \frac{1}{\sqrt{\sum_{k=1}^n a^2(t, Y_k)}} d\tilde{M}_n(t),$$

otrzymujemy:

$$d < M_n(t) > = \frac{1}{\sum_{k=1}^n a^2(t, Y_k)} d < \tilde{M}_n(t) > = dt$$

oraz

$$dX_n(t) = \alpha(t) \left(\sum_{k=1}^n a^2(t, Y_k) \right) dt + \sqrt{\sum_{k=1}^n a^2(t, Y_k)} dM_n(t).$$

Wykorzystując teraz Tw.5.12 oraz Tw.7.17 (Liptser i Shiryaev (1981)) proces $X_n(t)$ możemy ostatecznie przedstawić w postaci modelu Itô rozważanego w rozdziale 3.2:

$$dX_n(t) = \alpha(t)D_n(t) dt + V_n(X_n, t) dW_n(t).$$

3.5 Symulacje komputerowe

W badaniach symulacyjnych rozważono przykłady (I) i (II) przedstawione w 3.4. Dane były generowane dla następujących modeli:

$$(M1) \quad dZ_n(t) = \alpha(t) dt + \frac{\sigma}{\sqrt{n}} dW_n(t), \quad Z_n(0) = 0, \quad \sigma = 1,$$

$$(M2) \quad dX_n(t) = \sum_{k=1}^n Y_k(t) dY_k(t),$$

gdzie Y_k są niezależnymi kopiami procesu dyfuzyjnego $Y(t)$, dla którego:

$$dY_k(t) = \alpha(t) Y_k(t) dt + dW_k(t), \quad Y_k(0) \sim N(0, 1).$$

Dla obu modeli symulacje były przeprowadzone dla dwóch różnych funkcji dryfu:

- $\alpha(t) = \sin(4\pi t)$,
- $\alpha(t) = t^2$.

Rysunki 3.1–3.4 przedstawiają estymator histogramowy funkcji dryfu (linia kropkowana) dla modelu M1 i M2 uzyskany dla $n = 500$ oraz $m(n) = \sqrt{n}$.

Dodatkowo skonstruowano 95% przedziały ufności, w oparciu o asymptotyczny rozkład normalny i estymator wariancji.

Efektywność skonstruowanych przedziałów była badana z wykorzystaniem empirycznych prawdopodobieństw pokrycia wyznaczonych na bazie 1000 prób Monte Carlo.

W symulacjach uwzględniono trzy różne wybory ciągu $m(n)$:

- $m(n) = \sqrt{n}$,
- $m(n) = n^{4/9}$,
- $m(n) = n^{2/5}$,

spełniające założenia określone w Uwadze 3.3.1.

Tabele 3.1–3.6 zawierają procentowe wyniki pokrycia dla wybranych punktów przedziału $[0,1]$ oraz estymowane błędy standardowe (w nawiasach).

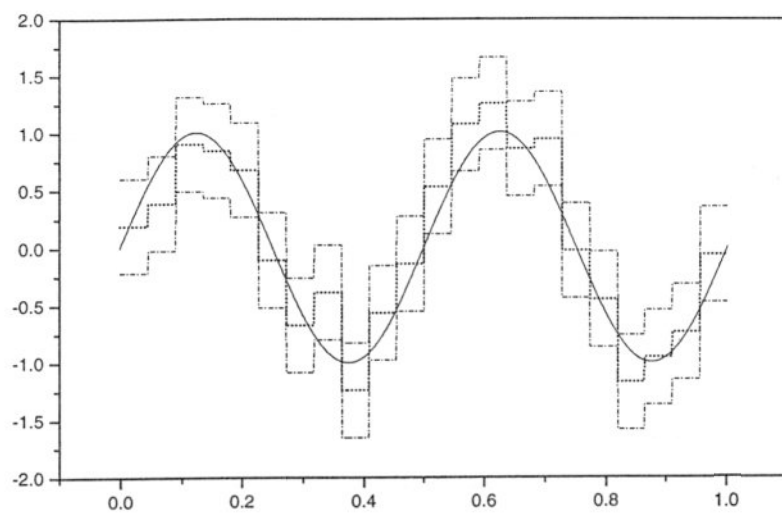
Możemy zauważyć, że efektywność przedziałów ufności dla modeli M1 oraz M2 i sinusoidalnej funkcji dryfu ($\alpha(t) = \sin(4\pi t)$) nie jest zadowalająca jeśli wybierzemy $m(n) = n^{2/5}$, a więc zmierzające do nieskończoności zbyt wolno. Dla pozostałych wyborów $m(n)$ otrzymujemy podobne pokrycie, bliskie nominalnemu (95%).

Dla pozostałych modeli (o parabolicznej funkcji dryfu) wyniki pokrycia empirycznego otrzymane dla wszystkich analizowanych ciągów $m(n)$ są podobne i wybierając $m(n)$ inne niż $\sqrt{n}$ nie uzyskujemy istotnej poprawy.

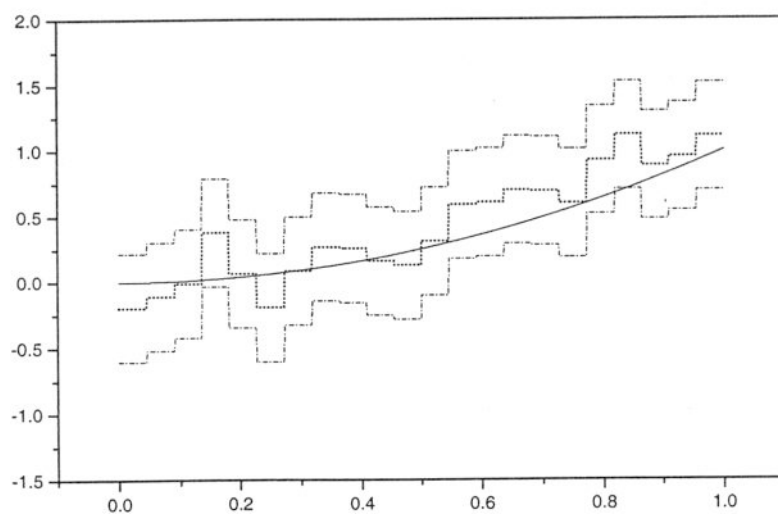
Niemniej jednak, prawdopodobieństwa pokrycia otrzymane dla $m(n) = \sqrt{n}$ charakteryzują się większą zmiennością dla różnych punktów odcinka $[0,1]$. Warto zauważyć, że dla wszystkich analizowanych przypadków mniejsza wartość $m(n)$ prowadzi do węższych przedziałów ufności.

Dokładność estymatora histogramowego może być analizowana również w odniesieniu do kryteriów dokładności dopasowania, takich jak $MIAE = E \int_0^1 |\hat{\alpha}(u) - \alpha(u)| du$, czy $MISE = E \int_0^1 (\hat{\alpha}(u) - \alpha(u))^2 du$. Przypomnijmy, że w rozdziale 3.3 przedstawione były wyniki teoretyczne związane ze zgodnością estymatora histogramowego w L^1 (Tw.3.3.3).

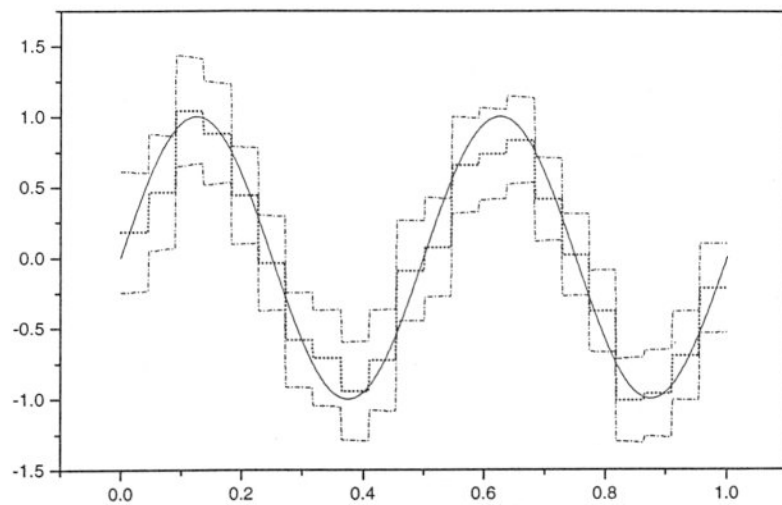
Tabele 3.7-3.10 zawierają wyniki błędów MIAE i MISE wyznaczonych na bazie 1000 realizacji Monte Carlo dla modeli M1 i M2 oraz dla trzech różnych wyborów ciągów $m(n)$. Możemy zaobserwować, że oba kryteria zachowują się w przybliżeniu podobnie w zależności od wyboru $m(n)$. Dla obu modeli i funkcji dryfu $\alpha(t) = t^2$ obserwujemy, że zarówno MIAE jak i MISE maleją wraz z $m(n)$. Z drugiej strony, dla sinusoidalnej funkcji dryfu nie obserwujemy tak regularnej tendencji, tzn. dla M1 i $\alpha(t) = \sin(4\pi t)$ najlepsza wartość jest osiągana dla ciągu $m(n) = n^{4/9}$, podczas gdy dla M2 wartości obu kryteriów są najmniejsze kiedy $m(n) = \sqrt{n}$.



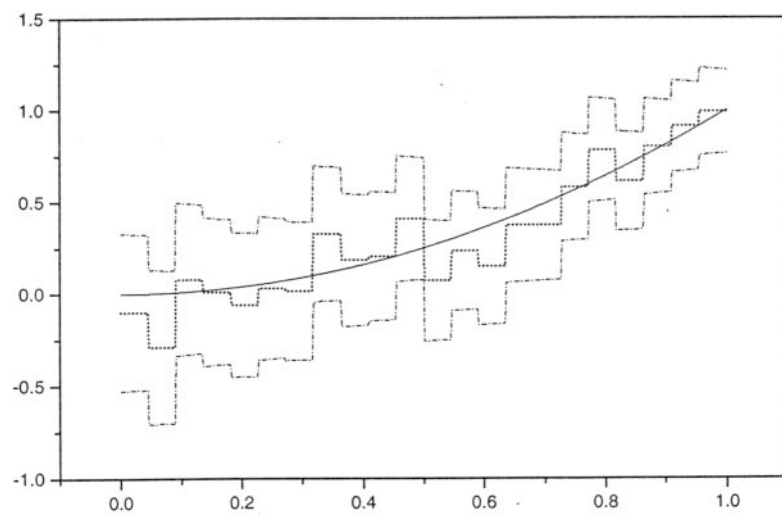
Rys. 3.1: Estymator histogramowy oraz przedziały ufności dla M1 i $\alpha(t) = \sin(4\pi t)$



Rys. 3.2: Estymator histogramowy oraz przedziały ufności dla M1 i $\alpha(t) = t^2$



Rys. 3.3: Estymator histogramowy oraz przedziały ufności dla M2 i $\alpha(t) = \sin(4\pi t)$



Rys. 3.4: Estymator histogramowy oraz przedziały ufności dla M2 i $\alpha(t) = t^2$

t	$\alpha(t) = \sin(4 \pi t)$	$\alpha(t) = t^2$
0.0227	93.9%(0.757)	95.8%(0.634)
0.0680	95.7%(0.641)	93.3%(0.791)
0.1133	94.8%(0.702)	95.9%(0.627)
0.1593	95.3%(0.669)	94.8%(0.702)
0.2047	94.1%(0.745)	94.1%(0.745)
0.2500	95.8%(0.634)	96.1%(0.612)
0.2953	94.6%(0.715)	95.3%(0.669)
0.3407	94.9%(0.696)	94.3%(0.733)
0.3867	94.4%(0.727)	95.7%(0.641)
0.4320	94.4%(0.727)	95.3%(0.669)
0.4773	94.8%(0.702)	94.8%(0.702)
0.5227	95.2%(0.676)	94.4%(0.727)
0.5680	94.0%(0.751)	94.7%(0.708)
0.6133	94.9%(0.696)	94.7%(0.708)
0.6593	95.4%(0.662)	96.0%(0.620)
0.7047	95.1%(0.683)	95.1%(0.683)
0.7500	94.2%(0.739)	94.4%(0.727)
0.7953	95.9%(0.627)	94.1%(0.745)
0.8407	94.4%(0.727)	95.4%(0.662)
0.8867	95.5%(0.656)	95.2%(0.676)
0.9320	95.0%(0.689)	95.9%(0.627)
0.9773	95.1%(0.683)	94.7%(0.708)

Tabela 3.1: Wyniki pokrycia empirycznego dla modelu M1 i $m(n) = \sqrt{n}$

t	$\alpha(t) = \sin(4 \pi t)$	$\alpha(t) = t^2$
0.0313	94.5%(0.721)	95.6%(0.649)
0.0940	94.1%(0.745)	95.7%(0.641)
0.1560	96.2%(0.605)	95.3%(0.669)
0.2187	95.0%(0.689)	94.8%(0.702)
0.2813	95.2%(0.676)	94.9%(0.696)
0.3440	95.1%(0.683)	95.0%(0.689)
0.4060	94.3%(0.733)	94.6%(0.715)
0.4687	95.1%(0.683)	94.4%(0.727)
0.5313	93.6%(0.774)	94.5%(0.721)
0.5940	95.7%(0.641)	94.3%(0.733)
0.6560	95.6%(0.649)	95.8%(0.634)
0.7187	95.9%(0.627)	95.2%(0.676)
0.7813	95.7%(0.641)	93.3%(0.791)
0.8440	95.0%(0.689)	95.7%(0.641)
0.9060	93.8%(0.763)	93.8%(0.763)
0.9687	95.2%(0.676)	94.2%(0.739)

Tabela 3.2: Wyniki pokrycia empirycznego dla modelu M1 i $m(n) = n^{4/9}$

t	$\alpha(t) = \sin(4 \pi t)$	$\alpha(t) = t^2$
0.0413	94.8%(0.702)	94.6%(0.715)
0.1253	94.1%(0.745)	95.5%(0.656)
0.2080	94.9%(0.696)	95.4%(0.662)
0.2920	94.8%(0.702)	94.6%(0.715)
0.3753	93.4%(0.785)	95.0%(0.689)
0.4580	95.1%(0.683)	94.7%(0.708)
0.5413	93.8%(0.763)	94.7%(0.708)
0.6247	93.7%(0.768)	95.4%(0.662)
0.7080	95.3%(0.669)	94.2%(0.739)
0.7913	95.4%(0.662)	93.9%(0.757)
0.8747	92.6%(0.828)	94.4%(0.727)
0.9580	93.7%(0.768)	95.4%(0.662)

Tabela 3.3: Wyniki pokrycia empirycznego dla modelu M1 i $m(n) = n^{2/5}$

t	$\alpha(t) = \sin(4\pi t)$	$\alpha(t) = t^2$
0.0233	94.0%(0.751)	94.4%(0.727)
0.0683	95.4%(0.662)	95.5%(0.656)
0.1133	93.4%(0.785)	94.0%(0.751)
0.1583	95.4%(0.662)	95.6%(0.649)
0.2050	94.9%(0.696)	95.1%(0.683)
0.2500	94.7%(0.708)	94.6%(0.715)
0.2950	93.5%(0.780)	93.2%(0.796)
0.3417	94.8%(0.702)	94.9%(0.696)
0.3867	95.8%(0.634)	95.5%(0.656)
0.4317	95.2%(0.676)	95.2%(0.676)
0.4767	95.5%(0.656)	95.4%(0.662)
0.5233	95.8%(0.634)	96.0%(0.620)
0.5683	93.9%(0.757)	94.4%(0.727)
0.6133	95.2%(0.676)	95.6%(0.649)
0.6583	95.0%(0.689)	95.2%(0.676)
0.7050	95.8%(0.634)	95.8%(0.634)
0.7500	96.9%(0.548)	96.5%(0.581)
0.7950	94.6%(0.715)	94.7%(0.708)
0.8417	94.8%(0.702)	94.4%(0.727)
0.8867	94.7%(0.708)	95.0%(0.689)
0.9317	94.3%(0.733)	94.3%(0.733)
0.9767	94.6%(0.715)	95.0%(0.689)

Tabela 3.4: Wyniki pokrycia empirycznego dla modelu M2 i $m(n) = \sqrt{n}$

t	$\alpha(t) = \sin(4\pi t)$	$\alpha(t) = t^2$
0.0317	95.0%(0.689)	95.1%(0.683)
0.0933	95.7%(0.641)	96.0%(0.620)
0.1567	95.0%(0.689)	94.5%(0.721)
0.2183	94.0%(0.751)	94.2%(0.739)
0.2817	94.5%(0.721)	94.4%(0.727)
0.3433	95.0%(0.689)	95.6%(0.649)
0.4067	95.0%(0.689)	95.0%(0.689)
0.4683	95.8%(0.634)	96.1%(0.612)
0.5317	95.4%(0.662)	95.5%(0.656)
0.5933	94.1%(0.745)	95.3%(0.669)
0.6567	95.4%(0.662)	95.2%(0.676)
0.7183	95.5%(0.656)	95.6%(0.649)
0.7817	94.9%(0.696)	95.3%(0.669)
0.8433	94.9%(0.696)	95.3%(0.669)
0.9067	94.8%(0.702)	95.3%(0.669)
0.9683	94.3%(0.733)	94.6%(0.715)

Tabela 3.5: Wyniki pokrycia empirycznego dla modelu M2 i $m(n) = n^{4/9}$

t	$\alpha(t) = \sin(4 \pi t)$	$\alpha(t) = t^2$
0.0417	95.3%(0.669)	95.8%(0.634)
0.1250	92.8%(0.817)	94.0%(0.751)
0.2083	94.9%(0.696)	95.0%(0.689)
0.2917	93.8%(0.763)	94.6%(0.715)
0.3750	93.1%(0.801)	94.2%(0.739)
0.4583	94.6%(0.715)	95.3%(0.669)
0.5417	94.2%(0.739)	94.8%(0.702)
0.6250	93.2%(0.796)	94.4%(0.727)
0.7083	95.1%(0.683)	95.0%(0.689)
0.7917	94.8%(0.702)	94.5%(0.721)
0.8750	93.0%(0.807)	95.5%(0.656)
0.9583	94.3%(0.733)	95.1%(0.683)

Tabela 3.6: Wyniki pokrycia empirycznego dla modelu M2 i $m(n) = n^{2/5}$

$\alpha(t)$	n	$m(n)$		
		$\sqrt{n}$	$n^{4/9}$	$n^{2/5}$
$\sin(4\pi t)$	500	0.192	0.191	0.208
	1000	0.156	0.151	0.163
t^2	500	0.167	0.144	0.125
	1000	0.142	0.118	0.102

Tabela 3.7: Wyniki MIAE dla modelu M1.

$\alpha(t)$	n	$m(n)$		
		$\sqrt{n}$	$n^{4/9}$	$n^{2/5}$
$\sin(4\pi t)$	500	0.161	0.169	0.193
	1000	0.129	0.131	0.150
t^2	500	0.131	0.112	0.099
	1000	0.111	0.093	0.081

Tabela 3.8: Wyniki MIAE dla modelu M2.

$\alpha(t)$	n	$m(n)$		
		$\sqrt{n}$	$n^{4/9}$	$n^{2/5}$
$\sin(4\pi t)$	500	0.058	0.057	0.068
	1000	0.038	0.036	0.041
t^2	500	0.044	0.033	0.025
	1000	0.032	0.022	0.016

Tabela 3.9: Wyniki MISE dla modelu M1.

$\alpha(t)$	n	$m(n)$		
		$\sqrt{n}$	$n^{4/9}$	$n^{2/5}$
$\sin(4\pi t)$	500	0.041	0.045	0.059
	1000	0.026	0.027	0.035
t^2	500	0.028	0.020	0.016
	1000	0.020	0.014	0.010

Tabela 3.10: Wyniki MISE dla modelu M2.

Dodatek A: Wybrane zagadnienia teorii prognozy dla procesów stacjonarnych

W dalszej części będziemy przyjmowali następujące oznaczenia:

$$\mathcal{M} = \overline{\text{sp}}\{X_t, t \in Z\},$$

$$\mathcal{M}_n = \overline{\text{sp}}\{X_t, -\infty < t \leq n\}$$

$$\mathcal{M}_{-\infty} = \bigcap_{n=-\infty}^{\infty} \mathcal{M}_n \text{- domknięta liniowa podprzestrzeń przestrzeni } \mathcal{M}$$

$$\sigma^2 = E[X_{n+1} - P_{M_n} X_{n+1}]^2 \text{ - jednokroowy średniokwadratowy błąd predykcji.}$$

$$P_{M_n} X_{n+1} \text{- rzut ortogonalny } X_{n+1} \text{ na podprzestrzeń } M_n.$$

Definicja 1 (Proces deterministyczny i regularny)

Proces $\{X_t\}_{t \in Z}$ jest nazywany deterministycznym wtedy i tylko wtedy, gdy:

$$\sigma^2 = E[X_{n+1} - P_{M_n} X_{n+1}]^2 = 0$$

(tzn. X_{n+1} może być całkowicie prognozowane przez element przestrzeni $\mathcal{M}_n$) lub równoważnie $\iff X_t \in \mathcal{M}_{-\infty}$ dla każdego t .

Proces, dla którego $\sigma^2 = E[X_{n+1} - P_{M_n} X_{n+1}]^2 > 0$ jest nazywany regularnym.

Twierdzenie 1 (Dekompozycja Wolda)

Jeżeli X_t jest regularnym i stacjonarnym procesem stochastycznym o wartości oczekiwanej $E X_t = 0$ wówczas możemy przedstawić go jednoznacznie jako:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t, \tag{3.12}$$

gdzie

- $\psi_0 = 1$ oraz $\sum_{j=0}^{\infty} \psi_j^2 < \infty$,
- Z_t - jest białym szumem o $E Z_t = 0$ i $E Z_t^2 = \sigma^2$,
- $Z_t \in \mathcal{M}_t$ dla każdego $t \in Z$,
- $E(Z_t V_s) = 0$ dla wszystkich $s, t \in Z$,

- $V_t \in \mathcal{M}_{-\infty}$ dla każdego $t \in Z$,
- $\{V_t\}$ jest procesem deterministycznym.

Definicja 2 (Proces całkowicie niedeterministyczny)

Proces stacjonarny X_t jest nazywany całkowicie niedeterministycznym wtedy i tylko wtedy, gdy: $\mathcal{M}_{-\infty} = \{0\}$.

W tym przypadku w dekompozycji Wolda nie występuje komponenta deterministyczna i proces może być przedstawiony jako $MA(\infty)$, $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$.

W dalszym ciągu będziemy zakładali, że $f(\lambda)$ jest gęstością spektralną stacjonarnego procesu X_t (o wartościach rzeczywistych i średniej zero). Zakładamy dodatkowo, że f jest odcięta od zera. Wówczas zachodzi:

Twierdzenie 2 (Wzór Kołmogorowa)

Jednokrokový średniokwadratowy błąd predykcji dla stacjonarnego procesu $\{X_t\}$ wyraża się jako:

$$\sigma^2 = 2\pi \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln f(\lambda) d\lambda \right\}.$$

Fakt 1 Proces X_t jest regularny (tzn. $\sigma^2 > 0$) $\iff \int_{-\pi}^{\pi} \ln f(\lambda) d\lambda > -\infty$.

Fakt 2 Jeżeli dla procesu X_t istnieje gęstość spektralna $f(\lambda)$ i spełniony jest warunek $\int_{-\pi}^{\pi} \ln f(\lambda) d\lambda > -\infty$, wówczas możemy przedstawić go jako proces $MA(\infty)$, tzn. $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, $\psi_0 = 1$, $t \in Z$, gdzie $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.

Twierdzenie 3 (Warunki gwarantujące odwracalność, Tw.7.6.9, Anderson (1971))

Jeżeli $f(\lambda)$ jest ograniczona wówczas:

możemy przedstawić proces X_t jako $\sum_{j=0}^{\infty} \phi_j X_{t-j} = Z_t$, gdzie $\phi_0 = 1$ oraz $\sum_{j=0}^{\infty} \phi_j^2 < \infty \iff$ funkcja $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$, $|z| < 1$ jest taka że $\int_0^{2\pi} \frac{1}{|\Psi(\rho e^{i\lambda})|^2} d\lambda$ jest ograniczona gdy $\rho \rightarrow 1$.

Bibliografia

- [1] Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4):701–726.
- [2] Alonso, A. M., Peña, D. and Romo, J. (2001). Introducing model uncertainty in time series bootstrap. *Working Paper 01-14*, Statistics and Econometric Series 9, Universidad Carlos III de Madrid
- [3] Alonso, A. M., Peña, D. and Romo, J. (2002). Forecasting time series with sieve bootstrap. *Journal of Statistical Planning and Inference*, 100:1–11.
- [4] An, H. Z., Chen, Z. G. and Hannan, E.J. (1982). Autocorrelation, autoregression and autoregressive approximation. *Ann. Statist.*, 10:926–936.
- [5] Andersen, P. K. and Borgan, O. (1985). Counting process models for life history. *Scand. J. Statist*, 12:97–158.
- [6] Andersen, P., Borgan, O., Gill, R., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.
- [7] Anderson, T. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.
- [8] Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, 68(1):326–328.
- [9] Belyaev, Y. K. (1997). The continuity theorem and its application to resampling from sums of random variables. *Theory of Stochastic Processes*, 3(19):100–109.
- [10] Belyaev, Y. K. and Sjöstedt - de Luna, S. (2000). Weakly approaching sequences of random distributions. *Journal of Applied Probability*, 37:807–822.
- [11] Berk, K.N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.*, 2:489–502.
- [12] Berkowitz, J. and Kilian, L. (1996). Recent developments in bootstrapping time series. *Working Paper No. 1996-45*, Finance and Economics Discussion Series, Board of Governors of the Federal Reserve.

- [13] Berkowitz, J. and Birgean, I. and Kilian, L. (1999). On the finite-sample accuracy of nonparametric resampling algorithms for economic time series. *Working Paper*, Finance and Economics Discussion Series, Board of Governors of the Federal Reserve.
- [14] Bhansali, R. J. (1983). A simulation study of autoregressive and window estimators of the inverse autocorrelation function. *Appl. Statist.*, 32:141-149.
- [15] Bickel, P. and Bühlmann, P. (1997). Closure of linear processes. *J. Theoret. Probab.*, 10 str. 445-479.
- [16] Billingsley, P. (1987). *Prawdopodobieństwo i miara*. Państwowe Wydawnictwo Naukowe, Warszawa.
- [17] Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons.
- [18] Breidt, F. J., Davis, R. A. and Dunsmuir, W. (1992). On backcasting in linear time series models. *New Directions in Time Series Analysis*, str. 25-40.
- [19] Brockwell, P. and Davies, R. (1987). *Time Series Theory and Method*. Springer-Verlag.
- [20] Bühlmann, P. (1995). Moving-average representation of autoregressive approximations. *Stochastic Processes and their Applications*, 60: 331-342.
- [21] Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli*, 3(2):123-148.
- [22] Bühlmann, P. (1998). Sieve bootstrap for smoothing in nonstationary time series. *The Annals of Statistics*, 26(1):48-83.
- [23] Bühlmann, P. (2002). Bootstraps for Time Series. *Statistical Science*, 17(1):52-72.
- [24] Cao, R. et al. (1997). Saving Computer time in constructing consistent bootstrap prediction intervals for autoregressive processes. *Comm. Statist. Simulation Comput.*, 26(3):961-978.
- [25] Carlstein, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Annals of Statistics*, 14:1171-1179.
- [26] Carlstein, E. (1992). Resampling techniques for stationary time series: some recent developments. *New Directions in Time Series Analysis*, str. 75-85.
- [27] Chang, Y. and Park, J. Y. (2001). A Sieve Bootstrap for the Test of a Unit Root. *Department of Economics, Rice University, School of Economics, Seoul National University*.

- [28] Cheung, S. H., Wu, K. H. and Chan, W. S. (1998). Simultaneous prediction intervals for autoregressive integrated moving average models: A comparative study. *Computational Statistics & Data Analysis*, 28:297–306.
- [29] Chatfield Ch. (1998). Prediction Intervals. *Department of Mathematical Sciences, University of Bath*, Working paper.
- [30] Choi, E. i Hall, P.(2000). Bootstrap confidence regions computed from autoregressions of arbitrary order. *J. Royal Statist. Soc. Ser. B*, 62:461–477.
- [31] Deistler, M. and Hannan, E. J. (1998). *The Statistical Theory of Linear Systems*. Wiley, New York.
- [32] DiCiccio, J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3):189–228.
- [33] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7:1–26.
- [34] El-Nouty, Ch. and Guillou, A.(1999). On the smoothed bootstrap. *Journal of Statistical Planning and Inference*, 83:203–220.
- [35] Geman, S. and Hwang, C.R. (1982) Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.*, 10:401-414.
- [36] Gibbs, A. L. and Su, F. E. (2002). On Choosing and Bounding Probability Metrics. *Manuscript version: Departament of Mathematics and Statistics, York University, Canada and Departament of Mathematics, Harvey Mudd College, Claremont, USA*.
- [37] Grenander, U. (1981). *Abstract Inference*. Wiley, New York.
- [38] Grigoletto, M.(1998). Bootstrap prediction intervals for autoregressions: some alternatives. *International Journal of Forecasting*, 14:447–456.
- [39] Guerra, R., Polansky, A. M. and Schucany, W. R. (1997). Smoothed bootstrap confidence intervals with discrete data. *Computational Statistics & Data Analysis*, 26:163–176.
- [40] Hall, P. (1985). Resampling a coverage process. *Stochastic Process Applications*, 19:259–269.
- [41] Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *The Annals of Statistics*, 16:927–985.
- [42] Hall, P., DiCiccio T. J. and Romano J. P. (1989). On smoothing and the bootstrap. *The Annals of Statistics*, 17(2):692–704.

- [43] Hannan, E. J. (1987). Rational transfer function approximation. *Statist. Sci.*, 5:105–138.
- [44] Härdle, W., Horowitz, J. and Kreiss, J.P. (2001). Bootstrap methods for time series. *Working paper*.
- [45] Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.
- [46] Hurvich, C.M., Tsai, C.L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76:297–307.
- [47] Hyndman, R.J. (1995). Highest-density forecast regions for non-linear and non-normal time series models. *Journal of Forecasting*, 14:431–441.
- [48] Karr, A.F. (1987) Maximum likelihood estimation in the multiplicative intensity model via sieves. *Ann.Statist.*, 15:473-490.
- [49] Kim, J. H.(2001). Bootstrap-after-bootstrap prediction intervals for autoregressive models. *Journal of Business & Economic Statistics*, 19(1):117–128.
- [50] Klein, J. P. (1998), Melvin L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York.
- [51] Künsch, H.R.(1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17:1217–1241.
- [52] Leśkow J. (1988). Histogram maximum likelihood estimator of a periodic function in the multiplicative intensity model. *Statistics and Decisions*, 6:79–88.
- [53] Leśkow, J. (1989). A note on a kernel regularization of a histogram estimator in the multiplicative intensity model. *Statistics and Probability Letters*, 7:395–400.
- [54] Leśkow, J. (1993). Sieve-based maximum likelihood estimator for almost periodic stochastic processes models. *Probability and Mathematical Statistics*, 14(1):11–24.
- [55] Leśkow, J. and Róžański, R. (1989a). Histogram maximum likelihood estimator in the multiplicative intensity model. *Stochastic Processes and their Applications*, 31:151–159.
- [56] Leśkow, J. and Róžański, R. (1989b) Maximum likelihood estimator of the drift function for a diffusion process. *Statistics and Decisions*, 7:243–262.
- [57] Lewis, R. and Reinsel G. C. (1985). Prediction of Multivariate Time Series by Autoregressive Model Fitting. *Journal of Multivariate Analysis*, 16:393–411.
- [58] Liptser, R. and Shirayev, A. (1981). *Statistics of Random Processes*. Polish Edition, Warsaw.

- [59] Loève, M. (1977). *Probability Theory*. Springer-Verlag.
- [60] Masarotto, G. (1990). Bootstrap prediction intervals for autoregressions. *International Journal of Forecasting*, 6:229–239.
- [61] McKeague, I. (1986). Estimation for a semimartingale regression model using the method of sieves. *Ann. Stat.* 14:579–589.
- [62] Nadaraya, E. A. (1964). Some new estimates for distribution functions. *Theory Prob. Applic.*, 15:497–500.
- [63] Nguyen, H.T. and Pham, D.P. (1982). Identification of nonstationary diffusion model by the method of sieves. *SIAM J. Contr. Optim.* 20:603–611.
- [64] Polansky, A. M. and Schucany W. R. (1997). Kernel Smoothing to Improve Bootstrap Confidence Intervals. *Journal of the Royal Statistical Society*, 59(4):821–838.
- [65] Polansky, A. M. (2000). Stabilizing bootstrap-t confidence intervals for small samples. *The Canadian Journal of Statistics*, 28(3).
- [66] Polansky, A. M. (2001). Bandwidth selection for the smoothed bootstrap percentile method. *Computational Statistics & Data Analysis*, 36:333–349.
- [67] Politis, D.N. and Romano, J.P. (1993). The stationary bootstrap. *Journal of the American Statistical Association*, 89:1303–1313.
- [68] Politis, D.N. and Romano, J.P. and Wolf, M. (1999). *Subsampling*. Springer-Verlag, New York.
- [69] Prakasa Rao B. L. S. (1999). *Semimartingales and Their Statistical Inference*. Chapman.
- [70] Rammlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel function. *Ann. Statist.*, 11:453–466.
- [71] Rebolledo, R. (1980). Central limits theorems for local martingales. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 51:269–286.
- [72] Róžański, R. and Zagdański, A. (2001). On estimation in the multiplicative intensity model via histogram sieve. *Probability and Mathematical Statistics*, 21:89–99.
- [73] Róžański, R. and Zagdański, A. (2002). Estimation of the drift function for Itô processes and some class of semimartingales via histogram sieve. Wysłane do publikacji w *Applicationes Mathematicae*.
- [74] Said, S.E. and Dickey, D.A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71:599–608.

- [75] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York.
- [76] Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.*, 8:147–164.
- [77] Stine, R. A. (1987). Estimating properties of autoregressive forecast. *Journal of the American Statistical Association*, 82(400):1073–1078.
- [78] Swanepoel, J. W. H. (1990). A review of bootstrap methods. *South African Statistical Journal*, 24:1–34.
- [79] Tibshirani, R. (1988) Variance stabilization and the bootstrap. *Biometrika*, 75:433–444.
- [80] Thombs, L. A. and Schucany, R. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, 85(410):486–492.
- [81] Topsøe, F. (1967). On the connection between P -continuity and P -uniformity in weak convergence. *Teor. Verojatnost. i Primenen*, 12(2):279–288.
- [82] Walter, G.G. and Blum, J.R. (1984) A simple solution to a nonparametric maximum likelihood estimation problem. *Ann. Statist.*, 12:372–379.
- [83] Wei, W.W.S. (1990) Time Series Analysis. *Addison-Wesley*.
- [84] Zagdański, A. (1999). Using the sieve bootstrap method in time series analysis. *Operations Research Proceedings 1999, Springer*:165–169.
- [85] Zagdański, A. (2001a). Prediction intervals for stationary time series using the sieve bootstrap method. *Demonstratio Mathematica*, 34(2):469–481.
- [86] Zagdański, A. (2001b). On the application of the method of sieves in estimation of the intensity of multiplicative point processes. *Prace Naukowe Instytutu Matematyki Politechniki Wrocławskiej*, 24:129–136.

