

Wstęp do ukrytych modeli Markowa i metody Bauma–Welcha.

Piotr Wiktor Zwiernik

5 stycznia 2005

Streszczenie

W artykule przedstawimy szeroką klasę modeli, jaką są ukryte modele Markowa. Pomyślany jest on jako samodzielne wprowadzenie do teorii. Zaprezentowana zostanie definicja procesu oraz najpopularniejsza metoda estymacji jego parametrów. Ze szczególnym uwzględnieniem potraktujemy przypadek, gdy model posiada dwuelementową przestrzeń stanów. Opisane zostaną procesy o warunkowych prawdopodobieństwach emisji z rozkładem wykładniczym, Poissona oraz normalnym.

Spis treści

1	Ukryte modele Markowa	3
1.1	Wprowadzenie	3
1.2	Definicja ukrytych modeli Markowa	4
1.2.1	Założenia wstępne	4
1.2.2	Łańcuchy Markowa	5
1.2.3	Uwagi dotyczące przypadku dwustanowego	7
1.2.4	Ukryte modele Markowa	8
1.3	Estymacja parametrów HMM	12
1.3.1	Wprowadzenie do algorytmu Bauma–Welcha	12
1.3.2	Algorytm prefiksowy–sufiksowy	14
1.3.3	Algorytm Bauma–Welcha	16
2	Ukryte modele Markowa rozkładów dyskretnych	19
2.1	Przełączanie rozkładów Poissona	19
2.1.1	Wprowadzenie	19
2.1.2	Podstawowe statystyki procesu	20
2.1.3	Modelowanie liczby ataków padaczki	22
2.2	Przełączanie rozkładów wielomianowych	24
2.2.1	Wprowadzenie	24
2.2.2	Podstawowe statystyki procesu	25
2.2.3	Modelowanie ryzyka na rynku obligacji	26
2.2.4	Modelowanie erupcji gejzera „Old Faithful”	28
3	Ukryte modele Markowa rozkładów ciągłych	31
3.1	Markowskie przełączanie rozkładów normalnych	31
3.1.1	Wprowadzenie	31
3.1.2	Stabilna alternatywa metody Bauma–Welcha	31
3.2	Model analizy cyklu gospodarczego	33
3.2.1	Wprowadzenie	33
3.2.2	Postać modelu koniunktury	35
3.2.3	Modelowanie zwrotów koniunktury	36
4	Zakończenie	41

1 Ukryte modele Markowa

1.1 Wprowadzenie

Ukryte modele Markowa (ang. *Hidden Markov Models*, HMM) to pojęcie doskonale znane w zastosowaniach inżynierskich od prawie pół wieku. Pierwsze poważne zastosowanie znalazły one w 1967 roku w laboratoriach IBM, gdzie wykorzystano je do rozpoznawania znaków. Jak dotąd najszerszej korzysta się z nich tworząc modele akustyczne w rozpoznawaniu mowy[26]. Istnieją znaczące publikacje zbiorcze traktujące o metodologii i zastosowaniu ukrytych modeli Markowa. Za dobry przykład może służyć niedawna publikacja autorstwa R. Elliota, L. Aggouna oraz J. Moore[16], artykuł Y. Ephraima i N. Merhava[17] oraz liczne publikacje w takich czasopismach naukowych jak *IEEE Transactions on Signal Processing*, *IEEE Transactions on Speech and Audio Processing* oraz *Biometrics*.

Z biegiem lat zaczęto odkrywać, że te same modele sprawdzają się także do opisu innych zjawisk w przyrodzie. Podobne rozwiązania teoretyczne wykorzystywano w genetyce i biochemii, patrz na przykład prace Thompsona¹. Churchilla^{2,3}, Baldiego[4]. Zucchini i Guthorp⁴ używali ukrytych modeli Markowa do modelowania sekwencji dni suchych i wilgotnych w niektórych miejscach naszego globu. Albert[1] oraz Le, Leroux i Puterman[25][24] stosowali natomiast w swoich publikacjach ukryte modele Markowa do modelowania szeregów czasowych liczby ataków epilepsji. Hamilton[18] stworzył szeroko upowszechniony w ekonomii model cykli koniunkturalnych.

To najbardziej istotne obszary wykorzystania tych metod w modelowaniu procesów w otaczającej nas rzeczywistości. Lista zastosowań jest jednak jeszcze daleka od wyczerpania. Istnieje wiele dość egzotycznych prób modelowania z użyciem ukrytych modeli Markowa, jak na przykład szacowanie czasu między kolejnymi erupcjami gejzera Old Faithful czy systemy rozpoznawania twarzy. Ukryte modele Markowa znalazły również zastosowanie w naukach społecznych. Pojawili się naukowcy, którzy próbowali przy ich po-

¹[30] cyt. w:[27]

²[13] cyt. w:[27]

³[14] cyt. w:[27]

⁴[32] cyt. w:[27]

mocy przewidywać wydarzenia polityczne. P. Schrodtr na przykład zbudował model, który między innymi na podstawie depesz z agencji Reuters szacował prawdopodobieństwa wystąpienia napięć, w regionie Bałkanów[28].

W rozdziale tym przedstawimy podstawy teorii ukrytych modeli Markowa. Punktem wyjścia w podrozdziale 1.2 będzie pobieżne przedstawienie wyników z teorii łańcuchów Markowa. Następnie zaprezentujemy definicję ukrytych modeli Markowa. W podrozdziale 1.3 zaprezentujemy algorytmy wykorzystywane do efektywnej estymacji parametrów HMM, czyli algorytm prefiksowy–sufiksowy oraz algorytm Bauma–Welcha.

1.2 Definicja ukrytych modeli Markowa

1.2.1 Założenia wstępne

Wszystkie zmienne losowe określone są na standardowej przestrzeni probabilistycznej (Ω, \mathcal{F}, P) . Będziemy używać wielkich liter by oznaczyć zmienne losowe, małych dla ich realizacji oraz liter „pisanych” dla określenia zbiorów, z których przyjmują wartości zmienne losowe. I tak na przykład możemy napisać, że zmienna X przyjęła wartość $\{x\}$ ze zbioru \mathcal{X} .

Dyskretny proces stochastyczny będziemy oznaczać $(X_t)_{t \in \mathcal{T}}$, gdzie u nas $\mathcal{T} = \mathbb{N}$. Założymy, że $x_t \in \mathcal{X}$ dla każdego $t \in \mathcal{T}$. Z procesem stochastycznym związana jest mierzalna przestrzeń produktowa $(\mathcal{X}^{\mathcal{T}}, \mathcal{B}_{\mathcal{X}^{\mathcal{T}}})$, gdzie $\mathcal{B}_{\mathcal{X}}$ oznacza borelowskie σ -ciało otwartych podzbiorów przestrzeni \mathcal{X} względem ustalonej miary. Nas w szczególności będzie interesować przypadek, gdy proces stochastyczny zdefiniowany jest przez rozkład na $(\mathcal{X}^{\mathcal{T}}, \mathcal{B}_{\mathcal{X}^{\mathcal{T}}})$, który zależy od pewnego parametru. Niech $\phi \in \Phi$ oznacza ten parametr, gdzie Φ jest zbiorem parametrów, który u nas będzie podzbiorem \mathbb{R}^d dla pewnego d . Jako P_ϕ będziemy oznaczać rozkład procesu określony przez parametr ϕ .

Ciąg n zmiennych losowych X_i ($i = 1, 2, \dots, n$) będziemy oznaczać przez $X^{(n)}$, a ich realizację przez $x^{(n)}$. Niech $P_\phi^{(n)}$ oznacza n -wymiarowy rozkład $X^{(n)}$ indukowany przez P_ϕ . Dla każdego n zakładamy, że rozkład $P_\phi^{(n)}$ jest absolutnie ciągły względem pewnej σ -skończonej miary $\mu^{(n)}$. Gęstość względem tej miary oznaczamy $p(x^{(n)}; \phi)$. Wartość oczekiwaną mierzalnej funkcji $g(X^{(n)})$ względem miary probabilistycznej $P_{\phi_0}^{(n)}$ oznaczamy przez $E_{\phi_0}(g(X^{(n)}))$. W szczególnie interesującym nas przypadku, gdy $g(X^{(n)}) =$

$\ln(p(X^{(n)}; \phi))$, dana jest ona wzorem:

$$E_{\phi_0} \left(\ln(p(X^{(n)}; \phi)) \right) = \int \ln(p(X^{(n)}; \phi)) P_{\phi_0}(dx^{(n)}). \quad (1)$$

1.2.2 Łańcuchy Markowa

O łańcuchach Markowa (będziemy czasem oznaczać skrótowo — ŁM) powstało szereg publikacji. W niniejszej pracy przedstawione zostaną jedynie elementy teorii niezbędne do zgłębienia dalszej części tekstu⁵.

Definicja 1 Ciąg zmiennych losowych (X_t) o wartościach w przeliczalnym zbiorze S_X (przestrzeni stanów) nazywamy łańcuchem Markowa wtedy i tylko wtedy, gdy dla każdego $t \in \mathbf{N}$ i każdego ciągu $x_1, x_2, \dots, x_t \in S_X$ mamy

$$\begin{aligned} P(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_2 = x_2, X_1 = x_1) = \\ = P(X_t = x_t | X_{t-1} = x_{t-1}) \end{aligned} \quad (2)$$

jeśli tylko $P(X_{t-1} = x_{t-1}, \dots, X_2 = x_2, X_1 = x_1) > 0$.

Na łańcuch Markowa możemy patrzeć jako na proces stochastyczny. Warunek w definicji 1 mówi wtedy tyle, że ewolucja procesu zależy jedynie od bieżącego stanu. Zatem na kształtowanie się zmiennej X_t w każdej chwili t ma wpływ tylko i wyłącznie wartość procesu w chwili $t - 1$ (interesuje nas przypadek dyskretny). Jeżeli prawdopodobieństwa przejścia między stanami nie zależą od chwili, w której rozpatrujemy proces, to mówimy, że łańcuch Markowa jest jednorodny w czasie. Prawdopodobieństwo przejścia do stanu j ze stanu i będziemy wtedy oznaczać jako p_{ij} (gdzie $p_{ij} = P(X_2 = j | X_1 = i) = P(X_{t+1} = j | X_t = i)$ dla każdego t). Macierz $P = [p_{ij}]$ jednorodnego ŁM będziemy nazywać macierzą przejść.

Definicja 2 Jeśli $(X_t)_{t=1}^{\infty}$ jest łańcuchem Markowa, to rozkład zmiennej losowej X_1 nazywamy rozkładem początkowym.

⁵Podstawą tej części pracy był podręcznik [22]

Dla każdego stanu i jednorodnego łańcucha Markowa definiujemy średni czas pozostawania w nim:

$$v(i) = E(\inf_{n \geq 2} \{X_n \neq i\} | X_1 = i). \quad (3)$$

Jeżeli stan jest pochłaniający (tzn. $p_{ii} = 1$) przyjmujemy $v(i) = \infty$. W przeciwnym przypadku

$$v(i) = \frac{1}{1 - p_{ii}}. \quad (4)$$

Definicja 3 (i) *Mówimy, że stan j jest osiągalny ze stanu i , gdy możliwe jest przejście z i do j czyli $P(X_n = j | X_1 = i) > 0$ dla pewnego $n > 1$.*

(ii) *Stany i oraz j nazywamy wzajemnie komunikującymi się jeśli stan i jest osiągalny ze stanu j oraz j jest osiągalny ze stanu i .*

(iii) *Łańcuch Markowa nazywamy nieprzywiedlnym, gdy wszystkie jego stany wzajemnie komunikują się.*

Wprowadźmy teraz ważną definicję.

Definicja 4 *Mówimy, że rozkład prawdopodobieństwa $(\delta_j)_{j \in S_X}$ jest rozkładem stacjonarnym, gdy*

$$\delta = \delta P$$

czyli gdy $\sum_{j \in S_X} \delta_j = 1$ oraz dla każdego j jest spełnione $\delta_j \geq 0$ i $\delta_j = \sum_i \delta_i p_{ij}$.

Definicja 5 *Niech $d(i)$ oznacza największy wspólny dzielnik liczb $n > 1$, takich że spełnione jest $P(X_n = i | X_1 = i) > 0$. Stan i nazywamy stanem okresowym wtedy i tylko wtedy, gdy $d(i) > 1$. Wtedy $d(i)$ jest okresem stanu. Gdy $d(i) = 1$, to i jest stanem nieokresowym.*

W przypadku nieprzywiedlnych ŁM zachodzi ważny fakt, że wszystkie stany ŁM mają takie same własności. W szczególności zatem o nieokresowym łańcuchu Markowa będziemy mówić, jeżeli wszystkie jego stany są nieokresowe.

Zachodzi następujące

Twierdzenie 1 Niech $(X_t)_{t=1}^{\infty}$ będzie nieprzywiedlnym i nieokresowym, jednorodnym łańcuchem Markowa ze skończoną, N -elementową, przestrzenią stanów S_X i z macierzą przejścia P . Wtedy:

- (i) istnieje rozkład stacjonarny δ
- (ii) $\lim_{t \rightarrow \infty} P(X_t = j | X_1 = i) = \delta_j > 0$ dla wszystkich stanów i, j ,
- (iii) rozkład stacjonarny jest jedyny i $\delta_j = \frac{1}{\mu_j}$, gdzie μ_j jest średnim czasem powrotu łańcucha do stanu j .

Dla kompletności tej krótkiej dyskusji ŁM dodajmy jedynie, że dowodzi się istnienia ŁM dla zadanych: przestrzeni stanów, macierzy stochastycznej i rozkładu początkowego⁶.

1.2.3 Uwagi dotyczące przypadku dwustanowego

W pracy tej będziemy się koncentrować na łańcuchach Markowa o dwuelementowej przestrzeni stanów. Oczywiście w tym przypadku o ile $p_{12} \neq 0$ oraz $p_{21} \neq 0$, to oba stany komunikują się wzajemnie, a ŁM jest nieprzewiedlny.

Macierz przejścia dla takiego ŁM możemy zapisać jako:

$$\begin{bmatrix} 1 - p_{12} & p_{12} \\ p_{21} & 1 - p_{21} \end{bmatrix} \quad (5)$$

Istnieje wtedy również rozkład stacjonarny δ i jest dany wzorem:

$$\delta = \frac{1}{p_{12} + p_{21}} [p_{21}, p_{12}]. \quad (6)$$

Poprzez diagonalizację macierzy przejścia dla ŁM, łatwo uzyskujemy wzór na P^k :

$$P^k = \begin{bmatrix} \delta_1 & \delta_2 \\ \delta_1 & \delta_2 \end{bmatrix} + (1 - p_{12} - p_{21})^k \begin{bmatrix} \delta_2 & -\delta_2 \\ -\delta_1 & \delta_1 \end{bmatrix}, \quad (7)$$

Formuła ta pozwala badać graniczne własności ŁM, a w szczególności szacować szybkość zbieżności do rozkładu stacjonarnego.

⁶Wykorzystując fakt zgodności rozkładów skończeniowymiarowych, dostaniemy tezę jako prosty wniosek z twierdzenia Kołmogorowa o istnieniu procesu. Sposób dowodzenia podobnych zagadnień można znaleźć w [31].

1.2.4 Ukryte modele Markowa

Niech $(X_t)_{t=1}^{\infty}$ będzie łańcuchem Markowa o wartościach z N -elementowej przestrzeni stanów $S_X = \{1, 2, \dots, N\}$. Niech $x_t \in S_X$ oznacza wartość, jaką może przyjąć X_t . Niech $\delta_i = P(X_1 = i)$ oznacza prawdopodobieństwo, że stanem początkowym jest i . Wtedy $\delta = [\delta_i]$ jest wektorem $(1 \times N)$ reprezentującym rozkład początkowy. Zauważmy, że wprowadziliśmy to samo oznaczenie dla rozkładu stacjonarnego. Wynika to z tradycji modelowania za pomocą HMM, gdzie zakłada się, że rozkład początkowy procesu jest również jego rozkładem stacjonarnym. Takie założenie w wielu przypadkach jest zasadne zarówno na gruncie merytorycznym (gdy proces rzeczywiście zachowuje się od początku stacjonarnie) i numerycznym (w oszacowaniach, wartość rozkładu początkowego nie ma znaczenia dla dostatecznie dużej liczby obserwacji).

Będziemy dalej zakładać, że rozpatrywany ŁM jest jednorodny. Macierzą przejścia będzie macierz $P = [p_{ij}]$, gdzie $i, j = 1, \dots, N$ oraz $p_{ij} = P(X_t = j | X_{t-1} = i)$.

Rozważmy złożony proces $(X_t, Y_t)_{t=1}^{\infty}$ taki, że $(X_t)_{t=1}^{\infty}$ jest łańcuchem Markowa, a $(Y_t)_{t=1}^{\infty}$ jest ciągiem zmiennych losowych przyjmujących wartości ze zbioru S_Y . Niech $y_t \in S_Y$ oznacza wartość, jaką może przyjąć Y_t . W każdej chwili t i dla każdego $i \in S_X$ na zbiorze S_Y zdefiniowaną mamy gęstość prawdopodobieństwa $f_i(y)$, która określa rozkład zmiennej losowej Y_t względem pewnej σ -skończonej miary μ . W pracy tej będą nas interesować tylko dwa przypadki — miara Lebesgue'a i miara licząca.. Gęstości takiej postaci będziemy nazywali gęstościami emisji.

Tak zdefiniowany proces $(X_t, Y_t)_{t=1}^{\infty}$ nazywamy ukrytym procesem Markowa. „Ukrytość” polega na założeniu, że obserwujemy jedynie realizacje procesu (Y_t) , podczas gdy realizacje (X_t) wpływając na realizacje (Y_t) pozostają niewidoczne. W związku z tym o (X_t) będziemy mówić jako o ukrytym ŁM. Zbiór jego wartości będziemy nazywać przestrzenią stanów ukrytego procesu Markowa. Wartości procesu (Y_t) będziemy natomiast nazywać znakami emitowanymi przez HMM lub po prostu obserwacjami.

W przypadku skończonego zbioru znaków $S_Y = 1, 2, \dots, M$ ukrytego procesu Markowa, S_Y będziemy nazywać alfabetem, a prawdopodobieństwa

emisji będziemy zapisywać w macierzy $\Pi = [\pi_{xy}]$, gdzie $x \in S_X$ i $y \in S_Y$. Macierz $(N \times M)$ prawdopodobieństw emisji będziemy nazywać macierzą emisji.

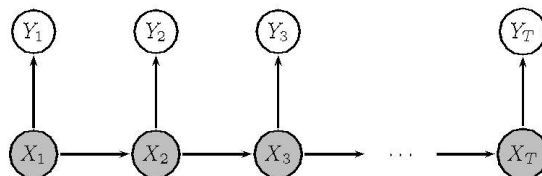
Zauważmy, że spełnione jest równanie:

$$\begin{aligned} P(Y_1, Y_2, \dots, Y_T | X_1, X_2, \dots, X_T) &= \\ &= \frac{P(X_1)P(Y_1|X_1) \cdot P(X_2|X_1) \cdot P(Y_2|X_2) \cdot \dots \cdot P(X_T|X_{T-1}) \cdot P(Y_T|X_T)}{P(X_1) \cdot P(X_2|X_1) \cdot \dots \cdot P(X_T|X_{T-1})} \\ &= P(Y_1|X_1) \cdot P(Y_2|X_2) \cdot \dots \cdot P(Y_T|X_T) \end{aligned} \quad (8)$$

Zatem możemy napisać:

$$P(Y_1, Y_2, \dots, Y_T | X^{(T)}) = P(Y_1|X^{(T)}) \cdot P(Y_2|X^{(T)}) \cdot \dots \cdot P(Y_T|X^{(T)}) \quad \text{dla każdego } T \quad (9)$$

gdzie $X^{(T)} = (X_1, X_2, \dots, X_T)$. Oznacza to, że zmienne losowe Y_1, Y_2, \dots są niezależne pod warunkiem znajomości procesu $(X_t)_{t=1}^\infty$. Sytuację tą przedstawia rysunek 1. Własność niezależności obserwacji względem procesu niewidzialnego to niezwykle ważna cecha HMM. Na bazie tej własności tworzone są efektywne filtry szacujące parametry modeli, z których najbardziej popularny zostanie przedstawiony w dalszej części pracy.



Rysunek 1: Węzły grafu reprezentują zmienne losowe. Zaciemnione zostały zmienne ukryte. Strzałki między węzłami mówią o bezpośredniej zależności stochastycznej. Brak strzałki między węzłami oznacza niezależność losową pod warunkiem węzłów-rodziców.

Rozpatrując HMM natrafiamy na szereg problemów, które zilustruje poniższy prosty przykład.

Przykład 1 (Nieuczciwe kasyno gry) *Przykładem ukrytego modelu Markowa jest tak zwane „nieuczciwe kasyno gry”. Przypuśćmy, że w kasyno mamy dwa rodzaje kości: uczciwą kostkę do gry (z prawdopodobieństwem $\frac{1}{6}$ wypada każda z sześciu możliwych wartości) oraz nieuczciwą kostkę (dla której*

prawdopodobieństwo wyrzucenia szóstki jest równe $\frac{1}{2}$, a dla pozostałych liczb $\frac{1}{10}$). Tak więc mamy dwa stany: F (uczciwa kostka) i L (nieuczciwa). Układ może zmieniać swój stan z pewnym prawdopodobieństwem, ale my stanu nie możemy zaobserwować (krupier zmienia kostki pod stołem!). To, którą kostkę rzucamy zależy tylko i wyłącznie od stanu ukrytego łańcucha Markowa. Niech jego macierzą przejścia będzie na przykład:

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.45 & 0.55 \end{bmatrix},$$

gdzie jako pierwszy oznaczymy stan F , a jako drugi stan L .

Średni czas nieprzerwanego rzucania kostką uczciwą jest równy $\frac{1}{1-0.9} = 10$ okresów, a kostką fałszywą $\frac{1}{1-0.55} = 2.2$ okresów. I choć nie wiemy, którą rzucamy kostką to mamy jednak jakąś o tym informację — choć nie bezpośrednio. Obserwujemy bowiem ciąg liczb będących wynikiem rzutów kostką. Macierz emisji dla tego przykładu wygląda tak:

$$\Pi = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{2} \end{bmatrix}.$$

Rozważając powyższy przykład możemy się natknąć na trzy ważne problemy:

1. Znając szereg 1000 kolejnych wyników rzutów, chcemy znaleźć prawdopodobieństwo pojawienia się takiego ciągu przy zadanych parametrach modelu. Albo bardziej formalnie: mając ciąg obserwacji $y^{(T)} = (y_1, y_2, \dots, y_T)$ oraz parametry modelu $\theta = (\delta, P, \Pi)$ chcemy znaleźć $P(y^{(T)}|\theta)$, czyli wartość, którą w przypadku ukrytych modeli Markowa rozpatrywanych w tej pracy będziemy traktować jako wartość funkcji wiarygodności $\mathcal{L}(\theta)$.
2. Znając szereg 1000 kolejnych wyników rzutów, chcemy znaleźć najbardziej prawdopodobny (nieobserwowalny!) ciąg typu $(\dots, F, F, F, L, F, L, F \dots)$, który ten szereg wygenerował. Albo bardziej formalnie: mając ciąg obserwacji $y^{(T)} = (y_1, y_2, \dots, y_T)$ chcemy znaleźć ciąg stanów $X^{(T)} = (x_1, x_2, \dots, x_T)$, który z największym prawdopodobieństwem mógł wyemitować rozważaną sekwencję znaków.

3. Znając szereg 1000 kolejnych wyników rzutów, chcemy oszacować, dla jakich wartości parametrów modelu, wygenerowanie takiego szeregu byłoby najbardziej prawdopodobne. Trochę formalniej: mając ciąg obserwacji $y^{(T)} = (y_1, y_2, \dots, y_T)$ chcemy znaleźć δ , P oraz Π tak by zmaksymalizować wartość prawdopodobieństwa $P(y^{(T)}|\theta)$.

W dalszej części tej pracy przedstawione zostaną sposoby rozwiązywania problemów pierwszego i trzeciego. Do rozwiązania drugiego problemu używa się zwykle tak zwanego algorytmu Viterbiego. Jednak problem postawiony w punkcie 2 nie interesuje nas w kontekście późniejszych analiz. Opis algorytmu Viterbiego zostanie zatem pominięty.

W przypadku pierwszego z problemów, nasuwa się prosta metoda liczenia prawdopodobieństwa $P(y^{(T)}|\theta)$. Znamy przecież parametry modelu, więc jeśli zaobserwowaliśmy $y^{(T)}$ to wystarczy zsumować po wszystkich możliwych ciągach x_1, x_2, \dots, x_T wyrażenie postaci:

$$P(X_1 = x_1)P(Y_1 = y_1|X_1 = x_1) \dots P(X_T = x_T|X_{T-1} = x_{T-1})P(Y_T = y_T|X_T = x_T).$$

Widać to wyraźnie, jak się spojrzy na graf na rysunku 1.

Dowodzi się jednak, że taka operacja wymaga rzędu $2T \cdot N^T$ operacji arytmetycznych, a to o wiele za dużo by algorytm ten miał rzeczywiste zastosowanie. Okazuje się, że procedura oparta na obliczaniu prawdopodobieństw prefiksowych i sufiksowych, opisana w podrozdziale 1.3, wymaga już tylko rzędu N^2T operacji. Jak ogromna to różnica można sobie uświadomić choćby na przykładzie z nieuczciwym kasynem. Dla $N = 2$ i $T = 1000$ w drugim przypadku mamy około 2^{12} operacji arytmetycznych, a w pierwszym już około 2^{1010} . Różnica jest więc kolosalna.

Podsumujmy podstawowe własności ukrytych modeli Markowa:

- stan HMM jest nieobserwowalny
- obserwacje Y_t w sposób losowy zależą od stanów procesu X_t
- wystąpienie każdego kolejnego stanu ukrytego łańcucha Markowa zależy wyłącznie od jego poprzednika.

Dyskretny ukryty model Markowa z N stanami i M znakami jest zdefiniowany przez $\theta = [\delta, P, \Pi]$, gdzie:

- $\delta \in \mathbb{R}^N$: $\delta_i = P(X_1 = i)$, wektor rozkładu początkowego
- P : $p_{ij} = P(X_{t+1} = j | X_t = i)$ macierz przejścia między stanami HMM
- Π : $\pi_{is} = P(Y_t = s | X_t = i)$ macierz emisji znaków

Ciągły ukryty model Markowa z N stanami jest zdefiniowany przez $\theta = [\delta, P, \pi_1, \pi_2, \dots, \pi_N]$, gdzie:

- $\delta \in \mathbb{R}^N$: $\delta_i = P(X_1 = i)$, wektor rozkładu początkowego
- P : $p_{ij} = P(X_{t+1} = j | X_t = i)$ macierz przejścia między stanami HMM
- π_i : $\pi_i = f_i(y)$ gęstość emisji znaków w i -tym stanie

1.3 Estymacja parametrów HMM

1.3.1 Wprowadzenie do algorytmu Bauma–Welcha

Najbardziej obecnie popularną metodą estymacji parametrów HMM zaprezentował L.E. Baum wraz ze współpracownikami w artykułach opublikowanych między 1966 a 1972 rokiem^{[7]⁷[9]⁸[5]}. Nazwisko Welcha pojawiło się jedynie w pracy z 1970 roku. Był tam wymieniony jako współautor pracy Bauma z wcześniejszego okresu. Ich praca nie została jednak nigdy opublikowana.

Algorytm Bauma–Welcha jest iteracyjnym sposobem lokalnej maksymalizacji funkcji wiarygodności $\mathcal{L}(\theta)$. Załóżmy, że ciąg obserwacji $y^{(T)}$ został wygenerowany przez HMM z prawdziwym parametrem $\theta_0 \in \Theta$. Niech $\mathcal{L}(\theta) = P(y^{(T)} | \theta)$ oznacza funkcję wiarygodności dla tego HMM, a $l(\theta) = \ln \mathcal{L}(\theta)$. Estymator największej wiarygodności parametru θ_0 jest wynikiem zadania:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} l(\theta). \quad (10)$$

Oznaczmy jako $P_\theta^{(T)}$ T -wymiarowy rozkład ciągu zmiennych $X^{(T)} = (X_1, X_2, \dots, X_T)$.

Załóżmy, że oszacowanie $\theta_n \in \Theta$ parametru θ_0 zostało obliczone na koniec n -tej iteracji. Niech $\hat{\theta}$ oznacza jakieś inne oszacowanie parametru θ_0 .

⁷[6] cyt. w:[27]

⁸[8] cyt. w:[27]

Dla danego ciągu obserwacji $y^{(T)}$ i każdej pary parametrów $\theta_n, \hat{\theta} \in \Theta$ zdefiniujemy funkcję pomocniczą Q określoną następująco (porównaj równanie 1):

$$Q(\hat{\theta}, \theta_n) = E_{\theta_n} \left(\ln P(X^{(T)}, y^{(T)} | \hat{\theta}) \middle| y^{(T)} \right). \quad (11)$$

Korzystając z nierówności Jensena dotyczącej funkcji wypukłych możemy oszacować różnicę $l(\hat{\theta}) - l(\theta_n)$:

$$\begin{aligned} l(\hat{\theta}) - l(\theta_n) &= \ln \frac{P(y^{(T)} | \hat{\theta})}{P(y^{(T)} | \theta_n)} \\ &= \ln E_{\theta_n} \left\{ \frac{P(X^{(T)}, y^{(T)} | \hat{\theta})}{P(X^{(T)}, y^{(T)} | \theta_n)} \middle| y^{(T)} \right\} \\ &\geq E_{\theta_n} \left\{ \ln \frac{P(X^{(T)}, y^{(T)} | \hat{\theta})}{P(X^{(T)}, y^{(T)} | \theta_n)} \middle| y^{(T)} \right\} \\ &= Q(\hat{\theta}, \theta_n) - Q(\theta_n, \theta_n). \end{aligned}$$

Równość w powyższym oszacowaniu zachodzi wtedy i tylko wtedy, gdy:

$$P(X^{(T)}, y^{(T)} | \hat{\theta}) = P(X^{(T)}, y^{(T)} | \theta_n) \quad \text{p.w. względem } P_{\theta_n}. \quad (12)$$

Nowe oszacowanie na parametr θ_0 w $(n+1)$ -iteracji uzyskujemy jako wynik działania:

$$\theta_{n+1} = \arg \max_{\theta \in \Theta} Q(\theta, \theta_n). \quad (13)$$

Ponieważ oczywiście $\theta_n \in \Theta$ zatem, można zauważyć, że z powyższych nierówności otrzymujemy $l(\theta_{n+1}) \geq l(\theta_n)$. W istocie zatem algorytm Bauma–Welcha jest wczesną wersją algorytmu EM (ang. *expectation-maximization*), który opisany został szczegółowo w 1977 roku przez Dempstera, Lairda i Rubina[2].

Pierwotnie algorytm B–W powstał dla zastosowań w procesach rozpoznawania mowy, w problemie podobnym do tego, o którym mówi poniższy przykład.

Przykład 2 (Słownikowa metoda rozpoznawania mowy) *Chcemy być w stanie rozpoznać wypowiedź, o której wiemy, że składa się ze znaków ze*

znanego słownika z V słowami. Na podstawie kilku powtórzeń słowa algorytm buduje statystyczny model danego słowa w słowniku (oszacowania rozkładu początkowego HMM, macierzy Π oraz P), który jest używany następnie przy rozpoznawaniu. Algorytm podobny do tego jest zastosowany w telefonach komórkowych ERICSSON T-18, gdzie osiąga on skuteczność rzędu 99%⁹. Każda wypowiedź generuje obserwację (sygnał akustyczny) y_1, y_2, \dots, y_T , którą rozważa się jako realizację długości T , procesu losowego (Y_t) o skończonej przestrzeni stanów. System posiada statystyczny model dla każdego słowa, jednak na skutek zaburzeń w momencie nadawania¹⁰ oraz niedoskonałości odbioru sygnału, różni się on od wzorca. Zadaniem słownikowej metody rozpoznawania mowy jest więc rozpoznanie wypowiedzi jak najskuteczniej.

Sam algorytm efektywnej estymacji parametrów HMM składa się z dwóch części: algorytmu prefiksowego–sufiksowego oraz właściwego algorytmu Bauma–Welcha. Czasem oba algorytmy łącznie nazywa się algorytmem Bauma–Welcha.

1.3.2 Algorytm prefiksowy–sufiksowy

O algorytmie wspominaliśmy już wcześniej. Motywacją w przykładzie z kasynem było efektywne policzenie prawdopodobieństwa pojawienia się danej sekwencji znaków $y^{(T)}$ przy zadanych parametrach modelu, czyli obliczenie funkcji wiarygodności $\mathcal{L}(\theta) = P(y^{(T)}|\theta)$.

Jest to metoda ogólna, my jednak przyjęliśmy konwencję zapisu zakładającą, że mamy do czynienia ze skończonym zbiorem S_Y . Jest to jednak kwestia tylko i wyłącznie oznaczeń i po przepisaniu wzorów z drobnymi poprawkami uzyskujemy formuły dla przypadku ciągłego (patrz np. Ephraim, Merhav[17]).

Procedurę rozpoczynamy obliczając prawdopodobieństwa prefiksowe $\alpha_t(i)$ i sufiksowe $\beta_t(i)$. Oblicza się je dla wszystkich stanów HMM oraz dla wszystkich okresów $t = 1, 2, \dots, T$:

⁹Szczegółowy opis algorytmu i znacznie szersze ujęcie problemu:

<http://helios.et.put.poznan.pl/~fkaluza/arm/opisprog.htm>

¹⁰Jaki model statystyczny poradzi sobie z płaczącym się językiem?

$$\alpha_t(i) = P(Y_1 = y_1, \dots, Y_t = y_t, X_t = i | \theta) \quad (14)$$

oraz

$$\beta_t(i) = P(Y_{t+1} = y_{t+1}, \dots, Y_T = y_T | X_t = i, \theta). \quad (15)$$

Zwróćmy uwagę, że przyjęta konwencja implikuje, że $\beta_T(i) = 1$ dla każdego stanu i .

Procedura obliczania prawdopodobieństw prefiksowych i sufiksowych jest rekurencyjna. Pierwsze wartości znamy dla zadanych parametrów modelu:

$$\alpha_1(i) = P(X_1 = i)P(Y_1 = y_1 | X_1 = i) = \delta_i \pi_{iy_1} \quad (16)$$

oraz

$$\beta_T(i) = 1. \quad (17)$$

Kolejne wartości obliczamy dla wszystkich $1 \leq t \leq T-1$ według wzorów.

$$\alpha_{t+1}(j) = \left(\sum_{i=1}^N \alpha_t(i) p_{ij} \right) \pi_{jy_{t+1}} \quad (18)$$

oraz

$$\beta_t(i) = \sum_{j=1}^N \pi_{jy_{t+1}} \beta_{t+1}(j) p_{ij}. \quad (19)$$

Powyższych wzorów dowodzi się stosunkowo prosto wychodząc z już wypisanych zależności¹¹.

Mając powyższe prawdopodobieństwa możemy wyznaczyć prawdopodobieństwo $P(y^{(T)} | \theta)$. Rzeczywiście: zauważmy przede wszystkim, że

$$\alpha_T(i) = P(Y_1 = y_1, \dots, Y_T = y_T, X_T = i | \theta),$$

a zatem spełnione jest:

$$P(y^{(T)} | \theta) = \sum_{i=1}^N \alpha_T(i). \quad (20)$$

¹¹Patrz na przykład: [27, s. 60–61].

W istocie istnieje wiele możliwości obliczenia tego prawdopodobieństwa, gdyż korzystając z wzajemnej niezależności Y_t otrzymujemy:

$$\begin{aligned}
\alpha_t(i)\beta_t(i) &= \\
&= P(X_t = i)P(Y_1 = y_1, \dots, Y_t = y_t | X_t = i)P(Y_{t+1} = y_{t+1}, \dots, Y_T = y_T | X_t = i) \\
&= P(X_t = i)P(Y_1 = y_1, \dots, Y_T = y_T | X_t = i) \\
&= P(Y_1 = y_1, \dots, Y_T = y_T, X_t = i)
\end{aligned} \tag{21}$$

zatem:

$$\sum_{i=1}^N \alpha_t(i)\beta_t(i) = P(Y_1 = y_1, \dots, Y_T = y_T) = \mathcal{L}(\theta) \tag{22}$$

Z równania 22 widać, że jeżeli umiemy obliczyć prawdopodobieństwa prefiksowe i sufiksowe dla każdego okresu t , to mamy T sposobów na wyznaczenie funkcji wiarygodności, gdyż równanie to jest prawdziwe dla każdego $t \in \{1, 2, \dots, T\}$.

Można, za sugestią Levisona, Rabinera, Sondhi [26], efektywniej zapisać ten algorytm za pomocą rachunku macierzowego. Definiując dla każdej chwili t wektory: $\alpha_t = [\alpha_t(1), \dots, \alpha_t(N)]$ oraz $\beta_t = [\beta_t(1), \dots, \beta_t(N)]$ funkcję wiarygodności możemy zapisać jako:

$$\mathcal{L}(\theta) = \alpha_t \beta_t' \tag{23}$$

dla dowolnego t , gdzie β_t' oznacza transpozycję wektora β_t .

Jeżeli zdefiniujemy natomiast macierz $B_t = P\lambda(y_t)$, gdzie $\lambda(y_t)$ jest macierzą diagonalną z kolejnymi elementami na przekątnej $\pi_{1y_t}, \dots, \pi_{Ny_t}$, to rekurencyjne wzory 18 oraz 19 możemy zapisać w postaci:

$$\alpha_{t+1} = \alpha_t B_{t+1} \tag{24}$$

oraz

$$\beta_t = \beta_{t+1} B_{t+1}' \tag{25}$$

1.3.3 Algorytm Bauma–Welcha

Algorytm Bauma–Welcha składa się z trzech formuł reestymacyjnych dla parametrów modelu. Zwiększają one w każdym kroku wartość $\mathcal{L}(\theta) = P(y^{(T)} | \theta)$. Sprowadza się to do rozwiązania zadania z wzoru 10.

Naszym celem jest przedstawienie wzorów na estymatory parametrów rozkładu początkowego, prawdopodobieństw przejść oraz estymatory parametrów warunkowych gęstości emisji ukrytego modelu Markowa. Jednak, jak to już zostało wcześniej wspomniane, w zastosowaniach opisanych w tej pracy, rozkład początkowy nie gra szczególnej roli i zakłada się, że jest równy rozkładowi stacjonarnemu. Pozwala to również ograniczyć liczbę szacowanych parametrów.

Zdefiniujmy wpierw dwa rodzaje prawdopodobieństw:

$$\gamma_t(i) = P(X_t = i | y^{(T)}, \theta) \quad (26)$$

oraz

$$\phi_t(i, j) = P(X_t = i, X_{t+1} = j | y^{(T)}, \theta). \quad (27)$$

W języku prawdopodobieństw prefiksowych i sufiksowych 26 oraz 27 możemy wyrazić jako:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(y^{(T)}|\theta)} \quad (28)$$

oraz

$$\phi_t(i, j) = \frac{\alpha_t(i)p_{ij}\pi_{iy_{t+1}}\beta_{t+1}(j)}{P(y^{(T)}|\theta)}. \quad (29)$$

Rzeczywiście równanie 28 wynika z zależności 21 oraz wzoru na prawdopodobieństwo warunkowe, gdyż:

$$P(X_t = i | y^{(T)}) = \frac{P(y_1, \dots, y_T, X_t = i)}{P(y_1, \dots, y_T)} = \frac{\alpha_t(i)\beta_t(i)}{P(y_1, \dots, y_T)} \quad (30)$$

Przy dowodzeniu drugiego równania skorzystamy z rekurencyjnego wzoru 19 oraz wzajemnej niezależności obserwacji względem ŁM:

$$\begin{aligned} P(X_t, X_{t+1} | y^{(T)}) &= \frac{P(y_1, \dots, y_T, X_t, X_{t+1})}{P(y_1, \dots, y_T)} = \\ &= \frac{P(y_1, \dots, y_T | X_t, X_{t+1})P(X_t, X_{t+1})}{P(y_1, \dots, y_T)} = \\ &= \frac{P(y_1, \dots, y_t | X_t, X_{t+1})P(y_{t+1}, \dots, y_T | X_t = i, X_{t+1} = j)P(X_t = i)p_{ij}}{P(y_1, \dots, y_T)} = \\ &= \frac{P(y_1, \dots, y_t | X_t)P(y_{t+1}, \dots, y_T | X_{t+1})P(X_t)p_{ij}}{P(y_1, \dots, y_T)} = \\ &= \frac{P(y_1, \dots, y_t | X_t)P(X_t)P(y_{t+1} | X_{t+1})P(y_{t+2}, \dots, y_T | X_{t+1})p_{ij}}{P(y_1, \dots, y_T)} = \\ &= \frac{\alpha_t(i)\pi_{jy_{t+1}}\beta_{t+1}(j)p_{ij}}{P(y_1, \dots, y_T)} \end{aligned} \quad (31)$$

Dysponując prawdopodobieństwami typu 26 i 27 możemy teraz zaproponować formuły reestymacyjne na parametry badanego HMM. W istocie są one intuicyjnym zastosowaniem powyższych obliczeń¹². Prawdopodobieństwo, że dla $t = 1$ układ znajduje się w stanie i jest równe $P(X_1 = i | \mathcal{I}_T) = \gamma_1(i)$, gdzie $\mathcal{I}_T = (y^{(T)}, \theta)$ jest całą dostępną informacją w każdej kolejnej iteracji. Na podstawie 28 możemy więc napisać:

$$\hat{\delta}_i = \frac{\alpha_1(i)\beta_1(i)}{\mathcal{L}(\theta)}. \quad (32)$$

Elementy macierzy przejść między stanami mówią nam dokładnie jakie jest prawdopodobieństwa przejścia ze stanu i do stanu j w jednym kroku. Najlepszym estymatorem dla p_{ij} powinna być częstość pobytów w i taka, że w następnym okresie układ znajdował się w stanie j , a więc po prostu:

$$\hat{p}_{ij} = \frac{\sum_{t=1}^{T-1} \phi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} = \frac{p_{ij} \sum_{t=1}^{T-1} \alpha_t(i) \pi_{jy_{t+1}} \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \alpha_t(i) \beta_t(i)}. \quad (33)$$

Elementy macierzy emisji mówią nam z jakim prawdopodobieństwem w stanie i zostanie wyemitowany znak s . Dobrym estymatorem powinna więc być częstość tych pobytów w stanie i dla których wyemitowanym znakiem było s , a więc po prostu.

$$\hat{\pi}_{jk} = \frac{\sum_{\{t:y_t=k\}} \gamma_t(i)}{\sum_{t=1}^{T-1} \gamma_t(i)} = \frac{\sum_{\{t:y_t=k\}} \alpha_t(i) \beta_t(i)}{\sum_{t=1}^T \alpha_t(i) \beta_t(i)}. \quad (34)$$

Widać, że o ile metoda prefiksowa–sufiksowa ma charakter ogólny, o tyle algorytm Bauma–Welcha wymaga pracowitego wyprowadzenia estymatorów największej wiarygodności dla konkretnych warunkowych rozkładów na obserwacje. Tutaj pokazaliśmy formuły reestymacyjne dla przypadku skończonej przestrzeni znaków. Przy omawianiu konkretnych zastosowań w następnych rozdziałach, również pokażemy odpowiednie formuły.

Okazuje się, że reestymowane parametry z równań 32, 34 oraz 33 zwiększają wartość funkcji wiarygodności \mathcal{L} . W swojej pracy Baum wraz ze współpracownikami [9] wykazali, że przy słabych założeniach ostra nierówność $\mathcal{L}(\hat{\delta}, \hat{\Pi}, \hat{P}) > \mathcal{L}(\delta, \Pi, P)$ jest zachowana poza przypadkiem, gdy (δ, Π, P) jest ekstremum funkcji wiarygodności, tzn. gdy $(\hat{\delta}, \hat{\Pi}, \hat{P}) = (\delta, \Pi, P)$.

¹²Część z tych obliczeń, a także dowody własności reestymowanych parametrów B–W można znaleźć w [27].

2 UKRYTE MODELE MARKOWA ROZKŁADÓW DYSKRETNYCH¹⁹

Cały proces estymacji parametrów HMM zaimplementowany na komputerze powinien więc wyglądać w skrócie tak:

1. oblicz prawdopodobieństwa prefiksowe i sufiksowe: GOTO 2
2. oblicz $\hat{\theta}$ według formuł reestymacyjnych algorytmu B–M: GOTO 3
3. jeżeli $|\mathcal{L}(\theta) - \mathcal{L}(\hat{\theta})| < \epsilon$ EXIT, w przeciwnym wypadku GOTO 4
4. podmień θ i $\hat{\theta}$: GOTO 1

Wystarczy do tego już tylko dodać, że ϵ wybieramy całkowicie arbitralnie tak aby z jednej strony oceny parametrów wyliczone przez algorytm były jak najbliższe tym prawdziwym, a z drugiej strony by liczba iteracji wykonywanych przez komputer nie musiała być za duża.

2 Ukryte modele Markowa rozkładów dyskretnych

2.1 Przełączanie rozkładów Poissona

2.1.1 Wprowadzenie

Przypuśćmy, że badamy dane zjawisko i obserwujemy ciąg liczb wystąpień jakiegoś zdarzenia. Łatwo możemy wyobrazić sobie odpowiednie przykłady. Niech to będzie liczba statków wpływających do portu w ciągu kolejnych dni, liczba żółtych kartek otrzymanych przez zawodników w danej kolejce polskiej 1. ligi piłkarskiej, liczba napadów rabunkowych na starej Pradze w kolejnych tygodniach etc. etc. Te przykłady można mnożyć w nieskończoność. Tego rodzaju zjawiska najprościej możemy modelować zakładając, że są generowane przez jednorodny proces Poissona¹³ i badać jego wartości w

¹³Przypomnijmy, że procesem Poissona z intensywnością $\lambda > 0$ nazywamy proces stochastyczny $N = (N_t)_{t \geq 0}$ taki, że:

1. $N_0 = 0$
2. proces ma przyrosty niezależne
3. dla $0 \leq s < t$ zmienna $N_t - N_s$ ma rozkład $Poiss(\lambda(t - s))$
4. trajektorie N są prawostronnie ciągłe.

ustalonych jednostkach czasu. Wynika z tego, że wyniki w tych zadanych parametrach rozłącznych okresach o tej samej długości są niezależnymi zmiennymi losowymi o identycznym rozkładzie Poissona, a więc ich wartość oczekiwana jest równa wariancji. Okazuje się jednak, że większość podobnych obserwacji wykazuje wariancję większą od średniej (ang. *overdispersion*). Co więcej nie spełniony jest również zwykle warunek wzajemnej niezależności (ang. *serial independence*) [27, s. 55–56].

Spróbujmy teraz wprowadzić alternatywny model, który pozwoli nam uwzględnić oba spostrzeżenia z empirii, a więc wzajemną zależność obserwacji (to co obserwowaliśmy w poprzednich okresach wpływa w jakiś sposób na obserwację okresu bieżącego) oraz wariancję przewyższającą średnią. Przypuśćmy, że każda pojedyncza obserwacja generowana jest przez jeden z dwóch procesów Poissona. Średnie obu procesów oznaczmy jako λ_1 oraz λ_2 , a wybór samej średniej zawdzięczamy jeszcze innemu procesowi losowemu, który nazwiemy procesem przełącznikowym. Niech λ_1 będzie wybierane z prawdopodobieństwem δ_1 , a λ_2 z prawdopodobieństwem $\delta_2 = 1 - \delta_1$. W tym przypadku wariancja wyników jest równa średniej, $\delta_1\lambda_1 + \delta_2\lambda_2$, plus nieujemne wyrażenie, $(\lambda_1 - \lambda_2)^2\delta_1\delta_2$. Uzyskaliśmy w ten sposób wariancję wyższą od średniej¹⁴.

Jeżeli wspomniany proces przełącznikowy jest ciągiem niezależnych zmiennych losowych, obserwacje również będą niezależne. Jeżeli natomiast procesem przełącznikowym będzie łańcuch Markowa, uzyskamy to czego szukamy, a więc wzajemną zależność obserwacji. Jest to prosty przykład wykorzystania ukrytych modeli Markowa w modelowaniu dyskretnych szeregów czasowych, który nadto inspirował wiele zastosowań dla tej klasy modeli.

2.1.2 Podstawowe statystyki procesu

Rozważamy zatem ukryty model Markowa, w którym warunkowy rozkład zmiennej Y_t jest rozkładem Poissona. Oznacza to, że jeśli $X_t = i$, Y_t ma rozkład Poissona ze średnią λ_i . Czyli warunkowa średnia zmiennej Y_t jest równa:

$$\mu(t) = \sum_{i=1}^N \lambda_i \mathbf{1}_{\{X_t=i\}}, \quad (35)$$

¹⁴Patrz: [27, 55–56]

gdzie N , jak poprzednio jest liczbą stanów ukrytego ŁM.

Prawdopodobieństwa warunkowe dla każdego $i = 1, 2, \dots, N$ oraz $s = 0, 1, 2, \dots$ dane jest wzorem:

$$\pi_{is} = e^{-\lambda_i} \frac{\lambda_i^s}{s!}.$$

Estymacja modelu wymaga oszacowania N^2 parametrów — N średnich λ_i i $N^2 - N$ pozadiagonalnych elementów macierzy przejścia dla ŁM. W pracy tej przyjmujemy $N = 2$, co oznacza potrzebę estymacji jedynie 4 parametrów. Odpowiednik formuły reestymacyjnej na warunkowe prawdopodobieństwo emisji (patrz wzór 34) dla przełącznikowych procesów Poissona, to:

$$\hat{\lambda}_i = \frac{\sum_{t=1}^T \alpha_t(i) \beta_t(i) y_t}{\sum_{t=1}^T \alpha_t(i) \beta_t(i)}. \quad (36)$$

Założmy teraz, że znamy parametry omawianego modelu. Jego średnia określona jest wzorem:

$$E(Y_t) = \sum_{i=1}^N E(Y_t | X_t = i) \delta_i = \sum_{i=1}^N \lambda_i \delta_i = \delta \lambda' \quad (37)$$

gdzie $\delta = (\delta_1, \delta_2, \dots, \delta_N)$, a $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)$.

A skoro:

$$E(Y_t^2) = \sum_{i=1}^N E(Y_t^2 | X_t = i) \delta_i = \sum_{i=1}^N (\lambda_i^2 + \lambda_i) \delta_i, \quad (38)$$

to wariancję możemy zapisać w postaci:

$$D^2(Y_t) = \sum_{i=1}^N (\lambda_i^2 - \lambda_i) \delta_i - \left(\sum_{i=1}^N \delta_i \lambda_i \right)^2 = \lambda D \lambda' + \delta \lambda' - (\delta \lambda')^2, \quad (39)$$

gdzie macierz D jest macierzą diagonalną kolejno z $\delta_1, \delta_2, \dots, \delta_N$ na przekątnej.

Ze znanego wzoru na wariancję warunkową,

$$D^2(Y_t) = E(D^2(Y_t | X_t)) + D^2(E(Y_t | X_t)) \quad (40)$$

możemy równoważnie napisać:

$$D^2(Y_t) = E(Y_t) + D^2(\mu(t)), \quad (41)$$

gdzie średnia warunkowa $\mu(t)$ jest dana wzorem 35.

W przypadku, gdy $N = 2$ wzór 41 prowadzi do poniższej zależności:

$$D^2(Y_t) = E(Y_t) + \delta_1 \delta_2 (\lambda_2 - \lambda_1)^2, \quad (42)$$

Z powyższego równania bezpośrednio widać dlaczego przełącznikowe modele Poissona charakteryzują się wariancją wyższą od średniej.

2.1.3 Modelowanie liczby ataków padaczki

Opisanie modelu przedstawionego w poprzednim podrozdziale zostało zainspirowane pracą Hopkins,Davies,Dobson[21]¹⁵. Hopkins i in. postulowali w niej, że istnieją dwa stany podatności pacjenta na ataki, z których każdy generuje intensywność ataków według rozkładu Poissona, oraz, że proces zmieniający stany jest łańcuchem Markowa. W pracy Hopkinsa i in. nie pojawił się jednak statystyczny opis modelu. Pełnego opisu dokonali dopiero Albert[1] oraz Le,Leroux,Puterman[24]. Jak przyznają MacDonald,Zucchini[27], modele tego typu wydają się niezwykle obiecującym sposobem analizy liczby ataków.

Trzynastu pacjentów, zostało poddanych badaniu przez okres od 3 miesięcy do 5 lat. Albert w swojej pracy wykazał, że ataki u wszystkich pacjentów poza jednym przypadkiem, charakteryzowały się wariancją wyższą od średniej procesu. Dwóch pacjentów z tej grupy zostało wybranych do ilustracji działania modelu. Pacjent I był obserwowany przez 422 dni i jego średnia liczba ataków była równa 0,68 na dzień. Pacjent II był obserwowany przez 351 dni ze średnią liczbą ataków na dzień równą 1,54.

Tabela 1 zawiera wyniki estymacji przedstawianego modelu wraz z odpowiednimi błędami standardowymi oszacowań. Z tabeli wynika, że najlepszy dwustanowy ukryty model Poissona dla pacjenta I charakteryzują średnie $\lambda_1 = 0,251$ i $\lambda_2 = 2$, a dla pacjenta II $\lambda_1 = 0,361$ i $\lambda_2 = 2,48$. Nie ta różnica sprawia jednak, że wyniki są niezwykle ciekawe. Dużo istotniejsza jest duża

¹⁵Cyt. w [1].

2 UKRYTE MODELE MARKOWA ROZKŁADÓW DYSKRETNYCH²³

Tabela 1: Estymatory największej wiarygodności dla dwustanowego przełącznikowego procesu Poissona. (Źródło: Albert[1, s. 1375]).

Parametr	Oszacowanie	Błąd St.
Pacjent I		
λ_1	0,251	0,045
λ_2	2	0,24
p_{12}	0,197	0,042
p_{21}	0,61	0,081
$\frac{\lambda_2}{\lambda_1}$	7,968	1,325
Pacjent II		
λ_1	0,361	0,074
λ_2	2,48	0,143
p_{12}	0,127	0,035
p_{21}	0,101	0,032
$\frac{\lambda_2}{\lambda_1}$	6,87	1,107

jakościowa różnica między prawdopodobieństwami zmian stanu przez ukryty ŁM. W przypadku drugiego pacjenta zarówno p_{12} jak i p_{21} kształtują się na niskim poziomie 0,1. W przypadku pacjenta pierwszego $p_{12} = 0,197$ a $p_{21} = 0,61$. Proces opisujący liczbę ataków epilepsji u pierwszego pacjenta znacznie częściej zmienia stan. Średnie czasy przebywania w stanach (patrz wzór 4) są równe dla Pacjenta I odpowiednio 5,08 dnia i 1,64 dnia, a dla Pacjenta II w pierwszym stanie 7,87 dnia, a w drugim 9,9 dnia.

Inną ważną statystyką przedstawioną w tabeli 1 jest iloraz średnich $\frac{\lambda_2}{\lambda_1}$. Iloraz ten jest miarą odpowiedniości modelu w porównaniu ze zwykłym rozkładem Poissona. Jeżeli jest bliski 1, oznacza to, że wprowadzenie dwóch osobnych stanów nie jest uzasadnione. Albert, wyprowadzając odpowiednie statystyki, wykazał w swojej pracy, że takie wartości ilorazów wskazują na to, że rzeczywiście model dwustanowy lepiej opisuje badane szeregi.

Widać to dodatkowo w wynikach przedstawionych w tabeli 2. Zostały w niej przedstawione rzeczywiste częstości obserwacji poszczególnych dziennych liczb ataków padaczki u pacjentów, i porównane z oczekiwanymi częstościami poszczególnych modeli.

Częstość E_i dla $Y = i$ (patrz tabela 2) była wyliczana według wzoru:

$$E_i = T \left\{ \frac{p_{12}}{p_{12} + p_{21}} \frac{\exp^{-\lambda_2} \lambda_2^i}{i!} + \frac{p_{21}}{p_{12} + p_{21}} \frac{\exp^{-\lambda_1} \lambda_1^i}{i!} \right\} \quad (43)$$

gdzie T to długość szeregu czasowego.

Widać zatem, że w analizie procesu generującego liczbę ataków epilepsji, dwustanowe przełącznikowe procesy Poissona spisują się bardzo dobrze.

2 UKRYTE MODELE MARKOWA ROZKŁADÓW DYSKRETNYCH²⁴

Tabela 2: Zaobserwowane i oczekiwane liczby ataków epilepsji. Porównanie wyników modelu dwustanowego i standardowego rozkładu Poissona. Źródło: Albert[1, s. 1375].

Y	obserwacje	Poisson	dwa stany
Pacjent I			
0	263	219,8	260,2
1	90	138	93,2
2	32	48,5	35,1
3	23	12,2	18,1
4	9	2,7	9,2
5	3	0,67	3,8
6	2	0,17	1,3
Pacjent II			
0	126	91,9	125,5
1	80	110,9	79,7
2	59	75,1	57,3
3	42	39	42,6
4	24	18,5	25,9
5	8	8,3	12,8
6	5	3,8	5,3
7	4	1,7	1,9
8	3	0,86	0,6

Osobną kwestią pozostają ich zdolności prognostyczne, co jest nierzadko pożądaną cechą modeli rzeczywistych zjawisk. Istnieje bardzo niewiele publikacji na temat zdolności prognostycznych tej klasy modeli i wydaje się, że to zagadnienie będzie wymagać jeszcze dalszych badań.

2.2 Przełączanie rozkładów wielomianowych

2.2.1 Wprowadzenie

W przypadku skończonej przestrzeni znaków modelowanie za pomocą rozkładu Poissona jest oczywiście niewłaściwie. Z drugiej strony podejście polegające na umieszczeniu warunkowych prawdopodobieństw emisji w macierzy o wymiarach $N \times M$ (gdzie N to liczba stanów ŁM, a M to liczba znaków w alfabecie HMM) wymaga szacowania w sumie aż $N^2 + N(M - 2)$, co robi się nieefektywnie dla większego M . Rozwiązaniem, które często stosuje się w takim przypadku jest wykorzystanie rozkładu wielomianowego. Pozwala to ograniczyć liczbę parametrów do oszacowania do liczby, która pojawiła się w przypadku rozkładu Poissona. Dla danej liczby prób n każdy warunkowy rozkład emisji dla stanu $i = 1, 2, \dots, N$ jest bowiem jednoznacznie dany przez prawdopodobieństwo sukcesu p_i . Zatem musimy oszacować N prawdopodobieństw sukcesu p_i oraz $N^2 - N$ pozadiagonalnych elementów macierzy przejścia.

2.2.2 Podstawowe statystyki procesu

Rozważmy zatem ukryty proces Markowa, dla którego obserwowany proces (Y_t) ma rozkład dwumianowy warunkowo na ŁM. Rozkład dwumianowy jest jednoznacznie zadany przez dwa parametry: n (liczba prób) i p (prawdopodobieństwo sukcesu w jednej próbie). Należy zwrócić uwagę, że w ogólnym przypadku modelowania przy pomocy przełączania rozkładów dwumianowych zakłada się, że n nie musi być stałe w czasie. W każdej chwili t parametrami warunkowych rozkładów dwumianowych są wtedy n_t oraz p_i dla $i = 1, 2, \dots, N$. My jednak przyjmiemy $n_t = n$ dla każdego $t = 1, 2, \dots, T$. Zatem średnia warunkowa dla widzialnego procesu zapisuje się poniższym wzorem:

$$p(t) = \sum_{i=1}^N p_i \mathbf{1}_{\{X_t=i\}}, \quad (44)$$

gdzie p_i jest prawdopodobieństwem sukcesu dla obserwacji związanym ze zdarzeniem $\{X_t = i\}$.

Zatem dla $s = 0, 1, \dots, n$ mamy:

$$\pi_{is} = \binom{n}{s} p_i^s (1 - p_i)^{n-s}. \quad (45)$$

W opisywanym przypadku formuła reestymacyjna na parametry rozkładów warunkowych przybiera postać:

$$\hat{p}_i = \frac{\sum_{t=1}^T \alpha_t(i) \beta_t(i) y_t}{n \sum_{t=1}^T \alpha_t(i) \beta_t(i)}. \quad (46)$$

Tak jak w przypadku poissonowskich ukrytych modeli Markowa, możemy łatwo wyprowadzić podstawowe statystyki omawianego procesu.

$$E(Y_t) = \sum_{i=1}^N (np_i) \delta_i = n\delta p', \quad (47)$$

gdzie $p = (p_1, \dots, p_N)$. Dodatkowo mamy:

$$E(Y_t^2) = \sum_{i=1}^N (np_i(1 - p_i) + n^2 p_i^2) \delta_i = n\delta p' + n(n - 1)pDp', \quad (48)$$

gdzie D jest macierzą diagonalną z elementami wektora δ na przekątnej.

Zatem wariancja obserwacji w dwustanowym wielomianowym HMM jest równa:

$$D^2(Y_t) = n^2(\delta\Phi p' - (\delta p')^2) + n(\delta p' - \delta\Phi p'), \quad (49)$$

gdzie Φ jest macierzą diagonalną z elementami wektora p na przekątnej.

2.2.3 Modelowanie ryzyka na rynku obligacji

Efekty współzależności między poszczególnymi papierami są kluczowym składnikiem ryzyka kredytowego portfela inwestycyjnego. Kwestią sporną pozostaje, jak modelować te współzależności. Można próbować modelować ryzyko niewypłacalności dla każdego emitenta oddzielnie i badać współczynniki korelacyjne. Podejście takie dla dużego portfela (powiedzmy $n > 50$) staje się mocno nieefektywne, wymagając ogromnej liczby zmiennych w modelu. W zamian tego proponowane są różne alternatywne podejścia prostszego opisu procesu interakcji między ryzykiem poszczególnych papierów. Wymogiem jest by model taki miał stosunkowo mało parametrów a także był uzasadniony na gruncie ekonomicznym i empirycznym (patrz [15]).

Jedną z propozycji, która zyskała sobie popularność, jest ukryty model Markowa z dwumianowymi warunkowymi prawdopodobieństwami emisji. Model jest takiej samej postaci, jak opisany w poprzedniej części. W tym przypadku zniesione zostało jednak założenie o tym, że liczba prób w rozkładzie dwumianowym jest stała. Jest to uzasadnione, gdyż liczba papierów w portfelu ulega zmianie. Dodać w tym miejscu należy, że w kwestii wyliczeń z poprzedniego rozdziału nic nie zmienia — dalej model opisują cztery parametry (średnie p_i są stałe w czasie), gdyż n_t dla każdego t jest dane i nie podlega estymacji.

Zakładamy, że portfel może znajdować się w jednym z dwóch stanów: ryzyko normalne i ryzyko podwyższone. Stany te są nieobserwowalne. Za pomocą dwumianowego HMM będziemy chcieli modelować liczbę „zdarzeń niewypłacalności” w portfelu.

Crowder, Davis, Giampieri [15] zastosowali powyższy model do dużego zbioru danych dotyczących historii zdarzeń niewypłacalności udostępnionych przez Standard & Poor's. Chcąc uzyskać wartościowszą interpretację, zmie-

2 UKRYTE MODELE MARKOWA ROZKŁADÓW DYSKRETNYCH 27

Tabela 3: Wyniki estymacji dla modelu ryzyka inwestycyjnego. Źródło: [15, s. 4]

Sektor	$N_{\text{całkowita}}$	$N_{\text{niewypłacalni}}$	λ	κ	p_{11}	p_{22}
samoch.	820	167	0,00121	2,7761	0,91367	0,96549
konsum.	1041	251	0,0018	7,38824	0,89472	0,7604
energ.	420	71	0,00123	8,25494	0,92074	0,84073
media	650	133	0,00269	6,7546	0,95939	0,82885

nili oni nieznacznie zapis prawdopodobieństw warunkowych do postaci:

$$\pi_{1s} = \binom{n_t}{s} \lambda^s (1 - \lambda)^{n_t - s} \quad (50)$$

$$\pi_{2s} = \binom{n_t}{s} (\kappa \lambda)^s (1 - \kappa \lambda)^{n_t - s}, \quad (51)$$

gdzie $0 < \lambda < 1$, a $1 \leq \kappa < \frac{1}{\lambda}$ i model degeneruje się dla $\kappa = 1$. Jest to zapis równoważny 45. Zwróćmy uwagę, że założenie $\kappa \geq 1$ oznacza wyróżnienie drugiego stanu, jako stanu większego ryzyka. Dodać jeszcze należy, że κ jest dobrą miarą tego, jak istotny jest podział na dwa stany. Analogiczna sytuacja pojawiała się w przypadku modelowania liczby ataków epilepsji i wartości statystyki $\frac{\lambda_2}{\lambda_1}$.

Celem badania były obligacje przedsiębiorstw działających w czterech dużych sektorach gospodarki amerykańskiej: samochodowym, konsumpcyjnym, energetycznym i mediach. Dane S&P dostarczają również informacji o momencie emisji obligacji. Momenty niewypłacalności zostały zagregowane w kwartalne okresy i tak przygotowany szereg poddany został standardowej dla HMM procedurze estymacyjnej.

W tabeli 3 przedstawione są dla każdego sektora: całkowita liczba emitentów, którzy pojawili się na rynku, liczba zdarzeń niewypłacalności, oraz parametry modelu. Zwraca uwagę fakt, że sektor samochodowy jest najmniej narażony na wahania ryzyka inwestycyjnego. W pozostałych sektorach natomiast zauważmy, że $p_{11} > p_{22}$. Może to oznaczać, że ważnym czynnikiem wpływającym na wahania ryzyka w tych sektorach jest czynnik koniunkturalny w gospodarce (kwestia ta będzie jeszcze dyskutowana w ostatnim rozdziale).

Wynikiem estymacji omawianego modelu są również wygładzone prawdopodobieństwa przebywania w poszczególnych stanach. Na ich podstawie można dokonać dynamicznej segmentacji badanego okresu na podokresy charakteryzujące się zwykłym i podwyższonym ryzykiem. Badanie zależności między występowaniem okresów o podwyższonym ryzyku inwestycyjnym w poszczególnych sektorach, pozwala dokonać analizy współzależności między ryzykiem inwestycyjnym na rynkach.

Badanie ryzyka inwestycyjnego z podziałem na sektory pozwala badać współzależności między obligacjami przedsiębiorstw z poszczególnych sektorów. Jednak estymacja modelu dla całego dostępnego zbioru danych może dodatkowo pomóc nam zanalizować globalne czynniki ekonomiczne wpływające na poziom ryzyka. W tym wypadku mieliśmy do czynienia (patrz [15]) z 9928 obligacjami, z których 1238 okazało się niewypłacalne. Procedura Bauma–Welcha wygenerowała wartości odpowiednich estymatorów: $\lambda = 0,0016$, $\kappa = 6,9012$, $p_{11} = 0,9262$ oraz $p_{22} = 0,5951$.

2.2.4 Modelowanie erupcji gejzera „Old Faithful”

Gejzer¹⁶ to rodzaj gorącego źródła, występujący w obszarze czynnego lub niedawno wygasłego wulkanu, charakteryzujący się okresowym wyrzucaniem gorącej wody i pary wodnej¹⁷. Gejzery wybuchają z różną częstotliwością (od kilkudziesięciu minut do kilkunastu dni) i siłą (woda jest wyrzucana na wysokość od kilku do 70 m). Gejzer „Old Faithful” w Narodowym Parku Yellowstone w stanie Wyoming jest najbardziej znanym gejzerem na świecie i drugą po misiu Yogi wizytówką parku. Jako atrakcja turystyczna działa regularnie od ponad stu lat, kiedy został odkryty. Co godzinę przez 4 minuty wyrzuca 40 tysięcy litrów wody na wysokość 45 metrów. Goście odwiedzający park chcą obejrzeć gejzer nie tracąc wiele czasu na czekanie. Dlatego właśnie motywacją w tym wypadku będzie jak najdokładniejsze określenie, kiedy nastąpi następna erupcja i ile będzie trwać.

Na zdjęciu 2 widzimy szacowany czas następnej erupcji. Na stronie Par-

¹⁶Termin „gejzer” pochodzi od islandzkiej nazwy największego kiedyś gejzera tej wyspy — Stóri Geysir (Wielka Fontanna), położonego w południowo-zachodniej części kraju (obecnie największy jest gejzer Strokkur).

¹⁷<http://wiem.onet.pl/wiem/002783.html>

2 UKRYTE MODELE MARKOWA ROZKŁADÓW DYSKRETNYCH²⁹



Rysunek 2: Na stronie Parku Yellowstone (<http://www.nps.gov/yell/oldfaithfulcam.htm>) możemy oglądać aktualny obraz gejzeru Old Faithful. Zwróćmy uwagę na liczby na górze — to godzina i szacowany czas następnej erupcji.

ku Yellowstone czytamy: „Wszystkie predykcje są generowane z wykorzystaniem formuły biorącej pod uwagę długość poprzedniej erupcji. Formuła ta podaje czas następnej erupcji z dokładnością do 10 minut.” Okazuje się, że ta tajemnicza formuła opiera się właśnie na dyskretnych przełącznikowych modelach Markowa. Azzalini i Bowman¹⁸ przedstawili interesujący model dotyczący powyższej sytuacji. Ich szeregi składały się z danych zbieranych co godzinę przez dwa tygodnie. Pierwszy szereg, d_t , dotyczył trwania poszczególnych erupcji, a drugi, w_t , dotyczył długości okresów pomiędzy kolejnymi początkami erupcji gejzera.

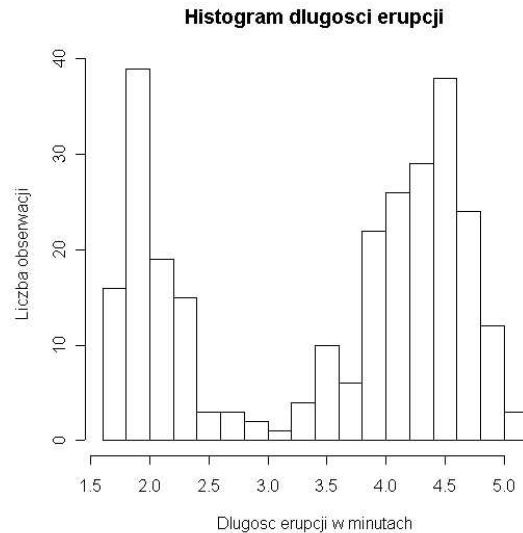
Oba szeregi charakteryzują się tym, że większość obserwacji da się opisać jako długie albo krótkie, a jedynie nieliczne obserwacje można uznać za średniej długości. Do tego w obu podgrupach (długie, krótkie) obserwowane są relatywnie niskie wariancje.

Jeżeli rozpatrzymy szereg czasowy długości kolejnych erupcji, to bardzo silnie odrzucamy hipotezę o tym, że długości te mają rozkład normalny. Jeżeli jednak podzielimy obserwacje na dwie grupy, to w każdej podgrupie hipoteza o normalności rozkładu już nie da się odrzucić. Na wykresie 3 przedstawiony został histogram dla tego szeregu czasowego. Bardzo wyraźnie widać dwumodalność rozkładu empirycznego. Granica podziału przebiega

¹⁸[3] cyt. w: [27]

2 UKRYTE MODELE MARKOWA ROZKŁADÓW DYSKRETNYCH³⁰

gdzieś około punktu 3 minut. Jak zatem widać, analiza histogramu może stać się kolejną ważną przesłanką za stosowaniem modeli podobnych do HMM.



Rysunek 3: Histogram długości erupcji gejzera Old Faithful (Źródło: opracowanie własne na podstawie danych z bazy programu statystycznego R)

Mamy zatem dwie jednorodne grupy o małej wariancji, a średnie między grupami znacząco się różnią. Z tego też powodu wydaje się naturalne traktować powyższe szeregi jako binarne. Azzalini i Bowman podzielili więc w obu szeregach obserwacje na dwie grupy, granicami stały się okresy 3 minutowe (dla długości erupcji) i 68 minutowe (długości okresów między erupcjami). Ciekawym jest fakt, że tak otrzymane dwa nowe szeregi były do siebie niezwykle podobne, prawie identyczne. Azzalini i Bowman postanowili więc w swojej analizie skoncentrować się na pierwszym z tych nowych szeregów (długość erupcji), twierdząc, że dostarcza on więcej informacji o stanie układu.

Ich podstawowy model opierał się na rozkładzie dwumianowym dla $n_t = n = 1$ i dla dwóch stanów. Stan pierwszy cechował się krótszymi erupcjami gejzera, a drugi dłuższymi. Oszacowane parametry dla tego modelu były równe:

$$\hat{P} = \begin{bmatrix} 0.000 & 1.000 \\ 0.827 & 0.173 \end{bmatrix}.$$

oraz

$$\hat{p} = [0.225, 1.000],$$

gdzie wektor p odpowiada prawdopodobieństwu sukcesu dla rozkładu dwumianowego w każdym ze stanów. Postać macierzy przejścia sugeruje, że po stanie odpowiadającym krótkiej erupcji następuje stan odpowiadający długiej erupcji. To bardzo ważna cecha szeregu i ciężko sobie wyobrazić prognozowanie czasu erupcji, jeżeli nie weźmie się pod uwagę tej jego własności.

3 Ukryte modele Markowa rozkładów ciągłych

3.1 Markowowskie przełączanie rozkładów normalnych

3.1.1 Wprowadzenie

Wyobraźmy sobie ciąg losowań przebiegających zgodnie z następującym schematem. Na każdym kroku wartość y_t jest losowana według jednego z rozkładów $N(\mu_i, \sigma_i)$, gdzie $i = 1, 2, \dots, N$.

O tym, z którego rozkładu odbywa się losowanie decyduje stan, w którym znalazł się ŁM (jest ich N). Dla obserwatora śledzącego proces, stany łańcucha pozostają niewidoczne. Odpowiednie gęstości warunkowe dla każdego okresu i dla $j = 1, 2, \dots, N$ możemy zapisać tak:

$$f(y_t | x_t = j, \theta) = \frac{1}{\sqrt{2\pi\sigma_j}} \exp\left\{-\frac{(y_t - \mu_j)^2}{2\sigma_j^2}\right\} \quad (52)$$

gdzie $\theta = [\mu_1, \dots, \mu_N, \sigma_1, \dots, \sigma_N]$. Naszym zadaniem będzie oszacowanie parametrów przełączanych rozkładów oraz macierzy przejścia, na podstawie ciągu obserwacji $y^{(T)}$.

3.1.2 Stabilna alternatywa metody Bauma–Welcha

Niech $\xi_{t|t} = [\xi_{t|t}(j)]$ będzie wektorem $(N \times 1)$ warunkowego prawdopodobieństwa przebywania w ukrytym stanie x_t pod warunkiem informacji do

chwili t , tzn. $\xi_{t|t}(j) = P(x_t = j|y^{(t)}; \theta)$. Dodatkowo niech $\xi_{t|t-1} = [\xi_{t|t-1}(j)]$ będzie wektorem $(N \times 1)$ warunkowego prawdopodobieństwa przebywania w ukrytym stanie x_t pod warunkiem informacji do chwili $t-1$, tzn. $\xi_{t|t-1}(j) = P(x_t = j|y^{(t-1)}; \theta)$.

Niech η_t będzie wektorem, którego kolejnymi elementami będą prawdopodobieństwa warunkowe postaci 52. Estymatory wartości $\xi_{t|t}$ przy zadanym $\hat{\xi}_{1|0}$ ¹⁹ i zadanych parametrach modelu θ mogą być wyliczone w bardzo efektywny sposób dla każdego $t = 1, 2, \dots, T$ (patrz: [20]):

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \odot \eta_t}{\mathbf{1}'(\hat{\xi}_{t|t-1} \odot \eta_t)} \quad (53)$$

$$\hat{\xi}_{t+1|t} = P\hat{\xi}_{t|t} \quad (54)$$

gdzie symbol \odot oznacza mnożenie wektorów element po elemencie, $\mathbf{1}'$ jest wektorem $(N \times 1)$ złożonym z jedynek, a P jest macierzą przejść ukrytego łańcucha Markowa.

Wartość funkcji wiarygodności jest produktem ubocznym tego filtru:

$$\mathcal{L}(\theta) = \sum_{t=1}^T \ln f(y_t, y^{(t-1)}; \theta), \quad (55)$$

gdzie

$$f(y_t, y^{(t-1)}; \theta) = \mathbf{1}'(\hat{\xi}_{t|t-1} \odot \eta_t). \quad (56)$$

Można pokazać (patrz: [17]), że ten sposób na efektywne wyliczenie funkcji wiarygodności dla zadanych parametrów modelu jest numerycznie stabilniejszy od metody zaproponowanej w rozdziale pierwszym. Jest to metoda powszechnie używana w przypadku ogólniejszej klasy modeli - przełącznikowych modeli Markowa, w których proces przełącznikowy rządzi przełączaniem się parametrów regresji.

Wyładzone prawdopodobieństwa mogą być wyliczone na podstawie algorytmu zaproponowanego przez Kima[23]. Używając tych samych oznaczeń co poprzednio możemy zapisać go tak:

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot [P \cdot (\hat{\xi}_{t+1|T}(\div) \hat{\xi}_{t+1|t})], \quad (57)$$

¹⁹Wydaje się uzasadnione przyjąć $\hat{\xi}_{1|0} = \delta$, gdzie δ jest rozkładem stacjonarnym dla ukrytego łańcucha Markowa.

gdzie znak (\div) oznacza dzielenie po współrzędnych. Zauważmy, że tutaj również obliczamy je rekurencyjnie z tym, że poruszamy się wstecz kolejno $t = T - 1, T - 2, \dots, 1$, a wartość $\hat{\xi}_{T|T}$ jest określona wzorem 53.

W przypadku stosowania algorytmu EM w każdej iteracji szacowania modelu uzyskujemy nowe oceny estymatorów średnich μ_i , wariancji σ_i^2 oraz elementów macierzy przejścia między stanami ŁM. Maksymalizację funkcji wiarygodności w każdej iteracji algorytmu, możemy przeprowadzać albo w drodze bezpośredniego numerycznego wyszukiwania jej ekstremów, albo poprzez zastosowanie wzorów reestymacyjnych.

Hamilton[19] wykazał, że estymatorami największej wiarygodności dla prawdopodobieństw przejścia są²⁰:

$$\hat{p}_{ij} = \frac{\sum_{t=2}^T P(x_t = j | y^{(T)}; \hat{\theta}) \frac{P(x_{t-1}=i | y^{(t-1)}; \hat{\theta})}{P(x_t=j | y^{(t-1)}; \hat{\theta})}}{\sum_{t=2}^T P(x_{t-1} = i | y^{(T)}; \hat{\theta})} \quad (58)$$

gdzie $\hat{\theta}$ oznacza pełny wektor nowych oszacowań dla parametrów modelu uzyskany w poprzedniej iteracji algorytmu.

Odpowiednie wektory średnich przełączanych rozkładów uzyskamy dla każdego stanu ukrytego ŁM według wzoru:

$$\hat{\mu}_j = \frac{\sum_{t=1}^T y_t \cdot P(x_t = j | y^{(T)}; \hat{\theta})}{\sum_{t=1}^T P(x_t = j | y^{(T)}; \hat{\theta})} \quad (59)$$

Na koniec trzeba jeszcze zauważyć, że w ogólności nie jesteśmy ograniczeni do używania HMM z przełączaniem pierwszego rzędu, z czego często korzysta się w zastosowaniach zaprezentowanych w następnym podrozdziale. Konstruuje się w takim przypadku łańcuch Markowa wyższego rzędu. W ten właśnie sposób skonstruowane zostaje użyteczne uogólnienie prostych HMM.

3.2 Model analizy cyklu gospodarczego

3.2.1 Wprowadzenie

Przypuśćmy, że wyróżniamy dwa stany gospodarki (niech się nazywają ekspansją i recesją). Badania empiryczne potwierdzają występowanie cykli, co

²⁰Zauważmy, że wzory te są identyczne jak w przypadku przełącznikowych modeli Poissona. Pojawia się jedynie dodatkowy wzór na wariancję, gdyż rozkład normalny jest zadany przez dwa parametry, a Poissona przez jeden.

w konsekwencji musi poprowadzić do stwierdzenia, że cechą immanentną gospodarek jest naprzemienne występowanie recesji i ekspansji. Wiemy również, że mechanizm wywołujący cykle może mieć charakter losowy²¹. Jeżeli więc będziemy budować modele objaśniające kształtowanie się podstawowych wielkości makroekonomicznych takich jak bezrobocie czy produkt krajowy brutto, które opisują w znacznym stopniu aktualny stan gospodarki, będziemy spodziewać się, że zależnie od stanu inne będą parametry wiążące zmienne objaśniające ze zmienną objaśnianą.

W przypadku modelowania koniunktury trzeba wziąć pod uwagę jeszcze jedną ważną jej cechę. Warto wspomnieć, że analitycy nie wiedzą z całą pewnością w jakim stanie znajduje się gospodarka. Wahania koniunkturalne są bowiem wahaniami w ogólnie rozumianej aktywności gospodarczej, nie ma natomiast możliwości bezpośredniego jej skwantyfikowania. Analityk więc staje wobec potrzeby wykorzystania szerokiego wachlarza znanych zmiennych ekonomicznych i różnych metod, tak by w jak najlepszym stopniu oszacować momenty wystąpienia zmiany stanu aktywności gospodarki — jest to postępowanie, które ma swoją genezę jeszcze w klasycznej pracy Burnsa, Mitchella[11]. Jest więc uzasadnione na gruncie teoretycznym by zakładać, że rzeczywiste stany gospodarki są dla nas niewidoczne i próbujemy jedynie oszacować momenty zmiany reżimów i prawdopodobieństwa ich występowania w danych okresach.

Zebranie powyższych wniosków może nasunąć pomysł, żeby wykorzystać w modelowaniu koniunktury coś na kształt ukrytych modeli Markowa. Mamy tam stany układu, mamy proces, który rządzi jego zmianami, zakładamy również, że stany są dla nas niewidoczne. Znakami emitowanymi przez układ mogłyby być na przykład stopa bezrobocia, stopa wzrostu produktu krajowego brutto czy produkcji przemysłowej albo jakiś inny wskaźnik zbiorczy. Spodziewalibyśmy się, że w ukrytym stanie odpowiadającym recesji z dużo mniejszym prawdopodobieństwem układ wygeneruje względnie wysoką stopę wzrostu PKB niż w stanie odpowiadającym ekspansji gospodarczej. I rzeczywiście w 1989 roku J.D. Hamilton[18] zaproponował rodzinę modeli, które stanowią uogólnienie ukrytych modeli Markowa — przełącznikowe

²¹Patrz na przykład: Burda, Wyplosz[10]

modele Markowa (ang. *Markov switching models*, MSM). Model ten zostanie pokrótce przedstawiony w następnym podrozdziale, po czym powrócimy do naszych rozważań dotyczących dwustanowych HMM, proponując w podrozdziale 3.2.3 model wychwytywania zwrotów koniunktury.

3.2.2 Postać modelu koniunktury

Hamilton założył, że istnieją dwa stany $x_t \in \{1, 2\}$, dla których parametry modelu opisującego zachowanie się gospodarki, istotnie się różnią. Jeżeli jako I_t oznaczymy wartość pewnego indykatora aktywności gospodarczej to model, może przybrać poniższą postać:

$$\begin{aligned} I_t - \mu_{x_t} &= \rho(I_{t-1} - \mu_{x_{t-1}}) + \epsilon_t \\ \epsilon_t &\sim N(0, \sigma^2) \\ \mu_1 &> \mu_2 \end{aligned} \tag{60}$$

Założmy, że warunek $\mu_1 > \mu_2$ możemy interpretować tak, że ukryty proces przełącza się pomiędzy stanami charakteryzujące się większym i mniejszym wzrostem oraz $x_t = 2$ charakteryzuje się mniejszą średnią aktywnością gospodarczą.

Hamilton w swoim modelu wykorzystał szereg produktu narodowego brutto dla gospodarki amerykańskiej. Zmienna objaśniana I_t jest stopą wzrostu PNB uzyskaną za pomocą przekształcenia $I_t = \log(\frac{Y_t}{Y_{t-1}})$, gdzie Y_t jest obserwacją pierwotnego szeregu dla okresu t .

Model pierwotnie zaproponowany przez Hamiltona[18] zakładał zarówno więcej opóźnień w autoregresji jak i dłuższą pamięć układu:

$$I_t - \mu_{x_t} = \phi_1(I_{t-1} - \mu_{x_{t-1}}) + \phi_2(I_{t-2} - \mu_{x_{t-2}}) + \phi_3(I_{t-3} - \mu_{x_{t-3}}) + \phi_4(I_{t-4} - \mu_{x_{t-4}}) + \epsilon_t. \tag{61}$$

Po zastosowaniu modelu oraz jego numerycznej maksymalizacji okazało się, że optymalne wartości parametrów były równe:

$$\mu_1 = 1.16, \mu_2 = -0.36, \phi_1 = 0.01, \phi_2 = -0.06, \phi_3 = -0.25, \phi_4 = -0.21, \sigma^2 = 0.59.$$

Macierz przejść między stanami gospodarki przedstawiała się natomiast tak:

$$\begin{bmatrix} 0.90 & 0.1 \\ 0.25 & 0.75 \end{bmatrix}.$$

Oznacza to, że średnia stopa wzrostu PNB w stanie $x_t = 1$ jest równa około 1,16% na kwartał, a w $x_t = 2$ średnia stopa spadku PNB jest równa 0,36%. Zauważmy, że zarówno prawdopodobieństwo pozostania w pierwszym jak i w drugim stanie są stosunkowo wysokie. Oznacza to, że oba stany są stosunkowo trwałe. Średni czas pozostawania gospodarki w pierwszym stanie jest równy $\frac{1}{1-p_{11}} = 10$ kwartałów, drugi stan natomiast trwa średnio $\frac{1}{1-p_{22}} = 4$ kwartały. Potwierdza to powszechnie przyjmowany fakt mówiący o bardzo wyraźnej asymetrii między fazą ekspansji gospodarczej a recesją. Do tego z tych oszacowań wynika, że pełny cykl (na który składają się obie fazy) trwa przeciętnie 14 kwartałów. Co również znajduje swoje potwierdzenie w teorii cykli koniunkturalnych. Zatem wynik modelu Hamiltona doskonale się z tym komponuje, co wskazuje na możliwość interpretowania stanów MSM, jako okresów recesji i ekspansji. Potwierdzają to dodatkowo bardzo zbliżone oszacowania wygładzonych prawdopodobieństw przebywania w stanie drugim, z oficjalnymi datami recesji ogłoszonymi przez NBER²².

Hamilton zaproponował by za recesję uznawać, każdy okres, który charakteryzuje się wygładzonym prawdopodobieństwem w stanie $x_t = 2$ większym niż z góry zadana liczba. Najbardziej naturalne jest przyjmować, że recesja wystąpiła w okresie t dla $P(x_t = 2 | \mathcal{I}_T) > 0.5$. Postępując właśnie w ten sposób Hamilton wykazał, że jego metoda wyznacza szczyty i dna cyklu praktycznie tak samo jak NBER (dane ogłaszane przez NBER przedstawiono w tabeli 4).

3.2.3 Modelowanie zwrotów koniunktury

Powstaje pytanie, czy bazując na podobnie prostym modelu, możemy w jakiś sposób sformalizować metodę wychwytywania zwrotów koniunktury. Tabela 4 zawiera podstawowe charakterystyki cykli dla gospodarki amerykańskiej — w szczególności zawiera okresy recesji w badanym przez nas czasie.

²²National Bureau of Economic Research jest organem odpowiedzialnym między innymi za badanie cyklu koniunkturalnego i oficjalne wyznaczanie zwrotów poszczególnych faz cykli. Komisja datująca cykle składa się z kilkunastu wybitnych ekonomistów i ogłasza swoje wyniki zwykle z wielomiesięcznym opóźnieniem.

Tabela 4: Szczegółowe zestawienie cykli zaproponowane przez NBER w interesującym nas okresie (Źródło: Opracowanie własne na podstawie danych NBER <http://www.nber.org/cycles.html/>).

Szczyt	Dno	Recesja		Ekspansja		Szczyt do	do
		Szczyt do dna	do dna	Dno do dna	Szczyt do		
Lipiec 1953(II)	Maj 1954 (II)	10	45	55	56		
Sierpień 1957(III)	Kwiecień 1958 (II)	8	39	47	49		
Kwiecień 1960(II)	Luty 1961 (I)	10	24	34	32		
Grudzień 1969(IV)	Listopad 1970 (IV)	11	106	117	116		
Listopad 1973(IV)	Marzec 1975 (I)	16	36	52	47		
Styczeń 1980(I)	Lipiec 1980 (III)	6	58	64	74		
Lipiec 1981(III)	Listopad 1982 (IV)	16	12	28	18		
Lipiec 1990(III)	Marzec 1991(I)	8	92	100	108		
Marzec 2001(I)	—	—	120	—	128		

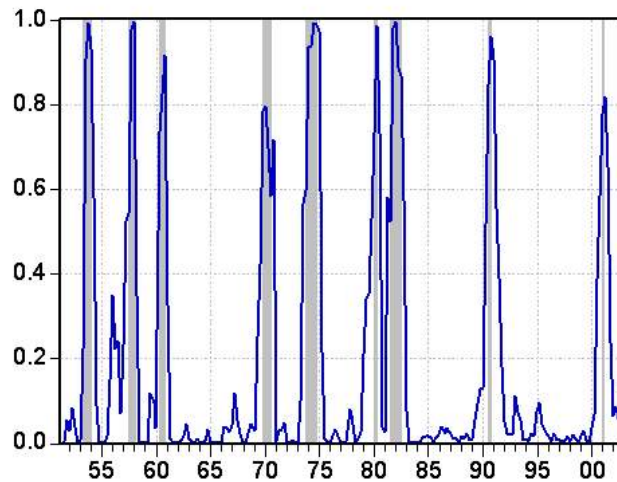
Pojęcie analizowania cykli w czasie rzeczywistym (ang. *Real-Time Business Cycle Dating*) w kontekście omawianych modeli wiąże się z szacowaniem prawdopodobieństwa, że nastąpił zwrot koniunktury (czyli przejście z jednego stanu ŁM do drugiego) na podstawie najświeższych danych. Okazuje się, że bardzo dobre możliwości wychwytywania zwrotów posiada niezwykle prosty model postaci[12]:

$$I_t = \mu_{x_t} + \epsilon_t, \quad (62)$$

gdzie $\epsilon_t \sim N(0, \sigma^2)$. Jest więc to zwykle przełączanie między dwoma różnymi rozkładami normalnymi o tej samej wariancji. Oczywiście podobny model nie ma prawa dobrze działać w przypadku gospodarek takich jak Polska. Wydaje się mało prawdopodobne, by średnie stopy wzrostu w dwóch różnych stanach aktywności gospodarczej, pozostawały stałe. Dojrzałe gospodarki takie jak amerykańska, rzeczywiście cechują się niezwykle stabilną stopą wzrostu[29]. Tymczasem gospodarki rozwijające się, jak polska, powinny wykazywać długookresowy trend spadkowy w stopie wzrostu PKB. W początkowych okresach po wyjściu z zapaści, Polska obserwowała wysokie stopy wzrostu PKB, co wynikało głównie z dynamicznie rosnącą efektywnością gospodarki. Z czasem jednak stopa wzrostu powinna stabilizować się na poziomie charakterystycznym dla krajów rozwiniętych.

Zanim sprawdzimy na ile dobrze model ten jest w stanie wychwytywać zwroty cyklu w czasie rzeczywistym dla Stanów Zjednoczonych, zastanówmy się nad kwestią dużo prostszą, a mianowicie, czy model jest w stanie wychwytywać zwroty na podstawie wszystkich dostępnych danych. Na rysunku 4 przedstawione są wygładzone prawdopodobieństwa przebywania przez

układ w stanie $x_t=2$, a więc w stanie, który odpowiada niższej malejącej aktywności gospodarczej. Model został oszacowany na podstawie danych kwartalnego PKB dla Stanów Zjednoczonych (dane Bureau of Economic Analysis) z okresem początkowym: drugi kwartał 1951 roku i końcowym: pierwszy kwartał 2003 roku. Ciemniejsze obszary na wykresie odpowiadają wyznaczanym przez NBER recesjom.



Rysunek 4: Wygładzone prawdopodobieństwa przebywania w stanie $x_t = 2$ skonfrontowane z okresami recesji datowanymi przez NBER, które zaznaczone są zaciemnionymi obszarami. Ta zaskakująca dobra zbieżność każe wierzyć, że podana metoda rzeczywiście może być alternatywą dla metod NBER (Źródło: opracowanie własne na podstawie danych Bureau of Economic Analysis)

Choć na podstawie analizy powyższych rysunków możemy powiedzieć, że zaproponowany model wykazuje zbieżność ze zwrotów koniunktury datowanymi przez NBER to można to jeszcze doprecyzować za pomocą narzędzi analitycznych. Założymy za Hamiltonem, że okres recesji następuje, gdy stan recesji jest bardziej prawdopodobny niż stan ekspansji gospodarczej. Czyli, gdy:

$$P(x_t = 2 | \mathcal{Y}_t) > 0,5. \quad (63)$$

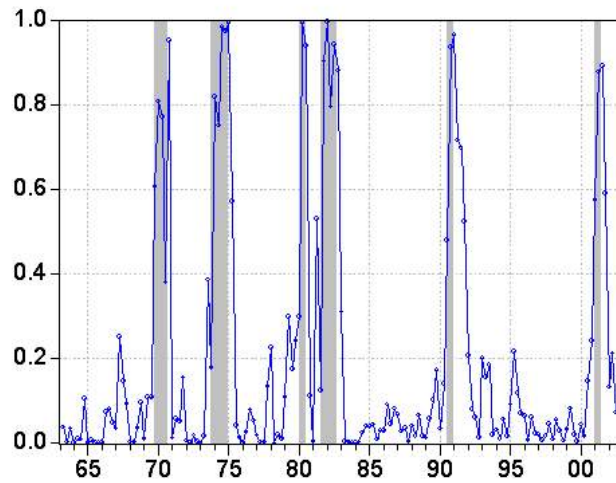
Przy takim intuicyjnym podejściu problem mógłby się pojawić w momencie, gdyby wartości prawdopodobieństw oscylowały w okolicy 0,5, gdyż wtedy każde niewielkie przekroczenie tej wartości byłoby przez model odczytywane jako punkt szczytowy. Model mógłby wtedy generować znacznie więcej

zwrotów koniunktury niż nastąpiło ich w rzeczywistości. Jednak dla danych amerykańskich te prawdopodobieństwa niezwykle rzadko przybierały wartości z przedziału $(0,3; 0,7)$, co oznacza, że z pomocą tego modelu możemy bardzo zdecydowanie podzielić badany okres na podokresy recesji i podokresy ekspansji.

Posługując się modelem 62 do datowania cykli w czasie rzeczywistym najlepiej brać nie wszystkie dostępne dane ale stałą ich liczbę. W ten sposób zabezpieczamy się w pewnym stopniu przed niestabilnością parametrów. M. Chauvet i J. Piger zaproponowali w swoim artykule[12] aby badać ostatnie 40 lat. Wykres przefiltrowanych prawdopodobieństw (uzyskanych według wzoru 53) dla danych między pierwszymi kwartałami 1963 roku i 2003 roku przedstawiono na wykresie 5. W niniejszej pracy jednak model został oszacowany jednorazowo dla badanego okresu. Jest to więc prostsze rozwiązanie niż w przypadku procedury proponowanej przez Chauvet i Piger, gdyż w ten sposób należałoby przeprowadzić 40 oddzielnych estymacji — dla każdego okresu badanym przedziałem byłoby 40 poprzedzających go lat. Wyniki jednak w obu przypadkach nie różnią się znacząco, dlatego zdecydowaliśmy się na prostsze rozwiązanie wierząc, że w dostatecznym stopniu zilustruje ono prezentowaną metodę.

Badanie objęło więc dokładnie 160 kwartalnych obserwacji. Interesuje nas jak wiele obserwacji, dla których prawdopodobieństwo przebywania w stanie recesji było równe ponad 0,5 znalazło się na wykresie 5 poza zacienionym obszarem oraz jak wiele obserwacji dla których to prawdopodobieństwo było równe poniżej 0,5 znalazło się błędnie wewnątrz zacienionego obszaru. Chodzi o to jak często występowały różnice pomiędzy oficjalnymi datowaniami NBER a datowaniami recesji generowanymi przez powyższy dwustanowy ukryty proces Markowa. Okazuje się, że jedynie w przypadku sześciu ze wszystkich badanych okresów stan szacowany przez HMM okazał się inny niż ten datowany przez NBER²³. I to aż trzy z tych sześciu okresów obejmowało okres bezpośrednio po recesji między trzecim kwartałem 1990 roku a pierwszym kwartałem 1991 roku. Szczegółowe zestawienie różnic przedstawia tabela 5.

²³Zakładamy nieomylność NBER, a więc tak zaproponowana metoda badania stanu aktywności gospodarczej myliła się w około 3,75% przypadków.



Rysunek 5: Wyznaczanie zwrotów koniunktury w czasie rzeczywistym. Tutaj prawdopodobieństwo w każdym okresie t były szacowane tylko na podstawie danych do tego okresu. Mimo to zbieżność z danymi NBER jest uderzająca. (Źródło: opracowanie własne na podstawie danych Bureau of Economic Analysis)

Poza drugim i trzecim kwartałem 1991 roku błędne decyzje HMM mogą wynikać z tego, że prawdopodobieństwa przebywania w drugim stanie są stosunkowo blisko połowy. Jak już wcześniej zostało wspomniane, jest to sytuacja niepożądana z punktu widzenia skuteczności modelu. Jedynie dla tych dwóch kwartałów 1991 roku można powiedzieć, że wyniki estymacji są zdecydowanie błędne. Kolejnym ważnym wnioskiem z tabeli 5 jest, że HMM w przypadku błędów, znacznie bardziej preferuje fazę ekspansji. Tylko raz w przypadku recesji model stwierdził, że jest faza ekspansji (na podstawie kryterium $P(x_t = 2 | \mathcal{Y}_t) < 0,5$), a w pozostałych przypadkach było na odwrót.

Tabela 5: Zestawienie błędów między stanami gospodarki datowanymi przez NBER a tymi generowanymi przez MSM dla wychwytywania zwrotów koniunktury.

okres	$P(x_t = 2 \mathcal{Y}_t)$	NBER
1970:3	0,38	recesja
1975:2	0,573	ekspansja
1981:2	0,532	ekspansja
1991:2	0,717	ekspansja
1991:3	0,7	ekspansja
1991:4	0,525	ekspansja

Nas jednak od początku najbardziej interesowała zdolność dwustanowego ukrytego procesu Markowa z normalnymi rozkładami emisji do wychwytywania zwrotów koniunktury. Okazuje się, że tylko w przypadku drugiego kwartału 1981 roku mielibyśmy do czynienia z takim „fałszywym” zwrotem, który o dwa kwartały wyprzedził prawdziwy początek recesji występującej pomiędzy trzecim kwartałem 1981 roku a czwartym kwartałem 1982 roku.

Wydaje się, że HMM mogą się stać bardzo cennym narzędziem w rękach analityków gospodarczych. Ich szybkie upowszechnienie się wśród ekonometryków jest najlepszym dowodem na ich skuteczność. Dlatego też ich badanie w przypadku polskiej gospodarki może się okazać bardzo pożyteczne.

4 Zakończenie

W pracy tej osiągnęliśmy dwa ważne cele. Opisaliśmy mianowicie w skrócie ogólną teorię ukrytych modeli Markowa. Prezentując ponadto ogólne metody estymacyjne. Po drugie przedstawiliśmy niektóre szczególnie ważne podklasy ukrytych modeli Markowa, których wykorzystanie w modelowaniu stało się w ostatnich latach szczególnie popularne. Przedstawione zastosowania szybko zdobyły popularność w swoich dziedzinach. Dzięki opisowi szerokiego wachlarza zastosowań, zdołaliśmy pokazać elastyczność ukrytych modeli Markowa w analizie najprzeróżniejszych zjawisk zachodzących w przyrodzie.

Skupiliśmy się w tej pracy na dwustanowych HMM. Modele tego typu szczególnie często pojawiają się w najróżniejszych zastosowaniach. Jak staraliśmy się pokazać, mają one często głęboką interpretację na gruncie teoretycznym, zawsze jednak dobrze sprawdzały się na gruncie empirycznym. Ponadto dwustanowe ŁM można łatwo badać pod kontem ich najprzeróżniejszych własności granicznych i dla skończonej ilości obserwacji skończenie. Jest to bardzo ważne, gdyż HMM dziedziczy własności takie jak stacjonarność i ergodyczność właśnie po swoim ukrytym procesie.

Literatura

- [1] P.S. Albert. A two-state markov mixture model for a time series of epileptic seizure counts. *Biometrics*, 47:1371–1381, 1991.

- [2] A.P.Dempster, N.M. Laird, D.B. Rubin. Maximum-likelihood from incomplete data via the em algorithm. *J. Royal Statist. Soc. Ser. B.*, 39, 1977.
- [3] A. Azzalini, A.W. Bowman. A look at some data on the old faithful geyser. *Appl. Statist.*, 39:357–365, 1990.
- [4] P. Baldi, Y. Chauvin, T. Hunkapiller, M.A. McClure. Hidden markov models of biological primary sequence information. *Proc. Nat. Acad. Sci. U.S.A.*, 91:1059–1063, 1994.
- [5] L.E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. O. Shisha, redaktor, *Proc. Third Symposium on Inequalities*, strony 1–8, New York, 1972. Academic Press.
- [6] L.E. Baum, J.A. Eagon. An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model of ecology. *Bull. Amer. Math. Soc.*, 73:360–363, 1967.
- [7] L.E. Baum, T. Petrie. Statistical inference for probabilistic functions of finite state markov chains. *Ann. Math. Statist.*, 37:1554–1563, 1966.
- [8] L.E. Baum, G.R. Sell. Growth transformations for functions on manifolds. *Pacific J. Math*, 27:211–227, 1968.
- [9] L.E. Baum, T.Petrie, G. Soules, N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *Ann. Math. Statist.*, 41:164–171, 1970.
- [10] M. Burda, C. Wyplosz. *Makroekonomia*. PWE, Warszawa, 2000.
- [11] A.F. Burns, W.C. Mitchell. *Measuring business cycles*. NBER, New York, 1946.
- [12] M. Chauvet, J. M. Piger. Identifying business cycle turning points in real time. *The Federal Reserve Bank of St. Louis.*, march/april:47–62, 2003.

- [13] G.A. Churchill. Stochastic models for heterogeneous dna sequences. *Bull. Math. Biol.*, 51:79–94, 1989.
- [14] G.A. Churchill. Hidden markov chains and the analysis of genome structure. *Computers Chem.*, 16:107–115, 1992.
- [15] M. Crowder, Mark Davis, Giacomo Giampieri. A hidden markov model of default interaction. <http://www.ma.ic.ac.uk/~mdavis/docs/CrowderDavisGiampieri.pdf>.
- [16] R.J. Elliot, L. Aggoun, J.B. Moore. *Hidden Markov Models: Estimation and control*. Springer, New York, 1995.
- [17] Y. Ephraim, N. Merhav. Hidden markov processes. *IEEE Transactions on Information Theory*, 48(6):1518–1569, lipiec 2002.
- [18] J.D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384, 1989.
- [19] J.D. Hamilton. Analysis of time series subject to changes in regime. *Journal of econometrics*, 45:39–70, 1990.
- [20] J.D. Hamilton. *Time series analysis*. Princeton University Press, Princeton, New Jersey, 1994.
- [21] A. Hopkins, P. Davies, C. Dobson. Mathematical models of patterns of seizures: Their use in the evaluation of drugs. *Archives of Neurology*, 42:463–467, 1985.
- [22] J. Jakubowski, R. Sztencel. *Wstęp do teorii prawdopodobieństwa*. SCRIPT, Warszawa, wydanie ii, 2001.
- [23] C.-J. Kim. Dynamic linear models with markov switching. *Journal of Econometrics*, 60:1–22, 1994.
- [24] N.D. Le, B.G. Leroux, M.L. Puterman. Reading reaction: Exact likelihood evaluation in a markov mixture model for time series of seizure counts. *Biometrics*, 48:317–323, 1992.

- [25] B.G. Leroux, M.L. Puterman. Maximum-penalized-likelihood estimation for independent and markov-dependent mixture models. *Biometrics*, 48:545–558, 1992.
- [26] S.E. Levison, L.R. Rabiner, M.M. Sondhi. An introduction to the application of the theory of probabilistic functions of markov process to automatic speech recognition. *Bell System Tech. J.*, 62:1035–1074, 1983.
- [27] I.L. MacDonald, L.B. Lerer. A time series analysis of trends in firearm-related homicide and suicide. *Int. J. Epidemiol.*, 23:66–72, 1994.
- [28] P.A. Schrodtt. Forecasting conflict in the balkans using hidden markov models. *Paper presented at the American Political Science Association meetings Washington, DC*, 1, 2000.
- [29] J.H. Stock, M.W. Watson. Understanding changes in international business cycle dynamics. maj 2003.
- [30] E.A. Thompson. Optimal sampling for pedigree analysis: parameter estimation and genotypic uncertainty. *Theor. Pop. Biol.*, 24:39–58, 1983.
- [31] A. D. Wentzell. *Wykłady z teorii procesów stochastycznych*. PWN, Warszawa, 1998.
- [32] W. Zucchini, P. Guttorp. A hidden markov model for spacetime precipitation. *Water Resour. Res.*, 27:1917–1923, 1991.