

Politechnika Wrocławska,
Instytut Matematyki

**Optymalność w semi-markowskich procesach decyzyjnych
i grach stochastycznych**

Anna Jaśkiewicz

Rozprawa doktorska napisana pod kierunkiem dr hab. Andrzeja Nowaka,
Profesora Uniwersytetu Zielonogórskiego

Wrocław, 2003



Politechnika Wrocławska,
Instytut Matematyki

**Optymalność w semi-markowskich procesach decyzyjnych
i grach stochastycznych**

Anna Jaśkiewicz

Rozprawa doktorska napisana pod kierunkiem dr hab. Andrzeja Nowaka,
Profesora Uniwersytetu Zielonogórskiego

Wrocław, 2003

9630

Biblioteka Główna i OINT
Politechniki Wrocławskiej



001758541



316064 L/1

30

2003/0555/D

*Serdecznie dziękuję Promotorowi
dr hab. Andrzejowi Nowakowi
za inspirację oraz opiekę naukową
w czasie pisania niniejszej pracy.*

**Opis podstawowych tez zawartych w rozprawie
doktorskiej opartej na artykułach:**

[1] A. Jaśkiewicz „An approximation approach to ergodic semi-Markov control processes”, Math. Methods of Oper. Res., vol. 54 (2001), str. 1-19.

[2] A. Jaśkiewicz „Zero-sum semi-Markov games”, SIAM J. Control Optim., vol. 41 (2002), str. 723-739.

[3] A. Jaśkiewicz „On the equivalence of two expected average cost criteria for semi-Markov control processes”, praca wysłana do druku.

1. Wstęp.

Teoria markowskich procesów decyzyjnych przyciągnęła uwagę wielu autorów z dwóch prostych powodów: praktycznego i intelektualnego. Po pierwsze dziedzina ta dostarcza narzędzia do rozwiązywania istotnych problemów pojawiających się w codziennym życiu: ekonomii i finansach [5, 6, 10, 43] (analiza portfelowa, teoria wzrostu ekonomicznego), naukach inżynierskich [1, 16, 17, 18] (teoria kolejek, telekomunikacja), problemach związanych z zarządzaniem [4, 16, 17, 18, 27, 31, 42] (alokacja zasobów, zarządzanie zapasami), jak i ubezpieczeniach. Okazuje się, że decyzje podejmowane przez człowieka, komputery mają zwykle wpływ na: ponoszone przez nas koszty bądź otrzymywane wypłaty, oszczędność czasu, innych zasobów oraz na przyszłość poprzez pewną (często stochastyczną) dynamikę. W wielu sytuacjach decyzje przynoszące natychmiastowy zysk nie muszą być dobre w świetle przyszłych wydarzeń. Markowskie procesy decyzyjne modelują te problemy i dostarczają wyniki na temat optymalnych strategii i ich istnienia oraz metody do ich wyliczenia. Po drugie dziedzina ta obfituje w szereg ciekawych zagadnień matematycznych i wykorzystuje wiele matematycznych metod.

Praca Shapley [47] o grze stochastycznej została uznana za początek rozwoju markowskich procesów decyzyjnych. W istocie, Shapley w 1953 roku zdefiniował grę stochastyczną o skończonym zbiorze stanów, która ma stałe i mniejsze od jedności prawdopodobieństwo kontynuacji. Taka gra jest rozszerzoną wersją markowskiego procesu decyzyjnego. Przez pierwsze trzydzieści lat, do wczesnych lat 80-tych, wiele badań było poświęconych równaniu optymalności i metodom jego rozwiązania: poprzez modele dyskontowane („vanishing discount approach”), iteracji strategii („policy iteration”), iteracji wartości („value iteration”). Ten ostatni algorytm i jego różne wersje znane są również pod innymi nazwami np. sukcesywna aproksymacja, indukcja wsteczna, metoda programowania dynamicznego. Koncepcja programowania dynamicznego została opisana w książce Bellmana [2], ale ta technika była już używana w latach 1940-50, a nawet wcześniej.

Uogólnieniem procesów markowskich są modele semi-markowskie, w których czas pobytu procesu w danym stanie zależy od tego stanu i podjętych w nim decyzji. Zostały one przedstawione przez Jewella [26] oraz Howarda [21] i często w literaturze nazywane są markowskimi procesami odnowy („Markov renewal programmes”). Zachowanie się takich procesów jest regulowane przez decyzje decydenta, podejmowane zgodnie z pewnymi regułami. Wybory dokonywane są w momentach przejścia od stanu do kolejnego stanu, wpływając na ewolucję procesu stochastycznego, z którym związany jest ciąg („strumień”) wypłat stanowiących podstawę do zdefiniowania wypłaty w modelu z nieskończonym horyzontem czasowym. Ze względu na liczne zastosowania oraz perspektywę uogólnienia rezultatów otrzymanych wcześniej dla modeli markowskich, semi-markowskie procesy decyzyjne (a także gry o sumie zerowej) zwróciły uwagę wielu autorów i badaczy. Nie będąc w stanie opisać całej szerokiej literatury z tej dziedziny, wymienimy tylko najbardziej reprezentatywne oraz związane z naszymi badaniami i studiami pozycje: [4, 12, 13, 14, 15, 20, 22, 23, 27, 28, 29, 34, 39, 40, 41, 44, 46, 49, 50, 51]¹. Jednym z bardzo

¹Dalsze szczegółowe komentarze i odnośniki literaturowe można znaleźć w załączonych pracach.

pomocnych narzędzi w studiowaniu tych procesów stała się transformacja przedstawiona przez Schweitzera [45] w końcu lat 70-tych. Pozwoliła ona na przekształcenie modelu semi-markowskiego w równoważny mu model markowski. Dzięki temu udało się rozwiązać równanie optymalności z kryterium tzw. średniej wypłaty dla procesów z przeliczalną przestrzenią stanów [13, 14]. Wcześniej takie próby podejmował Ross [41], a później tą ideę rozwinęli Sennott [46] oraz Vega-Amaya [50]. Niestety bezpośrednie podejście przez modele dyskontowane (bez transformacji Schweitzera) jest bardzo skomplikowane, a założenia przedstawione w pracach [41, 46, 50] są niezwykle trudne do zweryfikowania nawet w przypadku ergodycznym. Z drugiej strony zaczęto studiowanie równoważności dwóch kryteriów optymalności: średniej wypłaty i średniej wypłaty na jednostkę czasu (definicje podane są w rozdziale 2). To drugie kryterium jest bardziej naturalne i uwzględnia ciągłą naturę procesów semi-markowskich. Jednak praca z nim jest trudniejsza i wymaga dodatkowych metod i narzędzi m.in. faktów z teorii odnowy. Bazując na podstawowych twierdzeniach z tej dziedziny, Ross [41] po raz pierwszy pokazał, że dwa wyżej wspomniane kryteria pokrywają się dla polityk stacjonarnych w przypadku modeli z przeliczalnym zbiorem stanów i silnie ergodycznymi prawdopodobieństwami przejścia. Rezultat Rossa [41] został znacznie uogólniony w pracy Schälä [44] w 1992 roku, w której osłabiono założenia ergodyczności i wykazano, że w obu podejściach możemy mówić o tym samym równaniu optymalności Bellmana. Oprócz teorii odnowy Schälä zastosował w swojej analizie twierdzenie Dooba o opcjonalnym stopowaniu. Problemy związane z badaniem semi-markowskich procesów decyzyjnych znacznie się upraszczają, gdy zbiór stanów jest skończony. Jak opisano w książce Mine i Osaki [34], w takim przypadku można wykorzystywać jedynie narzędzia z analizy matematycznej oraz algebry macierzy.

Równanie optymalności (z kryterium średniej wypłaty) dla borelowskiej przestrzeni stanów przy ogólnych założeniach ergodyczności i stabilności stochastycznej zostało rozwiązane całkiem niedawno [20, 22, 51]. W pracach [20, 51] użyto transformacji Schweitzera a następnie algorytmu iteracji polityki (algorytmu Howarda). Natomiast w [22] przedstawiono zupełnie nową metodę polegającą na zaburzeniu funkcji prawdopodobieństwa przejścia. Następnie tą ideę zastosowano dla gier semi-markowskich [23]. W ten sposób otrzymano twierdzenia o istnieniu wartości gry, charakteryzując strategii optymalnych poprzez uogólnione równanie Bellmana, a także metodę przybliżania wartości gry poprzez pewne procedury iteracyjne, związane z wprowadzonymi prostymi grami pomocniczymi. Jedyna praca o grach semi-markowskich o sumie zerowej [28] rozważa modele z przeliczalnym zbiorem stanów i opiera się na metodzie Schweitzera. Ponadto zawiera ona znacznie silniejsze założenia ergodyczności i dotyczy tylko jednego z możliwych kryteriów optymalności.

Istotnym uzupełnieniem tez zawartych w pracy [23] jest równoważność kryteriów optymalności wykazana w [24]. Podobnie jak u Schälä opiera się ona na zastosowaniu elementów teorii odnowy i twierdzenia Dooba o opcjonalnym stopowaniu. Jednakże trochę inny rodzaj założeń ergodyczności oraz fakt, że przestrzeń stanów może być nieprzeliczalna zmusiły nas do pokonania pewnych nowych trudności związanych z teorią procesów markowskich z ogólną przestrzenią stanów. Praca [24] pozwala spojrzeć na wyniki z prac wcześniejszych [22, 23] z nowej perspektywy. Otrzymane tam rezultaty zachowują swoje

znaczenie również w przypadku kryterium średniej wypłaty na jednostkę czasu.

Rozprawa jest skomponowana w następujący sposób. W pierwszej części napisanej w języku polskim przedstawiamy opis modelu, główne tezy rozprawy wraz z komentarzami i dyskusją założeń. Druga część zawiera artykuły stanowiące główną część rozprawy.

2. Opis modelu.

Semi-markowska gra stochastyczna jest opisana przy pomocy następujących obiektów:

- (i) X jest borelowską² przestrzenią stanów;
- (ii) A i B są borelowskimi przestrzeniami decyzji graczy;
- (iii) K_A i K_B są niepustymi borelowskimi podzbiorami $X \times A$ i $X \times B$, odpowiednio.

Dla każdego $x \in X$ definiujemy

$$A(x) := \{a \in A : (x, a) \in K_A\}$$

jako zbiór możliwych decyzji gracza 1 w stanie $x \in X$. Analogicznie definiujemy zbiór $B(x)$. Oznaczmy:

$$K := \{(x, a, b) : x \in X, a \in A(x), b \in B(x)\}.$$

Łatwo wykazać, że K jest borelowskim podzbiorem w $X \times A \times B$;

(iv) q jest borelowskim prawdopodobieństwem przejścia ze zbioru K do X . Jeśli x jest stanem, w którym obecnie znajduje się gra semi-markowska zaś $a \in A(x)$ i $b \in B(x)$ są decyzjami graczy w tym stanie, to $q(\cdot | x, a, b)$ jest prawdopodobieństwem przejścia do następnego stanu;

(v) $Q(\cdot | x_n, a_n, b_n)$ jest dystrybuantą zmiennych losowych $T_{n+1} - T_n$, gdzie T_k oznacza moment podjęcia decyzji $a_k \in A(x_k)$, $b_k \in B(x_k)$, który następuje natychmiast po przejściu do stanu x_k ($n = 0, 1, \dots$ oraz $T_0 := 0$);

(vi) $r_1(x_n, a_n, b_n)$ jest wypłatą dla gracza 1 (kosztem dla gracza 2), związaną z przejściem do stanu x_n oraz podjętymi decyzjami $a_n \in A(x_n)$ i $b_n \in B(x_n)$;

(vii) $r_2(x_n, a_n, b_n)$ jest współczynnikiem zmiennej wypłaty dla gracza 1 (kosztu dla gracza 2).

Ewolucja gry semi-markowskiej o sumie zerowej przebiega w następujący sposób. W czasie $t = 0$ gracze obserwują stan $x_0 \in X$ i podejmują decyzje $a_0 \in A(x_0)$ i $b_0 \in B(x_0)$. Gra pozostaje w stanie x_0 przez pewien losowy czas T_1 , którego rozkład zależy od x_0 , a_0 i b_0 . Gracz 1 otrzymuje natychmiastową wypłatę $r_1(x_0, a_0, b_0)$ od gracza 2 (gracz 2 ma wypłatę $-r_1(x_0, a_0, b_0)$). Natomiast wypłata postaci $r_2(x_0, a_0, b_0)$ jest płaconą przez gracza 2 graczowi 1 aż do następnego skoku (patrz (3) poniżej). W momencie T_1 gra przechodzi do stanu x_1 zgodnie z rozkładem $q(\cdot | x_0, a_0, b_0)$. Sytuacja powtarza się nieskończenie wiele razy, dając trajektorię $(x_0, a_0, b_0, t_1, x_1, \dots)$ pewnego procesu stochastycznego. Aby zapewnić istnienie wartości gry dopuszcza się randomizowanie wyborów graczy na poszczególnych etapach gry. Formalny opis modelu i strategii przedstawiamy w dalszej

²W tym opracowaniu przez przestrzeń borelowską rozumie się borelowski podzbiór przestrzeni polskiej.

części.

Niech $\mathbf{R}$ ($\mathbf{R}_+$) oznacza zbiór liczb rzeczywistych (nieujemnych liczb rzeczywistych). Dla dowolnej przestrzeni borelowskiej D , przez $\mathcal{B}(D)$ oznaczamy σ -algebrę borelowskich podzbiorów zbioru D . Symbol $P(D)$ jest używany do oznaczenia przestrzeni wszystkich miar probabilistycznych na $\mathcal{B}(D)$. Zakłada się, że przestrzeń $P(D)$ jest wyposażona w słabą topologię i σ -algebrę $\mathcal{B}(P(D))$.

Niech H_n oznacza przestrzeń *dopuszczalnych historii* aż do n -tego kroku:

$$H_n = (K \times \mathbf{R}_+)^n \times X \quad \text{gdzie} \quad H_0 = X.$$

Element $h_n \in H_n$ jest nazywany *częściową historią* gry:

$$h_n = (x_0, a_0, b_0, t_1, \dots, x_{n-1}, a_{n-1}, b_{n-1}, t_n, x_n).$$

Zakłada się, że zbiór H_n jest wyposażony w σ -algebrę produktową.

Strategią albo *polityką* gracza 1 nazywamy ciąg $\pi = \{\pi_n\}$, w którym każdy wyraz π_n jest warunkowym prawdopodobieństwem $\pi_n(\cdot|h_n)$ na zbiorze decyzji $A(x_n)$, gdy znana jest cała historia h_n :

$$\pi_n(A(x_n)|h_n) = 1 \quad \forall h_n \in H_n, \quad n = 0, 1, \dots$$

Przez Π oznaczamy *zbiór wszystkich strategii* gracza 1. Niech F będzie zbiorem wszystkich borelowskich odwzorowań $f : X \mapsto P(A)$ takich, że $f(x) \in P(A(x))$ dla każdego $x \in X$. Jeśli zbiór $A(x)$ jest zwarty, to F jest zbiorem niepustym (Theorem 1 w [8]). Ciąg $\pi = \{\pi_n\}$ jest stacjonarną polityką gracza 1, jeśli istnieje funkcja $f \in F$ taka, że $\pi_n(\cdot|h_n) = f(\cdot|x_n)$ dla wszystkich $h_n \in H_n$. Zatem dowolną stacjonarną strategię $\pi = (f, f, \dots)$ możemy utożsamiać z $f \in F$. Podobnie definiujemy strategię $\gamma = \{\gamma_n\}$ gracza 2. Symbolem Γ (G) oznaczamy zbiór wszystkich strategii (strategii stacjonarnych) gracza 2.

Niech $\Omega := (X \times A \times B \times \mathbf{R}_+)^{\infty}$ i niech $(\Omega, \mathcal{F})$ będzie przestrzenią mierzalną, gdzie $\mathcal{F}$ jest σ -algebrą produktową. Z twierdzenia Ionescu Tulcea (Proposition V.1.1 w [35] lub [3, 17]) wynika, iż dla stanu początkowego $x \in X$ i dowolnych strategii $\pi \in \Pi$, $\gamma \in \Gamma$ istnieje jedyna miara probabilistyczna $P_x^{\pi\gamma}$ na $\mathcal{F}$ taka, że dla wszystkich $A' \in \mathcal{B}(A)$, $B' \in \mathcal{B}(B)$, $X' \in \mathcal{B}(X)$ i $h_n = (x_0, a_0, b_0, t_1, \dots, x_{n-1}, a_{n-1}, b_{n-1}, t_n, x_n)$ w H_n , $n = 0, 1, \dots$,

$$P_x^{\pi\gamma}(x_0 = x) = 1$$

$$P_x^{\pi\gamma}(a_n \in A'|h_n) = \pi_n(A'|h_n)$$

$$P_x^{\pi\gamma}(b_n \in B'|h_n) = \gamma_n(B'|h_n)$$

$$P_x^{\pi\gamma}(x_{n+1} \in X'|h_n, a_n, b_n, t_{n+1}) = q(X'|x_n, a_n, b_n)$$

i

$$P_x^{\pi\gamma}(T_{n+1} - T_n \leq t|h_n, a_n, b_n) = Q(t|x_n, a_n).$$

Przez $E_x^{\pi\gamma}$ oznaczamy operator wartości oczekiwanej względem miary probabilistycznej $P_x^{\pi\gamma}$.

Niech $\tau(x, a, b)$ będzie *średnim czasem przebywania* procesu (gry) w stanie x przy podjętych decyzjach $a \in A(x)$, $b \in B(x)$, czyli

$$\tau(x, a, b) := \int_0^\infty tQ(dt|x, a, b).$$

Ustalmy $\pi \in \Pi$, $\gamma \in \Gamma$, $x \in X$ oraz $t \geq 0$ i zdefiniujmy zmienną losową $N(t)$ w następujący sposób

$$N(t) := \max\{n \geq 0 : T_n \leq t\}.$$

Jeśli $t \in \mathbf{R}_+$, to $N(t)$ jest procesem stochastycznym zwanym procesem liczącym („counting process”). Poniżej podamy założenia, które implikują, że $N(t) < \infty$ $P_x^{\pi\gamma}$ -p.w. dla każdego $t \geq 0$ (Remark 2 w [24]).

Dla dowolnych strategii $\pi \in \Pi$ i $\gamma \in \Gamma$ oraz stanu początkowego $x \in X$ definiujemy *oczekiwaną średnią wypłatę gracza 1* jako

$$J(x, \pi, \gamma) := \liminf_{n \rightarrow \infty} \frac{E_x^{\pi\gamma}(\sum_{k=0}^{n-1} [r_1(x_k, a_k, b_k) + (T_{k+1} - T_k)r_2(x_k, a_k, b_k)])}{E_x^{\pi\gamma}(T_n)}$$

oraz *oczekiwaną średnią wypłatę na jednostkę czasu gracza 1* jako

$$j(x, \pi, \gamma) := \liminf_{n \rightarrow \infty} \frac{E_x^{\pi\gamma}(\sum_{k=0}^{N(t)} [r_1(x_k, a_k, b_k) + (T_{k+1} - T_k)r_2(x_k, a_k, b_k)])}{t}.$$

Stosując własności operatora wartości oczekiwanej, wyrażenia $J(x, \pi, \gamma)$ i $j(x, \pi, \gamma)$ możemy przepisać w postaci

$$J(x, \pi, \gamma) = \liminf_{n \rightarrow \infty} \frac{E_x^{\pi\gamma}[\sum_{k=0}^{n-1} r(x_k, a_k, b_k)]}{E_x^{\pi\gamma}[\sum_{k=0}^{n-1} \tau(x_k, a_k, b_k)]} \quad (1)$$

oraz

$$j(x, \pi, \gamma) = \liminf_{n \rightarrow \infty} \frac{E_x^{\pi\gamma}[\sum_{k=0}^{N(t)} r(x_k, a_k, b_k)]}{t}, \quad (2)$$

gdzie

$$r(x, a, b) := r_1(x, a, b) + \tau(x, a, b)r_2(x, a, b) \quad (3)$$

dla każdego $(x, a, b) \in K$. Założenia podane w rozdziale 3 zagwarantują, że $J(x, \pi, \gamma)$ i $j(x, \pi, \gamma)$ będą dobrze określone.

Dla dowolnego stanu początkowego $x \in X$, połóżmy

$$L^*(x) := \sup_{\pi \in \Pi} \inf_{\gamma \in \Gamma} J(x, \pi, \gamma) \quad \text{ i } \quad U^*(x) := \inf_{\gamma \in \Gamma} \sup_{\pi \in \Pi} J(x, \pi, \gamma).$$

Wówczas $L^*(x) \leq U^*(x)$, a $L^*(U^*)$ jest nazywana dolną (górną) wartością gry ze średnią wypłatą zdefiniowaną w (1). Jeśli $L^*(x) = U^*(x) =: V^*(x)$ dla wszystkich $x \in X$, to mówimy że gra ma wartość V^* . Strategia $\pi^* \in \Pi$ jest *optymalna dla gracza 1* w tej grze, jeśli

$$\inf_{\gamma \in \Gamma} J(x, \pi^*, \gamma) = V^*(x)$$

dla każdego $x \in X$. Analogicznie, strategia $\gamma^* \in \Gamma$ jest *optymalna dla gracza 2*, gdy

$$\sup_{\pi \in \Pi} J(x, \pi, \gamma^*) = V^*(x)$$

dla każdego $x \in X$. W podobny sposób możemy zdefiniować wartość dolną i górną gry oraz optymalne strategie graczy z kryterium średniej wypłaty na jednostkę czasu, danym wzorem (2).

Niech $x \in X$, $\nu \in P(A(x))$ i $\rho \in P(B(x))$. Dla dowolnej funkcji borelowskiej $u : X \times A \times B \mapsto \mathbf{R}$ kładziemy

$$u(x, \nu, \rho) := \int_{B(x)} \int_{A(x)} u(x, a, b) \nu(da) \rho(db),$$

pod warunkiem, że powyższa całka istnieje. Dodatkowo oznaczmy

$$u(x, f, g) := u(x, f(x), g(x))$$

dla $f \in F$ i $g \in G$.

3. Założenia.

B Podstawowe założenia:

- (i) dla każdego $x \in X$, zbiory decyzji $A(x)$, $B(x)$ są niepuste i zwarte;
- (ii) funkcja $r : K \mapsto \mathbf{R}$ jest borelowska oraz dla $(x, a, b) \in K$, funkcja $r(x, \cdot, b)$ jest górnie półciągła na $A(x)$ i $r(x, a, \cdot)$ jest dolnie półciągła na $B(x)$;
- (iii) dla $(x, a, b) \in K$, funkcje $\tau(x, \cdot, b)$, $\tau(x, a, \cdot)$ są ciągłe na $A(x)$ i $B(x)$, odpowiednio oraz istnieją stałe $m^* > 0$ i $M^* > 0$ takie, że

$$m^* \leq \tau(x, a, b) \leq M^*;$$

- (iv) dla każdego $(x, a, b) \in K$ i każdego zbioru $D \in \mathcal{B}(X)$, funkcje $q(D|x, \cdot, b)$ i $q(D|x, a, \cdot)$ są ciągłe na $A(x)$ oraz $B(x)$, odpowiednio;
- (v) istnieje stała $L > 0$ i funkcja borelowska $V : X \mapsto [1, \infty)$ taka, że $|r(x, a, b)| \leq LV(x)$ dla każdego $(x, a, b) \in K$;
- (vi) dla każdego $x \in X$, funkcje

$$\int_X V(y) q(dy|x, \cdot, b), \quad \int_X V(y) q(dy|x, a, \cdot)$$

są ciągłe na $A(x)$ oraz $B(x)$, odpowiednio.

GE Założenia geometrycznej ergodyczności:

- (i) istnieje borelowski zbiór $C \subset X$ taki, że dla pewnych stałych $\lambda \in (0, 1)$ i $\eta > 0$, mamy

$$\int_X V(y) q(dy|x, a, b) \leq \lambda V(x) + \eta 1_C(x)$$

dla każdego $(x, a, b) \in K$. 1_C oznacza funkcję charakterystyczną zbioru C , a V jest funkcją z założenia **(B, v)**;

(ii) funkcja V jest ograniczona na zbiorze C , tzn.

$$v_C := \sup_{x \in C} V(x) < \infty;$$

(iii) istnieje stała $\delta \in (0,1)$ i miara probabilistyczna μ na zbiorze borelowskim C , dla których zachodzi

$$q(D|x, a, b) \geq \delta \mu(D)$$

dla każdego zbioru borelowskiego $D \subset C$, $x \in C$, $a \in A(x)$ i $b \in B(x)$;

Uwaga 1: Założenie **(GE, iii)** w teorii łańcuchów Markowa oznacza, że proces generowany przez dowolne stacjonarne polityki graczy oraz prawdopodobieństwo przejścia q (zwany również zanurzonym procesem Markowa - „embedded Markov chain”) jest φ -nieredukowalny oraz nieokresowy. Miarę nieredukowalności φ można zdefiniować następująco:

$$\varphi(D) := \delta \mu(D \cap C) \quad \text{dla} \quad D \in \mathcal{B}(X).$$

Innymi słowy, jeśli tylko $\varphi(D) > 0$, to prawdopodobieństwo dotarcia do zbioru D jest zawsze dodatnie, niezależnie od tego z jakiego stanu proces startuje. Ponadto, istnieje maksymalna miara nieredukowalności ψ , dla której $\varphi \ll \psi$ (Theorem 4.2.2 w [32]). Zbiór C nazywany jest małym zbiorem („small set” [32, 33]).

Funkcja V z **(GE, i,ii)** z dokładnością do stałej multiplikatywnej ogranicza oczekiwany czas dojścia do zbioru C (Theorem 14.2.2 w [32]). Ten fakt wraz z założeniem **(GE, iii)** pociąga za sobą, iż

(a) proces markowski generowany przez $f \in F$, $g \in G$ oraz q posiada jedyną miarę niezmienniczą π_{fg} („positive chain”) taką, że

$$\int_X V(x) \pi_{fg}(dx) < \infty;$$

(b) jeśli tylko $\psi(D) > 0$, to prawdopodobieństwo dojścia do zbioru D jest równe 1, niezależnie od stanu początkowego $x \in X$ (Theorem 9.7.1 w [32]). Ta własność łańcucha Markowa jest nazywana w literaturze rekurencyjnością w sensie Harrisa („Harris recurrence”).

Jeśli małym zbiorem jest cała przestrzeń, to założenie **(GE, i)** zachodzi dla funkcji $V(\cdot) \equiv 1$ oraz dostatecznie dużego η . Jednakże dopuszczenie nieograniczonych funkcji V pozwala na objęcie w miarę szerokiej klasy różnych ciekawych przykładów np. modeli kolejkowych, sieciowych. ♣

Niech ν_1, ν_2 będą dowolnymi miarami probabilistycznymi na $\mathcal{B}(X)$. Wówczas V -normy totalnego wahania dla $\nu_1 - \nu_2$ i ν_1 są dane jako

$$\|\nu_1 - \nu_2\|_V := \sup_{|u| \leq V} \left| \int_X u(x) \nu_1(dx) - \int_X u(x) \nu_2(dx) \right|$$

oraz

$$\|\nu_1\|_V := \int_X V(x)\nu_1(dx).$$

Dla dowolnej borelowskiej funkcji $c : X \mapsto \mathbf{R}$ definiujemy V -normę jako

$$\|c\|_V := \sup_{x \in X} \frac{|c(x)|}{V(x)}.$$

Przez L_V^∞ będziemy oznaczać przestrzeń wszystkich borelowskich funkcji c , dla których $\|c\|_V < \infty$.

Prostym wnioskiem z powyższych rozważań, wynikającym z pracy [32] (Theorem 16.0.1) jest poniższy lemat.

Lemat 1: Załóżmy **(GE)**. Wówczas dla każdego $f \in F$, $g \in G$ indukowany proces $\{x_n\}$ jest V -jednostajnie ergodyczny, tzn., że istnieją stałe $\theta > 0$ i $\alpha \in (0, 1)$ takie, że

$$\|q^n(\cdot|x, f, g) - \pi_{fg}(\cdot)\|_V \leq R_x \alpha^n$$

z $R_x = V(x)\theta$, $x \in X$, $n \geq 1$. Przez $q^n(\cdot|x, f, g)$ oznaczamy prawdopodobieństwo przejścia w n -krokach, indukowane przez q oraz strategię stacjonarne $f \in F$ i $g \in G$.

Konsekwencją powyższego faktu jest Lemat 2.

Lemat 2: Dla dowolnych strategii stacjonarnych $f \in F$, $g \in G$ mamy

$$J_{fg} := J(x, f, g) = \frac{\int_X r(y, f, g)\pi_{fg}(dy)}{\int_X \tau(y, f, g)\pi_{fg}(dy)}.$$

W szczególności oznacza to, iż oczekiwana średnia wypłata nie zależy od stanu początkowego.

Kolejny lemat wynika z założenia **(GE)** i pokazuje między innymi jak otrzymać równanie Poissona związane z dowolną parą strategii stacjonarnych $f \in F$ i $g \in G$ [20, 18].

Lemat 3: Załóżmy **(GE)**. Dla $f \in F$ i $g \in G$

(a) funkcja $h_{fg} : X \mapsto \mathbf{R}$ zdefiniowana jako

$$h_{fg}(x) := E_x^{fg} \left(\sum_{n=0}^{\infty} [r(x_n, a_n, b_n) - J_{fg}\tau(x_n, a_n, b_n)] \right)$$

należy do przestrzeni L_V^∞ ;

(b) para (J_{fg}, h_{fg}) stanowi rozwiązanie równania Poissona

$$J_{fg}\tau(x, f, g) + h_{fg}(x) = r(x, f, g) + \int_X h_{fg}(y)q(dy|x, f, g).$$

Praca ze średnią wypłatą na jednostkę czasu, czyli wypłatą zdefiniowaną według wzoru (2), wymaga dodatkowych dwóch założeń:

R Warunek regularności procesu:

istnieją stałe $\epsilon > 0$ oraz $\beta < 1$ takie, że dla wszystkich $x \in C$, $a \in A(x)$ i $b \in B(x)$ zachodzi

$$Q(\epsilon|x, s, b) \leq \beta.$$

I „Warunek jednostajnej całkowalności”:

$$\lim_{t \rightarrow \infty} \sup_{x \in C} \sup_{a \in A(x), b \in B(x)} [1 - Q(t|x, a, b)] = 0.$$

Uwaga 2: Założenie (**R**) implikuje, iż nieskończona liczba skoków procesu nie zdarzy się w skończonym przedziale czasowym (nie dojdzie do tzw. eksplozji). Jednakże warunek ten wcale nie pokrywa się z (**B**, iii). Z (**R**) wynika jedynie, że

$$\tau(x, a, b) > \epsilon(1 - \beta) \quad \text{dla } x \in C.$$

Przytoczmy prosty przykład z pracy Sennott [46].

Niech $X = \{2, 3, \dots\}$, $A = B = \{1\}$ i $q(i+1|i, 1, 1) = 1$ dla każdego $i = 2, 3, \dots$. Czas pobytu w stanie i jest zmienną losową o rozkładzie

$$T_{i+1} - T_i := \begin{cases} \frac{1}{i^2} & \text{z prawdopodobieństwem } 1 - \frac{1}{i^2} \\ i^2 & \text{z prawdopodobieństwem } \frac{1}{i^2}. \end{cases}$$

Wówczas

$$1 < \tau(i, 1, 1) \leq \frac{19}{16}.$$

Jednak warunek regularności procesu nie jest spełniony. Mianowicie, jeśli proces startuje ze stanu $i = 2$, to w czasie 1 wykona nieskończenie wiele skoków z prawdopodobieństwem $\prod_{i=2}^{\infty} (1 - \frac{1}{i^2}) = \frac{1}{2}$.

Założenie (**I**) pomaga w udowodnieniu jednostajnej całkowalności pewnego półmartynału, która jest potrzebna do zastosowania twierdzenia Dooba o opcjonalnym stopowaniu.

Z matematycznego punktu widzenia założenie (**R**) nie jest potrzebne, jeśli rozpatrujemy kryterium (1). Jednak ze względu na liczne zastosowania warunek (**R**) zapobiega eksplozji procesu. We wcześniejszych pracach [41, 46, 49] zakładano, że nierówność (**R**) zachodzi dla wszystkich $x \in X$. W naszym przypadku, podobnie jak u Schälä [44], to nie jest konieczne. Z powyższych rozważań wiadomo bowiem, iż zanurzony proces Markowa z prawdopodobieństwem 1 powraca do zbioru C w skończonej ilości kroków. Zatem, wystarczy zażądać, aby nierówność w (**R**) była spełniona tylko dla $x \in C$. ♣

4. Aproksymacja gry semi-markowskiej o sumie zerowej.

W tym rozdziale opiszemy wyniki uzyskane w pracach [22, 23]. Głównym celem prezentacji jest pokazanie, że dla gry ze średnią wypłatą (1) spełnione jest równanie optymalności, a w szczególności, że ta gra ma wartość. Następnie na podstawie pracy [24] sprawdzimy, iż otrzymana wartość gry jest także wartością gry ze średnią wypłatą na jednostkę czasu oraz, że optymalne strategie graczy są takie same w obu przypadkach.

Zaprezentowana metoda rozwiązywania równania optymalności jest nowa, nawet w teorii markowskich procesów decyzyjnych. Polega ona na zaburzeniu prawdopodobieństwa przejścia q . Każda „ ε -zaburzona gra” może zostać rozwiązana za pomocą prostego algorytmu iteracyjnego, przypominającego procedurę znajdowania punktu stałego dla odwzorowania zwężającego (Twierdzenie 2). Gdy ε dąży do zera, to wartość gry w zaburzonym modelu dąży do wartości gry w oryginalnym modelu (Twierdzenie 1). W konsekwencji rozwiązanie wyjściowej gry może być otrzymane jako granica w pewien sposób zmodyfikowanych rozwiązań gier zaburzonych (Twierdzenie 3).

Standardowe podejście do procesów (gier) semi-markowskich dla przeliczalnej i borelowskiej przestrzeni stanów bazuje na transformacji Schweitzera ([13, 14, 20, 28]). Z wyjściowym modelem semi-markowskim jest kojarzony model markowski z odpowiednio zmodyfikowanymi wypłatami i prawdopodobieństwami przejścia. Wspomniany model markowski z czasem dyskretnym i kryterium średniej (granicznej) wypłaty jest badany przez aproksymację modelami dyskontowanymi gdy czynnik dyskonta dąży do jedności [19, 25, 28]. Metoda taka zwana „metodą znikającego dyskonta” jest bardzo trudna do zastosowania w przypadku procesów lub gier semi-markowskich z nieskończonym zbiorem stanów. Wynika to z postaci współczynnika dyskonta

$$\Delta_\alpha(x, a, b) := \int_0^\infty \exp(-\alpha t) Q(dt|x, a, b),$$

który zależy od $(x, a, b) \in K$. Próby rozwiązania równania optymalności tą metodą podejmowało wielu autorów, między innymi Ross [41], Sennott [46], Vega-Amaya [50]. Jednak podane założenia są bardzo trudne do zweryfikowania, nawet w przypadku ergodycznym. Dla skończonego zbioru stanów sytuacja jest nieco prostsza, bo można zastosować pewne klasyczne twierdzenia z algebry macierzy i analizy matematycznej [34].

Poprzednikiem naszej pracy [23] jest artykuł Lal i Sinhy [28]. Rozważali oni przeliczalną przestrzeń stanów i ograniczone funkcje wypłaty. Znacznie silniejsze założenia ergodyczności oraz podejście poprzez gry dyskontowane (transformacja Schweitzera) pozwoliło im na otrzymanie ograniczonego rozwiązania równania optymalności. Jednakże w tym przypadku, jak zauważył Cavazos-Cadena [9] (Theorem 5) gra przebiega w pewnym ograniczonym podzbiorze przestrzeni stanów, co redukuje możliwość otrzymania wielu interesujących zastosowań związanych z teorią kolejek [1, 16]. Innym artykułem dotyczącym gier semi-markowskich z borelowską przestrzenią stanów jest praca [39]. W pracy tej rozważane są skorelowane równowagi w klasie gier z ograniczonymi wypłatami przy założeniu, że $C = X$.

Niech $\varepsilon \in (0,1)$ i δ_s będzie miarą probabilistyczną skoncentrowaną w ustalonym stanie $s \in X$. Dla $(x, a, b) \in K$ zdefiniujemy

$$q_\varepsilon(\cdot|x, a, b) := (1 - \varepsilon)q(\cdot|x, a, b) + \varepsilon\delta_s(\cdot).$$

Zauważmy, że jeśli $\varepsilon \in (0, \varepsilon_0]$, gdzie

$$\varepsilon_0 := \frac{1 - \lambda}{2[V(s) - \lambda]},$$

to q_ε spełnia nierówność **(GE, i)** z λ zamienionym na $(\lambda + 1)/2 < 1$, $\eta > 0$ i tym samym zbiorem C . Z [33] wynika, że dla dowolnej funkcji $u \in L_V^\infty$ otrzymujemy

$$\left| \int_X u(y) q_\varepsilon^n(dy|x, f, g) - \int_X u(y) \pi_{fg}^\varepsilon(dy) \right| \leq V(x) \|u\|_V \theta_1 \alpha_1^n, \quad (4)$$

gdzie α_1 i θ_1 są stałymi niezależnymi od ε , a π_{fg}^ε jedyną niezmienniczą miarą probabilistyczną indukowaną przez q_ε , f i g . Lemat 1 zachodzi więc dla procesu markowskiego indukowanego przez $f \in F$, $g \in G$ i prawdopodobieństwo przejścia q_ε . W konsekwencji, dla dowolnych $f \in F$, $g \in G$ i $x \in X$ otrzymujemy

$$J_{fg}^\varepsilon := J^\varepsilon(x, f, g) = \frac{\int_X r(y, f, g) \pi_{fg}^\varepsilon(dy)}{\int_X \tau(y, f, g) \pi_{fg}^\varepsilon(dy)}.$$

Przez $J^\varepsilon(x, \pi, \gamma)$, dla dowolnych $\pi \in \Pi$, $\gamma \in \Gamma$, będziemy rozumieć oczekiwaną średnią wypłatę w „ ε -zaburzonej grze”.

Twierdzenie 1: (Theorem 1 w [22]). Załóżmy **(B, GE)**. Wówczas istnieje dodatnia stała d wyrażona przez stałe ustalone w powyższych założeniach taka, że dla dowolnego $\varepsilon \in (0, \varepsilon_0)$ mamy

$$\sup_{f \in F} \sup_{g \in G} |J_{fg} - J_{fg}^\varepsilon| \leq \varepsilon d.$$

Szkic dowodu: Ustalmy strategię $f \in F$ i $g \in G$. Przez indukcję matematyczną można sprawdzić, że

$$q_\varepsilon^n(\cdot|y, f, g) = \varepsilon\delta_s(\cdot) + (1 - \varepsilon)^n q^n(\cdot|y, f, g) + \sum_{k=1}^{n-1} \varepsilon(1 - \varepsilon)^k q^k(\cdot|y, f, g). \quad (5)$$

Korzystając z odpowiedniego lematu w [22] (Lemma 2), (5) i równości $(1 - \varepsilon)^n + \sum_{k=0}^{n-1} \varepsilon(1 - \varepsilon)^k = 1$ otrzymujemy oszacowanie dla każdego $n \geq 1$ i ustalonego stanu $s \in X$

$$\|q^n(\cdot|s, f, g) - q_\varepsilon^n(\cdot|s, f, g)\|_V \leq R_s \left[\alpha^n + \frac{\varepsilon}{1 - \alpha} \right] + \varepsilon \left[2V(s) + \frac{\eta}{1 - \lambda} \right].$$

Przypomnijmy, że $R_s = V(s)\theta$ (zobacz Lemat 1). Konsekwentnie, z nierówności trójkąta mamy

$$\begin{aligned} \|\pi_{fg}(\cdot) - \pi_{fg}^\varepsilon(\cdot)\|_V &\leq \|\pi_{fg}(\cdot) - q^n(\cdot|s, f, g)\|_V + \|q^n(\cdot|s, f, g) - q_\varepsilon^n(\cdot|s, f, g)\|_V \quad (6) \\ &\quad + \|q_\varepsilon^n(\cdot|s, f, g) - \pi_{fg}^\varepsilon(\cdot)\|_V \\ &\leq \frac{R_s\varepsilon}{1-\alpha} + \varepsilon \left[2V(s) + \frac{\eta}{1-\lambda} \right]. \end{aligned}$$

Korzystając z (6) możemy oszacować tempo zbieżności średniej wypłaty w ε -zaburzonej grze do średniej wypłaty w wyjściowym modelu gry, przy ε dążącym do zera. Stosując ponownie nierówność trójkąta oraz założenia (**B**, ii,iii) mamy

$$\begin{aligned} |J_{fg} - J_{fg}^\varepsilon| &= \left| \frac{\int_X r(x, f, g) \pi_{fg}(dx)}{\int_X \tau(x, f, g) \pi_{fg}(dx)} - \frac{\int_X r(x, f, g) \pi_{fg}^\varepsilon(dx)}{\int_X \tau(x, f, g) \pi_{fg}^\varepsilon(dx)} \right| \\ &\leq (S_1 + S_2) \|\pi_{fg} - \pi_{fg}^\varepsilon\|_V, \end{aligned}$$

gdzie

$$S_1 = \frac{M^* \|r\|_V \int_X V(x) \pi_{fg}(dx)}{(m^*)^2}, \quad S_2 = \frac{\|r\|_V}{m^*}.$$

Zatem

$$\sup_{g \in G} \sup_{f \in F} |J_{fg} - J_{fg}^\varepsilon| \leq \varepsilon d$$

z

$$d = (S_1 + S_2) \left[\frac{R_s}{1-\alpha} + 2V(s) + \frac{\eta}{1-\lambda} \right].$$

Zgodnie z Meyn i Tweedie [33] możemy wyrazić θ i α tylko za pomocą stałych $\eta, \delta, \lambda, v_C$. Z [33] (Theorem 2.3) otrzymujemy

$$\sup_{x \in X} \frac{\|q^n(\cdot|x, f, g) - \pi_{fg}(\cdot)\|_V}{V(x)} \leq \theta \alpha^n,$$

gdzie

$$\alpha := \rho = 1 - M_C^{-1}/2, \quad \theta := (1 + \gamma) \frac{\rho}{\rho - \vartheta} = (1 + \gamma)(2M_C - 1).$$

Dalej mamy

$$M_C = \frac{1}{(1 - \check{\lambda})^2} [1 - \check{\lambda} + \check{b} + \check{b}^2 + \bar{\zeta}_C(\check{b}(1 - \check{\lambda}) + \check{b}^2)];$$

$$\xi = \frac{4\eta + 2v_C\lambda\delta}{\delta^2};$$

$$\check{\lambda} = \frac{\lambda + \xi}{1 + \xi} < 1;$$

$$\check{b} = v_C + \xi < \infty;$$

$$\bar{\zeta}_C \leq \frac{4 - \delta^2}{\delta^5} \left(\frac{\eta}{1 - \lambda} \right)^2,$$

co kończy dowód. ♣

Uwaga 3: Zauważmy, iż oszacowanie odległości εd pomiędzy J_{fg}^ε i J_{fg} ustalone w Twierdzeniu 1 zostało wyrażone tylko przez stałe przyjęte w naszych podstawowych założeniach. Dlatego sądzimy, że ten wynik (razem z Twierdzeniem 2 poniżej) może mieć pewne praktyczne znaczenie. ♣

Niech $\varepsilon \in (0, \varepsilon_1)$, gdzie $\varepsilon_1 := (1 - \lambda)/(2V(s) + 1 - \lambda)$. Połóżmy

$$p(\cdot|x, a, b) := q_\varepsilon(\cdot|x, a, b) - \frac{\tau(x, a, b)\varepsilon}{M^*}\delta_s(\cdot)$$

dla $(x, a, b) \in K$.

Niech $T : L_V^\infty \mapsto L_V^\infty$ będzie operatorem zdefiniowanym następująco

$$(Tu)(x) = \min_{\rho \in P(B(x))} \max_{\nu \in P(A(x))} \left[r(x, \nu, \rho) + \int_X u(y)p(dy|x, \nu, \rho) \right], \quad (7)$$

dla wszystkich $x \in X$. Operator T jest dobrze określony i z twierdzenia Fana [11] wynika, że

$$(Tu)(x) = \max_{\nu \in P(A(x))} \min_{\rho \in P(B(x))} \left[r(x, \nu, \rho) + \int_X u(y)p(dy|x, \nu, \rho) \right] \quad (8)$$

dla każdego $x \in X$.

Twierdzenie 2: (Theorem 1 w [23]). Załóżmy **(B, GE)** i niech $\varepsilon \in (0, \varepsilon_1)$. Wówczas istnieje punkt stały $l_\varepsilon \in L_V^\infty$ operatora T , tzn.

$$l_\varepsilon(x) = (Tl_\varepsilon)(x) \quad (9)$$

dla każdego $x \in X$. Ponadto,

$$V_\varepsilon = \frac{\varepsilon l_\varepsilon(s)}{M^*} \quad (10)$$

jets wartością „ ε -zaburzonej gry” semi-markowskiej.

Szkic dowodu: Zdefiniujmy

$$l_n(x) := \inf_{\gamma \in \Gamma} \sup_{\pi \in \Pi} \mathcal{E}_x^{\pi\gamma} \left(\sum_{k=0}^{n-1} r(x_k, a_k, b_k) \right),$$

gdzie $\mathcal{E}_x^{\pi\gamma}$ oznacza operator wartości oczekiwanej względem miary indukowanej przez p i strategię $\pi \in \Pi$ oraz $\gamma \in \Gamma$. Korzystając z twierdzenia o selektorach minmaksowych [37] i zasady indukcji wstecznej można pokazać, że

$$l_n(x) = (T^n \mathbf{0})(x)$$

dla każdego $x \in X$. (Tutaj $0(x) := 0$). Łatwo sprawdzić, że ciąg $\{l_n\}$ jest ciągiem Cauchy'ego w L_V^∞ , a

$$l_\varepsilon(x) := \lim_{n \rightarrow \infty} l_n(x), \quad x \in X$$

jest punktem stałym operatora T . Pozostaje jeszcze wykazać, że V_ε jest wartością gry w modelu zaburzonym. Zatem, niech f i g będą strategiami stacjonarnymi: maksminową i minmaksową, odpowiednio, otrzymanymi z równania (9) (zobacz również (7), (8)). Wówczas wykorzystując metody programowania dynamicznego ([18, 22]) wnioskujemy, że

$$\begin{aligned} V_\varepsilon &= J_{fg}^\varepsilon = \sup_{\pi \in \Pi} \inf_{\gamma \in \Gamma} J^\varepsilon(x, \pi, \gamma) = \inf_{\gamma \in \Gamma} \sup_{\pi \in \Pi} J^\varepsilon(x, \pi, \gamma) \\ &= \sup_{\pi \in \Pi} J^\varepsilon(x, \pi, g) = \inf_{\gamma \in \Gamma} J^\varepsilon(x, f, \gamma) \end{aligned}$$

dla każdego $x \in X$. ♣

Parę $(V^*, h(\cdot))$, składającą się z liczby rzeczywistej V^* i borelowskiej funkcji $h \in L_V^\infty$ nazywamy *rozwiązaniem równania optymalności dla gry semi-markowskiej o sumie zerowej*, jeśli dla wszystkich $x \in X$ zachodzi

$$\begin{aligned} h(x) &= \min_{\rho \in P(B(x))} \max_{\nu \in P(A(x))} \left[r(x, \nu, \rho) + \int_X h(y) q(dy|x, \nu, \rho) - V^* \tau(x, \nu, \rho) \right] \\ &= \max_{\nu \in P(A(x))} \min_{\rho \in P(B(x))} \left[r(x, \nu, \rho) + \int_X h(y) q(dy|x, \nu, \rho) - V^* \tau(x, \nu, \rho) \right]. \end{aligned} \quad (11)$$

Przedstawiamy teraz główny rezultat tego rozdziału z pracy [23].

Twierdzenie 3: (Theorem 2 w [23]). Załóżmy **(B, GE)**. Wówczas

- (a) dla każdego $x \in X$ istnieją stała V^* , która jest wartością gry dla wyjściowego modelu oraz funkcja $h \in L_V^\infty$, które są rozwiązaniem równania optymalności (11);
- (b) istnieją borelowskie funkcje $f^0 \in F$ i $g^0 \in G$ takie, że

$$\begin{aligned} h(x) &= r(x, f^0, g^0) + \int_X h(y) q(dy|x, f^0, g^0) - V^* \tau(x, f^0, g^0) \\ &= \max_{\nu \in P(A(x))} \left[r(x, \nu, g^0) + \int_X h(y) q(dy|x, \nu, g^0) - V^* \tau(x, \nu, g^0) \right] \\ &= \min_{\rho \in P(B(x))} \left[r(x, f^0, \rho) + \int_X h(y) q(dy|x, f^0, \rho) - V^* \tau(x, f^0, \rho) \right] \end{aligned}$$

dla $x \in X$;

- (c) funkcja h jest jedynym rozwiązaniem równania (11) z dokładnością do stałej addytywnej.

Szkic dowodu:

(a) Ciąg rozwiązań $\{l_\varepsilon\}$, (przy $\varepsilon \rightarrow 0$) otrzymanych w Twierdzeniu 2 nie może być bezpośrednio użyty do konstrukcji rozwiązań równania optymalności w wyjściowym modelu gry. Poprzez prostą modyfikację (odjęcie pewnej stałej od l_ε) otrzymujemy ciąg $\{h_\varepsilon\}$, który jest jednostajnie ograniczony w przestrzeni L_V^∞ .

Zauważmy, że z Twierdzenia 1 wynika istnienie granicy:

$$V^* := \lim_{\varepsilon \rightarrow 0} V_\varepsilon = \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon l_\varepsilon(s)}{M^*}.$$

Ponadto, V^* jest częścią równania optymalności. Mianowicie z lematu Fatou dla zmieniających się miar [18] otrzymujemy

$$h_*(x) \geq \min_{\rho \in P(B(x))} \max_{\nu \in P(A(x))} \left[r(x, \nu, \rho) + \int h_*(y) q(dy|x, \nu, \rho) - \tau(x, \nu, \rho) V^* \right] \quad (12)$$

z $h_*(x) := \liminf_{n \rightarrow \infty} h_{\varepsilon_n}(x)$ ($\{\varepsilon_n\}$ jest tu pewnym ciągiem dążącym do zera) oraz

$$h^*(x) \leq \max_{\nu \in P(A(x))} \min_{\rho \in P(B(x))} \left[r(x, \nu, \rho) + \int h^*(y) q(dy|x, \nu, \rho) - \tau(x, \nu, \rho) V^* \right] \quad (13)$$

z $h^*(x) := \limsup_{n' \rightarrow \infty} h_{\varepsilon_{n'}}(x)$ ($\{\varepsilon_{n'}\}$ jest też pewnym ciągiem dążącym do zera, ale niekoniecznie tym samym co $\{\varepsilon_n\}$). Z metod programowania dynamicznego wynika, że

$$V^* \geq \sup_{\pi \in \Pi} J(x, \pi, g^*) \geq U(x),$$

gdzie g^* jest minmaksowym borelowskim selektorem w (12). Z drugiej strony

$$V^* \leq \inf_{\gamma \in \Gamma} J(x, f^*, \gamma) \leq L(x),$$

gdzie f^* jest maksminowym borelowskim selektorem w (13). Zatem V^* jest rzeczywiście wartością gry w wyjściowym modelu.

Kolejnym krokiem w dowodzie jest znalezienie funkcji $h \in L_V^\infty$, która spełniałaby (11) razem ze stałą V^* . Dzięki nierównościom (12), (13) oraz równaniu Poissona związanym z powyższymi strategiami $f^* \in F$, $g^* \in G$ (Lemat 3) można otrzymać równanie optymalności na pewnym borelowskim zbiorze $Z \subset X$, dla którego $\pi_{f^*g^*}(Z) = 1$. Następnie w sposób indukcyjny definiujemy ciąg borelowskich funkcji $h_k : X \mapsto \mathbf{R}$ i pokazujemy, że $h_k(x) \leq h_{k+1}(x)$ dla każdego $x \in X$ oraz $h_k \in L_V^\infty$ dla $k \geq 1$. Ostatecznie za rozwiązanie równania optymalności przyjmuje się funkcję $h(x) := \lim_{k \rightarrow \infty} h_k(x)$.

(b) Każdy maksminowy selektor $f^0 \in F$ otrzymany z równania (11) i każdy minmaksowy selektor $g^0 \in G$ w (11) są optymalnymi strategiami dla graczy i $V^* = J(x, f^0, g^0)$ dla wszystkich $x \in X$. Ponadto zachodzi

$$J(x, f^0, g^0) \geq \limsup_{n \rightarrow \infty} \frac{E_x^{\pi g^0} [\sum_{k=0}^{n-1} r(x_k, a_k, b_k)]}{E_x^{\pi g^0}(T_n)} \geq J(x, \pi, g^0)$$

oraz

$$J(x, f^0, g^0) \leq \liminf_{n \rightarrow \infty} \frac{E_x^{f^0 \gamma} [\sum_{k=0}^{n-1} r(x_k, a_k, b_k)]}{E_x^{f^0 \gamma}(T_n)} \leq J(x, f^0, \gamma)$$

dla $x \in X$ oraz $\pi \in \Pi$, $\gamma \in \Gamma$.

(c) Jednoznaczność równania optymalności opiera się na jednym z twierdzeń w [32] (Theorem 4.2.3), z którego pośrednio wynika, że dwa dowolne absorbujące zbiory korespondujące z różnymi strategiami stacjonarnymi nie mogą mieć pustego przekroju. Gdyby te zbiory były rozłączne, to wówczas indukowany proces $\{x_n\}$ miałby dwie klasy ergodyczne. ♣

Wykorzystując główny rezultat z naszej pracy [24] (zaprezentowany w rozdziale 5 jako Twierdzenie 5) możemy przedstawić następujący wynik.

Twierdzenie 4: Niech $f^0 \in F$, $g^0 \in G$ będą jak w Twierdzeniu 3. Załóżmy, że $(\mathbf{B}, \mathbf{GE}, \mathbf{R}, \mathbf{I})$ zachodzą. Wtedy (f^0, g^0) stanowi parę strategii optymalnych według kryterium (2).

Dowód: Na podstawie Twierdzenia 5 możemy napisać, że

$$j(x, f^0, g^0) = J(x, f^0, g^0) = V^* = \sup_{\pi \in \Pi} J(x, \pi, g^0) = \sup_{\pi \in \Pi} j(x, \pi, g^0)$$

oraz

$$j(x, f^0, g^0) = J(x, f^0, g^0) = V^* = \inf_{\gamma \in \Gamma} J(x, f^0, \gamma) = \inf_{\gamma \in \Gamma} j(x, f^0, \gamma)$$

dla wszystkich $x \in X$. ♣

5. Równoważność dwóch kryteriów optymalności w semi-markowskich procesach decyzyjnych.

W tym rozdziale opiszemy wyniki uzyskane w pracy [24]. Pokażemy, że kryteria średniej wypłaty (1) oraz średniej wypłaty na jednostkę czasu (2) pokrywają się dla stacjonarnych polityk oraz, że optymalne polityki stacjonarne graczy ze względu na pierwsze kryterium są również optymalnymi politykami w grze z drugim kryterium. Rdzeniem dowodu jest twierdzenie Dooba o opcjonalnym stopowaniu. Dodatkowe, ale niezbędne lematy zostały uzyskane dzięki założeniom ergodyczności $(\mathbf{GE})$ oraz znanym faktom z teorii odnowy [7, 41]. Równoważność kryteriów (dla polityk stacjonarnych) dla procesów ze skończonym zbiorem stanów i akcji została opisana w książce Mine i Osaki [34] w 1970 roku. Następnie w latach 70-tych Ross [41] uzyskał równość wypłat (1) i (2) dla procesów z przeliczalną przestrzenią stanów zakładając mocne warunki ergodyczności. Schäl w 1992 roku badał równanie optymalności, pokazując, że jest ono spełnione dla dwóch wspomnianych kryteriów. Założenia Schäla są znacznie ogólniejsze niż u Rossa i dopuszczają skończoną liczbę klas ergodycznych dla zanurzonego łańcucha Markowa. Dla borelowskiej przestrzeni stanów równanie optymalności z kryterium średniej wypłaty (1) przy ogólnych założeniach ergodyczności i stabilności stochastycznej zostało rozwiązane całkiem niedawno różnymi sposobami w [20, 22, 51]. Jednak żadna z tych prac nie wspomina o równoważności obu kryteriów dla polityk stacjonarnych. Nasz dowód z [24] zapożycza pewne metody z prac

Schäla [44] i Rossa [41], zwłaszcza w częściach wykorzystujących fakty z teorii odnowy.

Zaprezentowany opis wyników będzie dotyczył procesu decyzyjnego. W tym celu założymy, że gracz drugi ma ustaloną strategię stacjonarną $g \in G$. W ten sposób otrzymujemy model decyzyjnego procesu semi-markowskiego, w którym prawdopodobieństwo przejścia i wypłaty dane są jako

$$\tilde{q}(\cdot|x, a) := q(\cdot|x, a, g) \quad \text{oraz} \quad \tilde{r}(x, a) := r(x, a, g),$$

dla $x \in X$ i $a \in A(x)$. Podobnie definiujemy $\tilde{J}, \tilde{j}, \tilde{P}, \tilde{E}, \tilde{\tau}$. Z poprzedniej części wiadomo, że dla modelu spełniającego warunki **(B, GE)** istnieje rozwiązanie równania optymalności: $h \in L_V^\infty$ oraz stała $\tilde{g}$ takie, że

$$h(x) = \max_{a \in A(x)} \left[\tilde{r}(x, a) + \int_X h(y) \tilde{q}(dy|x, a) - \tilde{g} \tilde{\tau}(x, a) \right] \quad (14)$$

dla wszystkich $x \in X$. Ponadto,

$$\tilde{g} = \sup_{\pi \in \Pi} \tilde{J}(x, \pi) = \tilde{J}(x, f^*),$$

gdzie f^* jest borelowskim selektorem realizującym maksimum w (14).

Twierdzenie 5: ([24]). Załóżmy **(B, GE, R, I)**. Wówczas

(a)

$$\tilde{g} = \sup_{\pi \in \Pi} \tilde{j}(x, \pi),$$

(b)

$$\tilde{J}(x, f) = \tilde{j}(x, f) \quad \text{dla każdego } f \in F.$$

Dowód Twierdzenia 5 opiera się na analizie powrotów procesu stanów (generowanego przez prawdopodobieństwo przejścia $\tilde{q}$, politykę decydenta π) do zbioru C . W tym celu definiuje się indukcyjnie nowe zmienne losowe $N_C(k)$, określające losowe liczby skoków procesu potrzebne do k -tego powrotu do zbioru C , a mianowicie

$$N_C(1) := N_C = \min\{n \geq 1 : x_n \in C\}$$

oraz

$$N_C(k) := \min\{n > N_C(k-1) : x_n \in C\}.$$

Dzięki założeniu **(GE, i)** i formule Dynkina, które implikują twierdzenie porównawcze (Comparison Theorem 14.2.2 w [32]), można wyprowadzić następujące nierówności (Lema 2 w [24])

$$\tilde{E}_x^\pi \left(\sum_{n=0}^{N_C-1} V(x_n) \right) \leq \vartheta(x) := \frac{1}{1-\lambda} V(x) + \frac{\eta}{1-\lambda} 1_C(x),$$

$$\tilde{E}_x^\pi N_C \leq \phi(x) := \frac{1}{\ln(1/\lambda)} \left(\ln V(x) + \frac{\eta}{\lambda} 1_C(x) \right).$$

Przypomnijmy, że $\tilde{E}_x^\pi$ oznacza operator wartości oczekiwanej względem miary $\tilde{P}_x^\pi := P_x^{\pi g}$ indukowanej przez $\tilde{q}$ oraz π .

Proces z prawdopodobieństwem 1 zawsze powraca do zbioru C (w skończonej ilości kroków), a oczekiwana liczba skoków potrzebna do powrotu do C jest ograniczona przez funkcję V . Niestety, ten fakt wcale nie musi gwarantować regularności procesu, ponieważ nie znamy jego zachowania na zbiorze C . Dlatego przy pomocy założenia **(R)** definiujemy nowe zmienne losowe ograniczające z dołu czas pobytu procesu w kolejnych stanach. Dla stanów $x \in X \setminus C$ te zmienne losowe są równe 0. Natomiast jeśli proces znalazł się w $x \in C$, to te zmienne losowe zawsze wynoszą 0, gdy czas pobytu w x nie przekroczył ϵ , bądź są równe 0 lub ϵ z pewnymi prawdopodobieństwami. Formalnie, nowe zmienne mają rozkład dwupunktowy o dystrybuancie

$$H(t) := \begin{cases} \beta, & t \in [0, \epsilon) \\ 1, & t \geq \epsilon, \end{cases}$$

gdzie β and ϵ są stałymi z założenia **(R)**. Dla $t < 0$ kładziemy $H(t) := 0$. Przez $M(t)$ oznaczmy funkcję odnowy korespondującą ze zmiennymi losowymi o rozkładzie H , a mianowicie

$$M(t) := \sum_{n=0}^{\infty} H^{n*}(t).$$

H^{n*} jest n -tym splotem H ($H^{0*} := 1$ na $[0, +\infty]$ i $H^{0*} := 0$ dla $t < 0$). Dla dowolnego $x \in X$ oraz polityki $\pi \in \Pi$ określamy

$$M_x^\pi(t) := \tilde{E}_x^\pi \left(\sum_{m=1}^{\infty} 1_{[x_m \in C, T_m \leq t]} \right) = \sum_{m=1}^{\infty} \tilde{P}_x^\pi(x_m \in C, T_m \leq t).$$

Jest to oczekiwana liczba wizyt procesu w zbiorze C w czasie $[0, t]$. Dzięki standardowym metodom stosowanym w teorii odnowy oraz z faktu, że różnice $\{T_{n+1} - T_n\}$, $n = 0, 1, \dots$ stanowią ciąg warunkowo niezależnych zmiennych losowych (pod warunkiem danego procesu stanów i akcji $(x_0, a_0, g(x_0), x_1, a_1, g(x_1), \dots)$ można pokazać nierówności (Lemma 3 w [24], [7, 44])

$$M_x^\pi(t) \leq M(t) < \infty, \quad M_x^\pi(t+h) - M_x^\pi(t) \leq M(h).$$

Dodatkowo dla funkcji ograniczonej, nierosnącej i nieujemnej $z(t)$, dla której $\lim_{t \rightarrow \infty} z(t) = 0$, mamy

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t z(t-u) M_x^\pi(du) = 0. \quad (15)$$

W dowodzie tej granicy należy wykorzystać wcześniejsze nierówności oraz twierdzenie odnowy [7, 41] dla zmiennych losowych o rozkładzie H

$$\frac{M(t)}{t} \rightarrow \frac{1}{\epsilon(1-\beta)},$$

gdy $t \rightarrow \infty$.

Zanim podamy szkic dowodu Twierdzenia 5, przytoczymy najważniejsze fakty w poniższym lemacie. Niech $\mathcal{F}_n$ oznacza σ -algebrę generowaną przez wszystkie zdarzenia kończące się na n -tym stanie.

Lemat 4: Dla dowolnej polityki $\pi \in \Pi$ oraz $x \in X$

(a) (Lemma 5 w [24])

$$\begin{aligned} \tilde{E}_x^\pi \left(\sum_{n=0}^{N(t)} |\tilde{r}(x_n, a_n)| \right) &\leq L\vartheta(x) + L\vartheta_C M(t) < \infty, \quad \vartheta_C := \sup_{x \in C} \psi(x), \\ \tilde{E}_x^\pi \left(\sum_{n=0}^{N(t)} \tilde{r}(x_n, a_n) \right) &\leq M^* \phi(x) + M^* \phi_C M(t) < \infty, \quad \phi_C := \sup_{x \in C} \phi(x). \end{aligned}$$

Stałe L , M^* są używane w założeniach (B, iii,v).

(b) (Lemma 6 w [24])

$$\lim_{t \rightarrow \infty} \frac{1}{t} \tilde{E}_x^\pi V(x_{N(t)+1}) = 0.$$

(c) (Lemma 7 w [24])

$$\lim_{t \rightarrow \infty} \frac{1}{t} \tilde{E}_x^\pi T_{N(t)+1} = 1.$$

Szkic dowodu:

(a) Dowód polega na odpowiednim podziale przestrzeni Ω . Mianowicie, patrzymy na proces stanów i liczymy wartość oczekiwaną $\tilde{E}_x^\pi$ na trajektoriach pomiędzy kolejnymi powrotami procesu do zbioru C . Następnie korzystamy z uogólnionej wersji mocnej własności Markowa. Polega ona na przedefiniowaniu przestrzeni stanów i funkcji prawdopodobieństwa przejścia tak, aby z procesu indukowanego przez $\tilde{q}$ oraz π (niekoniecznie markowskiego) otrzymać proces markowski. Wówczas dla takiego procesu stosujemy już mocną własność Markowa.

(b) Dowód tej części jest dość skomplikowany dlatego dla przejrzystej analizy został podzielony na kilka kroków. Zaprezentujemy więc tylko główne punkty. Dla wygody półośmy

$$w_x^\pi(t) := \tilde{E}_x^\pi(V(x_{N(t)+1}); N_C > N(t) + 1) \quad \text{ i } \quad w_x(t) := \sup_{\pi \in \Pi} w_x^\pi(t).$$

Funkcja $w_x(t)$ jest uniwersalnie mierzalna względem x dla każdego $t \geq 0$ (Lemma 4 w [24] oparty na [3, 48]). Korzystając z twierdzenia Dooba o opcjonalnym stopowaniu [30, 36] dla dwóch czasów zatrzymania $N(t_1) + 1$ i $N(t_2) + 1$ ($t_1 < t_2$ a stąd $N(t_1) + 1 \leq N(t_2) + 1$) można pokazać, że $w_x^\pi(t)$ jest funkcją nierosnącą zmiennej t . W tym celu należy jedynie zauważyć, iż $\{V(x_n)1_{[N_C > n]}\}$, $n \geq 1$ jest jednostajnie całkowalnym supermartyngałem względem $\{\mathcal{F}_n\}$. Ponadto z monotoniczności funkcji $t \mapsto w_x^\pi(t)$, pracy [38] oraz założenia (I) wynika

$$\lim_{t \rightarrow \infty} \sup_{x \in C} w_x(t) = 0.$$

Definiujemy

$$z_x^\pi(t) := \tilde{E}_x^\pi(V(x_{N(t)+1}); N_C \geq N(t) + 1) < \infty \quad \text{i} \quad z(t) := \sup_{x \in C} \sup_{\pi \in \Pi} z_x^\pi(t).$$

Postępując analogicznie jak wyżej można pokazać, że funkcja $z(t)$ jest nierosnąca względem t i $z(t) < \infty$. Ponadto, korzystając z uogólnionej mocnej własności Markowa otrzymamy

$$\tilde{E}_x^\pi V(x_{N(t)+1}) \leq z_x^\pi(t) + \int_0^t z(t-u) M_x^\pi(du).$$

Aby zastosować (15) pozostaje sprawdzić, że $\lim_{t \rightarrow \infty} z(t) = 0$. To wynika z

$$z(t) \leq \sup_{x \in C} \sup_{\pi \in \Pi} w_x^\pi(t) + \sup_{x \in C} \sup_{\pi \in \Pi} E_x^\pi(V(x_{N(t)+1}); N_C = N(t) + 1)$$

i zastosowania nierówności Markowa dla drugiego składnika powyższej sumy.

(c) Dowód wynika z faktu, iż $[N(t) + 1 = n] \in \mathcal{F}_n$ oraz

$$\frac{t}{t} \leq \tilde{E}_x^\pi \frac{T_{N(t)+1}}{t} \leq \frac{\tilde{E}_x^\pi T_{N(t)} + \tilde{E}_x^\pi (T_{N(t)+1} - T_{N(t)})}{t} \leq \frac{t + M}{t}. \quad \clubsuit$$

Szkic dowodu Twierdzenia 5:

(a) Zauważmy, że $\{S_n\}$ zdefiniowany jako

$$S_n := \sum_{k=0}^{n-1} (\tilde{r}(x_k, a_k) - \tilde{g}\tilde{\tau}(x_k, a_k)) + h(x_n)$$

jest supermartyngałem względem $\{\mathcal{F}_n\}$. Wynika to z równania optymalności (14):

$$h(x) \geq \tilde{r}(x, a) - \tilde{g}\tilde{\tau}(x, a) + \int_X h(y) q(dy|x, a).$$

Łatwo sprawdzić na podstawie Lematu 4, że $\tilde{E}_x^\pi |S_{N(t)+1}| < \infty$ oraz, że $\{S_n\}$ jest jednostajnie całkownym supermartyngałem, tzn.

$$\lim_{n \rightarrow \infty} \tilde{E}_x^\pi [|S_n|; N(t) \geq n] = 0.$$

Zatem stosując dla tego supermartyngału twierdzenie Dooba o opcjonalnym stopowaniu [30, 36] dla dwóch czasów zatrzymania 0 i $N(t) + 1$ otrzymujemy

$$h(x) \geq \tilde{E}_x^\pi \left(\sum_{k=0}^{N(t)} (\tilde{r}(x_k, a_k) - \tilde{g}\tilde{\tau}(x_k, a_k)) \right) + \tilde{E}_x^\pi h(x_{N(t)+1}),$$

a w konsekwencji

$$\tilde{g} \frac{1}{t} \tilde{E}_x^\pi T_{N(t)+1} \geq \frac{1}{t} \tilde{E}_x^\pi \left(\sum_{k=0}^{N(t)} \tilde{r}(x_k, a_k) \right) + \tilde{E}_x^\pi \left(\frac{h(x_{N(t)+1})}{t} \right) - \frac{h(x)}{t}.$$

Lewa strona dąży do $\tilde{g}$ na podstawie Lematu 4, a prawa strona dąży do średniej wypłaty na jednostkę czasu $\tilde{j}(x, \pi) = j(x, \pi, g)$.

(b) W tej części jako punkt wyjścia wystarczy wziąć równanie Poissona (korespondujące z dowolną polityką stacjonarną $f \in F$ - Lemat 3) i zastosować twierdzenie Dooba o opcjonalnym stopowaniu dla odpowiedniego jednostajnie całkowalnego martyngału [30, 36]. ♣

6. Przykłady.

1. Model opisujący zamianę urządzenia. Pewne pracujące urządzenie narażone jest na gwałtowne wstrząsy, które pojawiają się w losowym czasie. Każdy taki wstrząs powoduje losową ilość uszkodzeń, które gromadzone są przez cały czas pracy urządzenia. Czas pomiędzy kolejnymi wstrząsami jest zmienną losową o dystrybuancie H . Przez $G(\cdot|x)$ oznaczmy warunkową dystrybuantę wielkości wstrząsu związaną ze stanem x . Jeśli nagromadzona ilość uszkodzeń wynosi x i pojawia się wstrząs o wielkości y , to urządzenie przestaje pracować z prawdopodobieństwem $1 - r(x + y)$, gdzie r funkcją nierosnącą, zwaną także „funkcją przetrwania.” System jest kontrolowany w następujący sposób: w czasie wstrząsu bądź po wymianie urządzenia jest wybierany planowany czas pracy a tej maszyny, tzn. czas do kolejnej wymiany. W ten sposób moment skoku systemu to czas wstrząsu bądź koniec planowanego czasu pracy, który został wcześniej wyznaczony dla tego urządzenia. Opisany model może być postrzegany jako proces semi-markowski, dla którego

- $X = [0, \infty)$ (przestrzeń stanów);
- $A = A(x)$ dla każdego $x \in X$ (przestrzeń decyzji); A jest zwartym podzbiorem zbioru $(0, M^*]$;
- prawdopodobieństwo przejścia $\tilde{q}$ dane jest wzorem

$$\begin{aligned} \tilde{q}(B|x, a) &:= \delta_0(B) \left[1 - H(a) + H(a) \int_0^\infty (1 - r(x + y)) G(dy|x) \right] \\ &\quad + H(a) \int_0^\infty 1_B(x + y) r(x + y) G(dy|x), \end{aligned}$$

dla każdego zbioru $B \in \mathcal{B}(X)$;

- dystrybuanta czasu pobytu w stanie x , przy podjętej decyzji a jest postaci

$$\tilde{Q}(t|x, a) := \begin{cases} H(t), & t \leq a \\ 1, & t > a; \end{cases}$$

- funkcją kosztu jest

$$c(x, a) := x\tilde{\tau}(x, a) + k(1 - H(a)) + KH(a),$$

gdzie $k < K$ są pewnymi stałymi; innymi słowy, koszt naprawy jest proporcjonalny do ilości zgromadzonych uszkodzeń; poza tym, jeśli urządzenie przestanie pracować to koszt naprawy jest dużo wyższy niż w przypadku, gdy użytkownik zdecyduje się je wymienić wcześniej.

Założmy, że $m^* = \inf A$.

A1 $H(\cdot)$ jest absolutnie ciągła względem miary Lebesgue'a na $\mathbf{R}$ oraz $H(m^*) > 0$ i $H(M^*) < 1$;

A2 istnieje $s > 0$ takie, że

$$\int_0^\infty e^{sy} r(x+y) G(dy|x) \leq 1,$$

(w szczególności można założyć, że $e^{sy} r(y) \leq 1$ dla każdego $y \in X$).

Zauważmy, iż

$$\tilde{\tau}(x, a) = \int_0^a z H(dz) + a(1 - H(a))$$

oraz

$$\begin{aligned} \int_0^\infty V(y) \tilde{q}(dy|x, a) &= \int_0^\infty e^{sy} \tilde{q}(dy|x, a) = H(a) \int_0^\infty (1 - r(x+y)) G(dy|x) \\ &\quad 1 - H(a) + H(a) \int_0^\infty e^{s(x+y)} r(x+y) G(dy|x). \end{aligned}$$

Zatem podstawowe założenia **(B)** są spełnione, a $V(x) := e^{sx}$. Ponadto,

$$c(x, a) \leq V(x) \frac{K + k + M^*}{s} \quad \text{i} \quad m^* \leq \tilde{\tau}(x, a) \leq M^*.$$

Niech teraz

$$\lambda := \frac{1 + H(M^*)}{2} < 1, \quad \eta := 1.$$

Łatwo sprawdzić, że nierówność **(GE, i)** zachodzi z powyższymi stałymi oraz $C := [0, \frac{\ln 2 - \ln(1 - H(M^*))}{s}]$, bowiem

$$\int_0^\infty e^{sy} \tilde{q}(dy|x, a) \leq 1 + H(M^*) e^{sx} \leq \lambda e^{sx} + 1_C(x).$$

Założenie **(GE, ii)** jest spełnione z $v_C := \frac{2}{1 - H(M^*)}$. Jeśli przyjmiemy, że $\delta := 1 - H(m^*)$ oraz $\mu(D) := \delta_0(D)$, wówczas

$$\tilde{q}(D|x, a) \geq \delta \mu(D),$$

co daje **(GE, iii)**. Warunek **(I)** jest spełniony, a warunek regularności **(R)**, zapobiegający eksplozji procesu zachodzi z $\beta := H(m^*)$ i $\epsilon := m^*$:

$$\tilde{Q}(m^*|x, a) = H(m^*), \quad \text{bo} \quad m^* \leq a.$$

Zatem wszystkie założenia dla modelu są spełnione ³. ♣

2. Zaprezentujemy przykład ilustrujący fakt, iż dla łańcucha nieergodycznego wspomniane dwa kryteria optymalności (1) i (2) nie są identyczne, nawet w przypadku polityk stacjonarnych. Niech $X = \{1, 2, 3\}$, $A = a = A(i)$ dla $i = 1, 2, 3$ oraz

$$T_{i+1} - T_i := \begin{cases} 2, & i = 3 \\ 1, & i \in \{1, 2\}. \end{cases}$$

Prawdopodobieństwo przejścia dane jest wzorem

$$\tilde{q}(j|i, a) := \begin{cases} \frac{1}{2}, & i = 1, j \in \{2, 3\} \\ 1, & i = j = 2, i = j = 3 \\ 0, & \text{poza tym.} \end{cases}$$

Funkcja wypłaty jest postaci

$$\tilde{r}(i, a) := \begin{cases} 0, & i \in \{1, 3\} \\ 1, & i = 2. \end{cases}$$

Gdy stanem początkowym jest $i_0 = 1$, proces spędza tam jedną jednostkę czasu a gracz nic nie otrzymuje. Potem gra przeskakuje z jednakowym prawdopodobieństwem do stanu 2 lub 3. Ze stanu $i \geq 2$ nie można przejść do żadnego innego stanu. Jeśli proces jest w stanie $i = 2$, to co jednostkę czasu mamy „skok”, w którym jednak nie występuje zmiana stanu. Za każdym razem gracz otrzymuje wypłatę 1. Jeśli $i = 3$, to skok ma miejsce co dwie jednostki czasu, przy czym wypłaty gracza są zerowe. Zbiór $A(i)$ jest jednoelementowy, zatem istnieje tylko jedna strategia stacjonarna f , dla której $J(1, f) = \frac{1/2}{3/2} = \frac{1}{3}$ i $j(1, f) = \frac{1}{2}$. Stąd $J(1, f) \neq j(1, f)$. ♣

Literatura

- [1] E. ALTMAN I A. HORDIJK, *Zero-sum Markov games and worth case optimal control of queueing systems*, invited paper, QUESTA, 21 (1995), special issue on optimization of queueing systems, Ed. S. Stidham, str. 415–447.
- [2] R.E. BELLMAN, *Dynamic Programming*, Princeton University Press, Princeton, 1957.
- [3] D.P. BERTSEKAS I S.E. SHREVE, *Stochastic Optimal Control: the Discrete Time Case*, Academic Press, New York, 1978.
- [4] R.N. BHATTACHARYA I M. MAJUMDAR, *Controlled semi-Markov models under long-run average rewards*, J.Statist. Plann. Inference, 22 (1989), str. 223–242.

³Inne przykłady związane z klasą założeń przyjętych w pracy można znaleźć w [16, 18, 22, 32].

- [5] T.R. BIELECKI I M. FREI, *On identification and control in the partially known Merton's portfolio selection model*, J. Optim. Theory Appl., 77 (1993), str. 399–420.
- [6] T. BIELECKI, D. HERNÁNDEZ-HERNÁNDEZ I S.R. PLISKA, *Risk sensitive control of finite state Markov chains in discrete time, with applications to portfolio management*, Technical Raport, Dept. of Finance, Univ. of Illinois at Chicago, 1999.
- [7] D. BLACKWELL, *A renewal theorem*, Duke Math. J., 15 (1948), str. 145–150.
- [8] L.D. BROWN AND R. PURVES, *Measurable selections of extrema*, Ann. Stat., 1 (1973), str. 902–912.
- [9] R. CAVAZOS-CADENA, *Recent results on conditions for the existence of average optimal stationary policies*, Ann. Oper. Res., 28 (1991), pp. 3–28.
- [10] D. DUFFIE, *Security Markets Stochastic Models*, Academic Press, New York, 1988.
- [11] K. FAN, *Minimax theorems*, Proc. Nat. Acad. Sci. USA, 39 (1953), str. 42–47.
- [12] A. FEDERGRUEN, H. HORDIJK I H.C. TIJMS, *Denumerable state semi-Markov decision process with unbounded cost criterion*, Stoch. Process. Appl., 9 (1979), str. 223–235.
- [13] A. FEDERGRUEN, P.J. SCHWEITZER I H.C. TIJMS, *Denumerable undiscounted semi-Markov decision processes with unbounded rewards*, Math. Oper. Res., 8 (1983), str. 298–313.
- [14] A. FEDERGRUEN I H.C. TIJMS, *The optimality equation in average cost denumerable state semi-Markov decision problems. Recurrence conditions and algorithms*, J. Appl. Probab., 15 (1978), str. 356–373.
- [15] E.A. FEINBERG, *Constrained semi-Markov decision processes with average rewards*, Math. Methods Oper. Res., 39 (1994), str. 257–288.
- [16] EDS. E.A. FEINBERG I A. SHWARTZ, *Handbook of Markov Decision Processes Methods and Applications*, Kluwer Academic Publishers, 2002.
- [17] O. HERNÁNDEZ-LERMA I J.B. LASSERRE, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York, 1996.
- [18] O. HERNÁNDEZ-LERMA I J.B. LASSERRE, *Further Topics on Discrete-Time Markov Control Processes*, Springer-Verlag, New York, 1999.
- [19] O. HERNÁNDEZ-LERMA I J.B. LASSERRE, *Zero-sum stochastic games in Borel spaces: average payoff criteria*, SIAM J. Control Optim., 39 (2001), str. 1520–1539.
- [20] O. HERNÁNDEZ-LERMA I F. LUQUE-VÁSQUEZ, *Semi-Markov control models with average costs*, Applicationes Mathematicae, 26 (1999), str. 315–331.
- [21] R.A. HOWARD, *Research in semi-Markovian decision structures*, J. Oper. Res. Soc. Japan, 6 (1964), str. 163–199.
- [22] A. JAŚKIEWICZ, *An approximation approach to ergodic semi-Markov control processes*, Math. Methods of Oper. Res., 54 (2001), str. 1–19.

- [23] A. JAŚKIEWICZ, *Zero-sum semi-Markov games*, SIAM J. Control Optim., 41 (2002), str. 723–739.
- [24] A. JAŚKIEWICZ, *On the equivalence of two expected average cost criteria for semi-Markov control processes*, submitted.
- [25] A. JAŚKIEWICZ I A.S. NOWAK, *On the optimality equation for zero-sum ergodic stochastic games*, Math. Methods of Oper. Res., 54 (2001), str. 291–301.
- [26] W.S. JEWELL, *Markov renewal programming: I and II*, Oper. Res., 11 (1963), str. 938–971.
- [27] M. KURANO, *Semi-Markov decision processes and their applications in the replacement models*, J. Oper. Res. Soc. Japan, 28 (1985), str. 18–29.
- [28] A.K. LAL I S. SINHA, *Zero-sum two person semi-Markov games*, J. Appl. Prob., 29 (1992), str. 56–72.
- [29] S.A. LIPPMAN, *Semi-Markov decision processes with unbounded rewards*, Manage. Sci., 19 (1973), str. 1225–1233.
- [30] A.P. MAITRA AND W.D. SUDDERTH, *Discrete Gambling and Stochastic Games*, Springer-Verlag, New York, 1996.
- [31] R. MENDELSSOHN I M.J. SOBEL, *Capital accumulation and the optimization of renewable resource models*, J. Econ. Theory, 23 (1980), str. 242–260.
- [32] S.P. MEYN I R.L. TWEEDIE, *Markov Chains and Stochastic Stability*, Springer-Verlag, New York, 1993.
- [33] S.P. MEYN I R.L. TWEEDIE, *Computable bounds for geometric convergence rates of Markov chains*, Ann. Appl. Probab., 4 (1994), str. 981–1011.
- [34] H. MINE I S. OSAKI, *Markovian Decision Processes*, Elsevier, New York, 1970.
- [35] J. NEVEU, *Mathematical Foundations of the Calculus of Probability*, Holden-Day, San Francisco, 1965.
- [36] J. NEVEU, *Discrete-Parameter Martingales*, Elsevier, New York, 1975.
- [37] A.S. NOWAK, *Measurable selection theorems for minimax stochastic optimization problems*, SIAM J. Control Optim., 23 (1985), str. 466–476.
- [38] A.S. NOWAK, *Universally measurable strategies in zero-sum stochastic games*, Ann. Probab., 13 (1985), str. 269–287.
- [39] A.S. NOWAK, *Some remarks on equilibria in semi-Markov games*, Applicationes Mathematicae, 27 (2000), str. 385–394.
- [40] M.L. PUTERMAN, *Markov decision Processes*, John Wiley, New York, 1994.
- [41] S.M. ROSS *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, 1970.

- [42] W.J. Runggaldier I A. Zaccaria, *A stochastic control approach to risk management under restricted information*. Technical Raport, Dept. of Math., Univ. of Padova, Italy, 1999.
- [43] P.A. Samuelson, *Lifetime portfolio selection by dynamic stochastic programming*, Rev. Econ. Studies, 51 (1984), str. 239–246.
- [44] M. Schäl, *On the second optimality equation for semi-Markov decision models*, Math. Oper. Res., 17 (1992), pp. 470–486.
- [45] P.J. Schweitzer, *Iterative solutions of the functionals equations of undiscounted Markov renewal programming*, J. Math. Anal. Appl., 34 (1971), str. 495–501.
- [46] L.I. Sennott, *Average cost semi-Markov decision processes and the control of queueing system*, Probability in the Engineering and Informational Sciences, 3 (1989), str. 247–272.
- [47] L.S. Shapley, *Stochastic games*, Proc. Nat. Acad. Sci. USA, 39 (1953), str. 1095–1100.
- [48] R.E. Strauch, *Negative dynamic programming*, Ann. Stat., 37 (1966), str. 871–890.
- [49] A. Yushkevich, *On semi-Markov controlled models with an average reward criterion*, Theory Probab. Appl., 26 (1981), str. 796–803.
- [50] O. Vega-Amaya, *Average optimality in semi-Markov control models on Borel spaces: unbounded costs and controls*, Bol. Soc. Matem. Mexicana, 38 (1993), str. 47–68.
- [51] O. Vega-Amaya and F. Luque-Vásquez, *Sample-path average cost optimality for semi-Markov control processes on Borel spaces: unbounded costs and mean holding times*, Applicationes Mathematicae, 27 (2000), str. 343–367.