

JOANNA TYRCHA

Warszawa

Model stałego prawdopodobieństwa przejścia w badaniach cyklu komórkowego*

(Praca wpłynęła do Redakcji 1986.03.03)

1. Wstęp. W biologii cyklu komórkowego przyjmuje się powszechnie, że tempo proliferacji jest regulowane przy przechodzeniu komórki od mitozy do syntezy DNA (czyli w tak zwanej fazie G_1 cyklu). Stąd im wolniejsze tempo proliferacji danej populacji, tym komórki dłużej pozostają w fazie przed syntezą DNA, podczas gdy suma czasów pozostałych faz cyklu pozostaje w zasadzie stała, różniąc się nieznacznie, nawet przy dużych zmianach tempa wzrostu. Między innymi, te właśnie obserwacje doprowadziły do przyjęcia hipotezy istnienia w cyklu swoistego „przełącznika”, który w zależności od warunków (zarówno wewnętrznych, jak i zewnętrznych) pozwala (pozycja włączona) lub nie (pozycja wyłączona) na inicjację syntezy DNA.

Znaczna zmienność czasowa okresu poprzedzającego syntezę DNA, a także względna czasowa stałość pozostałych faz cyklu komórkowego pozwoliła rozpatrywać cykl komórkowy jako złożony z dwóch przedziałów: jednego zmiennego w czasie, trwającego od zakończenia podziału do momentu zajścia zjawiska decydującego o rozpoczęciu syntezy DNA, oraz drugiego względnie stałego czasowo, zajmującego pozostałą część cyklu.

Taki podział cyklu komórkowego oraz dodatkowa obserwacja kinetyki proliferacji różnych typów populacji komórkowych doprowadziły do powstania wielu modeli próbujących tłumaczyć zjawiska regulacji wzrostu populacji komórkowych. Jednym z bardziej znanych jest tzw. *model prawdopodobieństwa przejścia* (ang. *transition probability model*, w skrócie model TP). Rozpatrzymy model TP dla komórek siostrzanych oraz przeprowadzimy dyskusję kilku założeń modelowych spotykanych najczęściej w literaturze. Wykażemy przy tym sprzeczność pewnego często przyjmowanego układu aksjomatów.

Niniejszy artykuł nawiązuje do pracy Ćwika i in. [3] oraz do pracy Ćwika

* Problem MR I.3.

i Gołembiewskiej [4]. Dziękuję dr Gołembiewskiej-Skoczylas z Instytutu Zoologii UW, która pozwoliła mi wykorzystać zebrane przez siebie dane oraz z którą konsultowałam moją pracę.

2. Cykl komórkowy. Centralnym problemem w biologii cyklu komórkowego jest odpowiedź na pytanie, jakie są mechanizmy regulujące wzrost komórki i jej reprodukcję. Jak dotąd brak jest jednolitej koncepcji opisującej te mechanizmy, a tym samym wyjaśniającej, co dzieje się przy niekontrolowanym wzroście prowadzącym do patologicznych zmian nowotworowych.

Cykl komórkowy większości komórek można podzielić (głównie ze względu na zachowanie się materiału genetycznego) na następujące fazy (rys. 1):

- G_1 — faza presyntetyczna, poprzedzająca okres syntezy DNA (ang. *gap one* — przerwa pierwsza),
- S — okres syntezy DNA,
- G_2 — faza postsyntetyczna, następująca po okresie syntezy DNA (ang. *gap two* — przerwa druga),
- M — mitozą (proces przekazywania informacji genetycznej komórki macierzystej do nowopowstających komórek potomnych); jej przebieg związany jest m.in. ze zmianami morfologicznymi w chromatinie jądrowej (wyróżnicowanie chromosomów),
- D — cytokineza, czyli rozdział substancji cytoplazmatycznej; cytokineza kończy się fizycznym rozdzieleniem komórek potomnych.

W większości, fazy te nie dają się obserwować bezpośrednio, a określenie, w jakim momencie cyklu znajduje się komórka, wymaga na ogół jej utrwalenia, to znaczy uśmiercenia przy możliwie jak najlepszym zachowaniu jej struktury. Ze względu na to, że populacje komórkowe (poza nielicznymi wyjątkami) rosną asynchronicznie, w ustalonym momencie czasowym mamy do czynienia z mieszaniną komórek o różnych fazach cyklu, co gorsze, określenie faz pozostaje poza możliwością obserwowania. W związku z tym powstają modele próbujące tłumaczyć mechanizmy przechodzenia komórki przez poszczególne etapy cyklu. Można wyróżnić kilka kierunków badań.

W niektórych badaniach stawia się bezpośrednio hipotezy dotyczące przyczyn, które determinują przebieg cyklu komórkowego. Na przykład, usiłuje się go związać z procesami chemicznymi i fizycznymi zachodzącymi w błonie komórkowej, tj. z procesami utleniania lipidów (proces chemiczny) lub ze zmianą stanu warstwy lipidowej wchodzącej w skład błony (proces fizyczny). Takie podejście przedstawione jest m.in. w monografii [2]. Podejście to trafia jednak na rozmaite przeszkody. Należą do nich niedostateczny poziom wiedzy z zakresu biofizyki i biochemii oraz wspomniana już konieczność utrwalenia komórki przy pomiarze interesujących wielkości. Co więcej, ogromną trudność sprawia fakt, że mierzone wielkości mają w populacjach komórek rozkłady prawdopodobieństwa zmieniające się na ogół z upływem czasu.

Innym kierunkiem są badania wiążące zjawiska cyklu z produktami

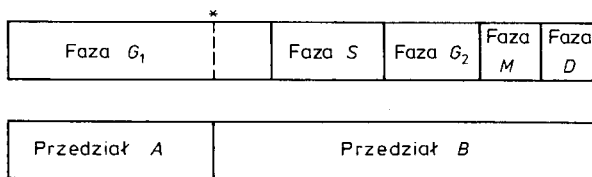
określonych genów. Badania te są jednak ograniczone tylko do specyficznych komórek, a wyprowadzone wnioski nie mogą być uogólnione.

Bardzo szerokim kierunkiem badawczym jest eksperymentalne zakłócanie cyklu pewnym czynnikiem. Najczęściej wybiera się dogodny czynnik i bada jego wpływ na rozkład mierzonych wielkości (np. na czas trwania cyklu komórkowego). Usiłuje się w ten sposób stworzyć narzędzie do badania różnych hipotez o cyklu. Poszukuje się wtedy stacjonarnych parametrów tych rozkładów dla odpowiednio dobranych planów doświadczenia.

Jako cechę obserwowalną wybiera się często czas trwania cyklu komórkowego, zwany też *czasem generacji* (cecha T). Jest to podstawowa opisująca cykl cecha, która daje się obserwować bezpośrednio i jest wrażliwa na wpływy wielu czynników. Cecha ta nie jest całkowicie zdeterminowana genetycznie, lecz zależy od szybkości przebiegu zjawisk wewnątrzkomórkowych warunkujących szybsze lub wolniejsze przechodzenie komórki przez kolejne etapy cyklu. Na podstawie zmian cechy T można więc wnioskować o tym, co dzieje się w komórce podczas cyklu.

Zróznicowanie czasów generacji komórek o identycznych genotypach zwróciło uwagę na stochastyczną naturę procesu determinującego długości poszczególnych faz cyklu. Zaczęto więc szukać adekwatnego statystycznego modelu cyklu komórkowego.

W 1973 r. zaproponowany został przez Smitha i Martina model [15], który uwzględniał znany fakt empiryczny, że łączny czas trwania faz $S + G_2 + M + D$ jest względnie stały w porównaniu z czasem trwania fazy G_1 . Był on oparty na przekonaniu, że o rozpoczęciu syntezy DNA decyduje jakies krytyczne „punktowe” zjawisko zachodzące w fazie G_1 . To krytyczne zjawisko o naturze probabilistycznej dzieli cały cykl na dwa przedziały A i B (por. rys. 1). Czas T



Rys. 1. Relacje między dwoma sposobami podziału cyklu komórkowego (* oznacza punkt krytyczny)

jest więc sumą czasów przebywania komórki w przedziałach A i B :

$$T = T_A + T_B.$$

Dotychczas nie ma przekonującej teorii wyjaśniającej naturę przejścia z fazy A do fazy B , a zatem nie może być mowy o *obserwowaniu* poszczególnych czasów T_A i T_B .

Rozkłady czasów generacji cechuje duża zmienność, która wyklucza na ogół prowadzenie badań na losowo dobranych komórkach. Jednym z propo-

nowanych kierunków badań jest wnioskowanie na podstawie czasów życia komórek *siostrzanych*, gdyż w komórkach tych obserwowana jest tendencja do przechodzenia podziału w cyklach o zbliżonych czasach trwania. Komórki potomne *in vivo* nie różnią się na ogół morfologicznie. Jeśli nawet różnice te istnieją (np. u orzęsków), to zanikają bardzo szybko po podziale. Mamy więc do czynienia z populacją *losowo uporządkowanych* par komórek *siostrzanych*.

3. Formalizacja modelu siostrzanego. Omówimy szczegółowe założenia, które rozpatruje się w ogólnym modelu cyklu komórkowego ($T = T_A + T_B$) dla populacji losowo uporządkowanych par komórek *siostrzanych*.

Oznaczmy przez (X, Y) obserwowaną parę czasów generacji dla losowo uporządkowanych sióstr. Czasy T_A i T_B oznaczmy symbolami X_A i X_B dla pierwszej siostry w losowo uporządkowanej parze, a symbolami Y_A i Y_B dla drugiej siostry. Mamy zatem model *siostrzany*:

$$X = X_A + X_B, \quad Y = Y_A + Y_B.$$

Niektóre założenia o modelu *siostrzanym* dotyczą zmiennej T (co odpowiada założeniom dotyczącym tak zmiennej X , jak i zmiennej Y). Występują też założenia o zmiennej T_A (lub zmiennej T_B), którym odpowiadają założenia o zmiennych X_A i Y_A (lub zmiennych X_B i Y_B).

Do sformułowania jednego z założeń konieczne jest zdefiniowanie tak zwanych *krzywych alfa i beta*, które zajmują dużo miejsca w literaturze biologicznej. Krzywa *alfa* jest wykresem logarytmu funkcji przeżycia zmiennej T (funkcją przeżycia zmiennej losowej nazywamy dopełnienie do 1 dystrybuanty tej zmiennej). Krzywa *beta* jest wykresem logarytmu funkcji przeżycia zmiennej $|X - Y|$. W dalszym ciągu postaramy się wyjaśnić przyczyny, dla których tyle uwagi poświęcono w literaturze relacjom między krzywymi *alfa* i *beta* na podstawie ich odpowiedników z próby (empirycznych krzywych *alfa* i *beta*). Pewną parę empirycznych krzywych *alfa* i *beta* przedstawia rysunek 2.

Przedstawimy teraz założenia modelu *siostrzanego* podzielone na trzy grupy: założenia oparte na materiale doświadczalnym, ogólne założenia teoretyczne oraz szczegółowe założenia teoretyczne.

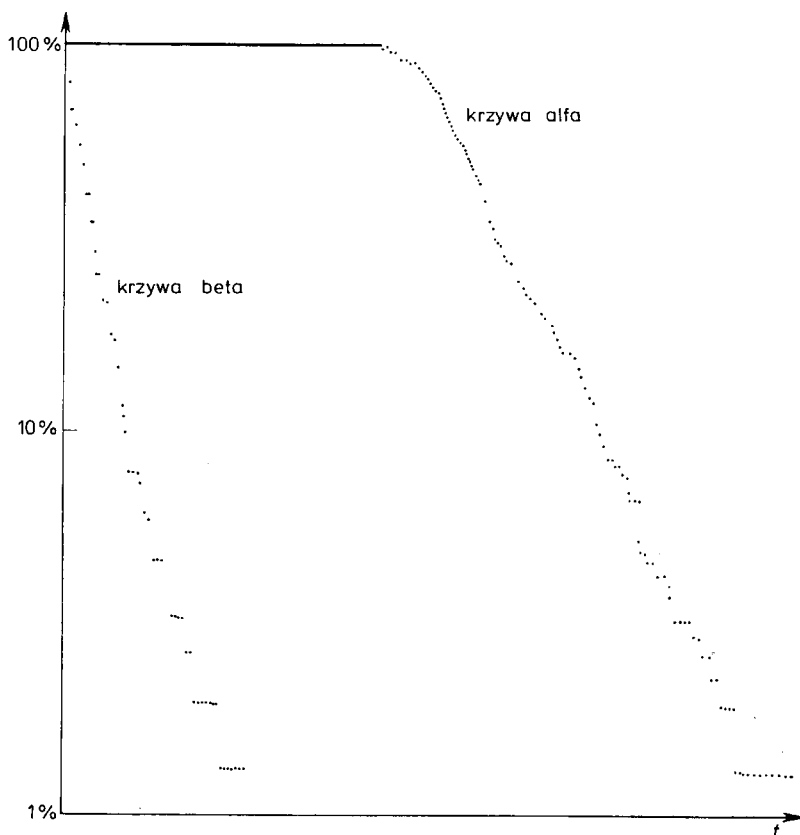
I. Założenia oparte na materiale doświadczalnym:

- 1a) przybliżona liniowość krzywej *beta* oraz przybliżona liniowość krzywej *alfa* dla czasów t większych od pewnego t_0 ;
- 1b) przybliżona równoległość krzywej *beta* i krzywej *alfa* dla $t \geq t_0$;
- 2) duża wartość współczynnika korelacji pomiędzy zmiennymi X i Y ;
- 3) liniowa regresja zmiennej X względem zmiennej Y .

II. Ogólne założenia teoretyczne:

- 1) (X_A, X_B) ma taki sam rozkład jak (Y_A, Y_B) ;
- 2) mała zmienność czasu T_B w porównaniu ze zmiennością czasu T_A :

$$\text{var } T_A \gg \text{var } T_B.$$



Rys. 2. Empiryczne krzywe alfa i beta (próbna 154 par komórek siostrzanych *Chilonella steini*)

III. Szczegółowe założenia teoretyczne występujące w literaturze biologicznej:

- 1) rozkłady par (X_A, Y_A) i (X_B, Y_B) są niezależne;
- 2) T_A ma rozkład wykładniczy z parametrem λ (co oznacza, że intensywność przejścia z fazy A do fazy B jest stała i równa λ);
- 3) $X_B = Y_B$;
- 3') $X_B = Y_B = t_0$ dla pewnej stałej t_0 .

Pierwsza grupa założeń dotyczy tylko danych obserwowalnych, a więc zmiennych T , X i Y .

Pierwsze z nich jest najszerzej dyskutowane w literaturze. Na podstawie materiału doświadczalnego stwierdza się, że wykresy empirycznych krzywych alfa i beta są zwykle zbliżone do przedstawionych na rysunku 2. Stąd wynika sformułowanie założeń I 1a i I 1b.

Duża wartość współczynnika korelacji zmiennych X i Y postulowana w I 2 jest wspomniana przez wielu autorów ([3], [12], [14], [9]).

Liniowość regresji czasu X względem czasu Y (założenie I 3) została

stwierdzona w badaniach orzęsków przez Gołembiewską [3]. (Gołembiewska skorzystała z tego, że w symetrycznych rozkładach dwuwymiarowych dla zmiennych ciągłych regresja jest liniowa wtedy i tylko wtedy, gdy tak zwana funkcja zależności monotonicznej jest tożsamościowo równa współczynnikowi korelacji; opis funkcji zależności monotonicznej i jej własności można znaleźć w [8], [3]. Badanie związków między współczynnikiem korelacji a funkcją zależności monotonicznej jest mniej zawodne niż badanie liniowości regresji (dla prób o małej liczności).)

Z kolei omówimy założenia tworzące grupę II. Pierwsze z nich, orzekające równość rozkładów par (X_A, X_B) i (Y_A, Y_B) , wynika z losowej kolejności sióstr w parze: rozkład czwórki (X_A, X_B, Y_A, Y_B) jest taki sam jak rozkład czwórki (Y_A, Y_B, X_A, X_B) . Założenie II 2 natomiast nawiązuje do małej zmienności faz $S + G_2 + M + D$ w porównaniu z fazą G_1 . Jest to fakt potwierdzony przez badania biologiczne [15]. Musimy jednak dodatkowo założyć (aby otrzymać II 2), że rozpoczęcie syntezy następuje wkrótce po zajściu zjawiska krytycznego. To dodatkowe założenie jest zwykle akceptowane przez biologów.

Połączone założenia grup I i II są naturalnie zbyt ogólne, aby można było na ich podstawie zbudować praktycznie użyteczny model cyklu komórkowego dla komórek siostrzanych. Z tego względu konieczne jest uzupełnienie obu grup dodatkowymi założeniami. W literaturze biologicznej dyskutowano rozmaite założenia szczegółowe. W grupie III zestawione są te spośród nich, które odegrały (i odgrywają nadal) ważną rolę w toczącej się dyskusji.

Założenie III 1 nie jest zwykle explicite formułowane, ale jest przyjmowane — przynajmniej milcząco — przez wszystkich autorów.

Założenie III 2 jest ściśle związane ze wspomnianym poprzednio modelem Smitha–Martina. W modelu tym zakłada się, że prawdopodobieństwo przejścia przez punkt krytyczny w czasie $[t, t + \Delta t)$ pod warunkiem, że długość fazy A jest nie mniejsza od t , jest takie same dla każdego t :

$$(1) \quad P(T_A \in [t, t + \Delta t) | T_A \geq t) = \lambda \Delta t + r(\Delta t),$$

gdzie $\lim_{\Delta t \rightarrow 0} r(\Delta t) / \Delta t = 0$, a λ jest dodatnią stałą zwaną intensywnością przejścia.

Model ten nazywa się *modelem stałego prawdopodobieństwa przejścia*; w literaturze anglojęzycznej używa się nazwy *transition probability model* i skrótu „model TP”. Warunek (1) charakteryzuje rozkład wykładniczy z parametrem λ .

Założenie III 3 nawiązuje do tego, że zmienność czasu T_B jest mała w porównaniu ze zmiennością czasu T_A (założenie II 2); skrajnym przypadkiem jest przyjęcie, że zmienność ta w ogóle nie istnieje. Założenie III 3 jest naturalnym (dla symetrycznego układu siostrzanego) uogólnieniem założenia III 3’.

Z III 3’ wynikają oczywiście III 1 i II 2. W dalszym ciągu rozważa się bądź układ [III 1, 2, 3, II 2], bądź układ [III 2, 3’].

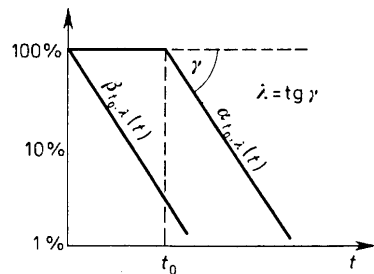
4. Dyskusja założeń szczegółowych. Skonfrontujmy teraz założenia grupy III z założeniami grup I i II. Tego rodzaju próby były w pewnym zakresie podejmowane wcześniej w literaturze biologicznej. Poszczególni autorzy rozpatrywali wybrane założenia szczegółowe i konfrontowali je z niektórymi założeniami z grup I i II; jeśli dochodzono do sprzeczności, to próbowano ją usunąć ad hoc, co polegało głównie na przyjmowaniu dodatkowych założeń szczegółowych (często bez zbadania ich niesprzeczności z poprzednimi założeniami). W literaturze biologicznej istnieje obszerny zapis tej dyskusji. Autorów intrygował przede wszystkim charakter krzywych alfa i beta. Opisywanie tych krzywych, weryfikowanie na ich podstawie słuszności założenia III 2 i ocenianie wartości parametrów λ i t_0 zajmuje w literaturze najwięcej miejsca (por. np. Shields [13] i [14], Smith i Martin [15]).

Przedstawimy najistotniejsze fragmenty tej dyskusji, a następnie pokażemy, że układ założeń [I 1-2, II, III] jest sprzeczny.

Zauważmy najpierw, że przy założeniach III 2 i III 3' funkcja przeżycia zmiennej $T = T_A + t_0$ jest postaci

$$(2) \quad \alpha_{t_0, \lambda}(t) = P(T \geq t) = \begin{cases} 1, & t \leq t_0, \\ e^{-\lambda(t-t_0)}, & t \geq t_0. \end{cases}$$

Wykres tej funkcji przy logarytmicznej skali na osi rzędnych — czyli krzywa alfa — jest więc linią łamaną⁽¹⁾ przedstawioną na rysunku 3.



Rys. 3. Wykresy funkcji $\alpha_{t_0, \lambda}(t)$ i $\beta_{t_0, \lambda}(t)$ dla $\varrho = 0$

Gdyby oprócz III 2 i III 3' przyjąć dodatkowo

Założenie III 4: X_A i Y_A są niezależne,

wtedy funkcja przeżycia zmiennej $|X - Y|$ miałaby postać

$$\beta_{t_0, \lambda}(t) = P_{t_0, \lambda}(|X_A - Y_A| \geq t) = 2 \int \lambda e^{-\lambda y} e^{-\lambda(y+t)} dy = e^{-\lambda t}.$$

Wykres tej funkcji przy logarytmicznej skali na osi rzędnych — czyli krzywa beta — byłby więc półprostą przedstawioną na rysunku 3. Zatem krzywa beta byłaby równoległa do krzywej alfa dla $t \geq t_0$, gdzie t_0 jest stałą z założenia III 3'.

⁽¹⁾ Wprowadzenie skali logarytmicznej przy sporządzaniu empirycznych funkcji przeżycia wynikało właśnie z postaci funkcji przeżycia w modelu TP.

Oznaczmy przez (I 1a)* i (I 1b)* następujące zawężenia założeń I 1a i I 1b:

Założenie (I 1a)*: krzywa beta jest liniowa dla $t \geq 0$, a krzywa alfa jest liniowa dla $t \geq t_0$;

Założenie (I 1b)*: krzywa beta jest równoległa do krzywej alfa dla $t \geq t_0$.

Założenia III 2, III 3' i III 4, rozpatrywane przez Shieldsa w pracy [13] nazwiemy *modelem Shieldsa*. W modelu tym spełnione są założenia (I 1a)* i (I 1b)*. Jednocześnie jednak z modelu Shieldsa wynika, że zmienne X i Y są niezależne, co jest sprzeczne z założeniem I 2. Przedstawia to poniższy schemat:

$$\{\text{model Shieldsa: III 2, III 3', III 4}\} \Rightarrow \left\{ \begin{array}{l} \text{(I 1a)*, (I 1b)*} \\ \sim \text{I 2 (tzn. założenie I 2)} \\ \text{nie jest spełnione} \end{array} \right\}$$

Model Shieldsa wywołał sprzeciw innych badaczy. W szczególności Green [7] zmodyfikował go, zastępując III 3' założeniami III 3 i III 1. W takim modelu zmienne X i Y są dodatnio regresyjnie zależne⁽²⁾ i mogą być silnie skorelowane. Rozkład czwórki (X_A, X_B, Y_A, Y_B) nie jest w modelu Greena określony jednoznacznie, lecz tylko z dokładnością do rozkładu zmiennej X_B . Krzywa beta nie zależy oczywiście od rozkładu zmiennej X_B i pozostaje taka sama jak w modelu Shieldsa; zmienia się natomiast krzywa alfa, gdyż czas T_B nie jest już stały. Jeśli jednak prawdziwe jest założenie II 2, to odstępstwo krzywej alfa od postaci (2) jest nieznaczne, a więc zachodzi postulowana w I 1 przybliżona liniowość i równoległość krzywych alfa i beta.

Z rozważań tych wynika, że model Greena należałoby uzupełnić założeniem II 2. To jednak nie jest możliwe, gdyż następuje wtedy kolizja z założeniem I 2. Z III 4 wynika bowiem, że

$$\text{corr}(X, Y) = \frac{1}{1+w},$$

gdzie $w = \text{var } T_A / \text{var } T_B$; ale na podstawie II 2 $w \gg 1$, a zatem wartość $\text{corr}(X, Y)$ musi być mała, co przeczy założeniu I 2.

Mamy więc

$$\{\text{model Greena: III 1, 2, 3}\} \Rightarrow \{\text{I 2}\},$$

$$\{\text{model Greena uzupełniony założeniem II 2}\} \Rightarrow \{\text{I 1, } \sim \text{I 2}\}.$$

Wydaje się więc, że założenie III 4 trzeba usunąć z proponowanych modeli. Co więcej, w dowolnym modelu czwórki (X_A, X_B, Y_A, Y_B) , przy założeniu III 1 mamy

$$\text{corr}(X, Y) = \frac{w}{w+1} \text{corr}(X_A, Y_A) + \frac{1}{w+1} \text{corr}(X_B, Y_B),$$

⁽²⁾ *Positive regression dependent*, J. Lehmann [10].

a stąd

$$\begin{aligned} \text{corr}(X_A, Y_A) &= \{(w+1)\text{corr}(X, Y) - \text{corr}(X_B, Y_B)\}/w \geq \\ &\geq \{(w+1)\text{corr}(X, Y) - 1\}/w. \end{aligned}$$

Specyfikując dolne ograniczenia dla $\text{corr}(X, Y)$ i w , znajdujemy z tej nierówności dolne ograniczenie dla $\text{corr}(X_A, Y_A)$. Na przykład dla $\text{corr}(X, Y) \geq 0.8$ i $w \geq 2$ otrzymujemy, że $\text{corr}(X_A, Y_A) \geq 0.7$. Zatem

z założeń I 2, II 2 i III 1 wynika istnienie dużej dodatniej korelacji między zmiennymi X_A i Y_A .

Wykażemy teraz, że założeń III 2, III 3' i II 1 nie można tak uzupełnić, aby jednocześnie spełnione były postulaty (I 1a)*, (I 1b)* i I 2.

Przy założeniu III 3' $\text{corr}(X, Y) = \text{corr}(X_A, Y_A)$, a więc z I 2 wynika, że X_A i Y_A muszą być dodatnio skorelowane z dostatecznie dużym współczynnikiem korelacji ϱ . Przy założeniu II 1 i III 2 mamy

$$\begin{aligned} \lambda &= 1/EX_A, \\ EX_A Y_A &= (EX_A)^2 + \varrho D^2 X_A = (1 + \varrho)/\lambda^2, \\ E(X_A - Y_A)^2 &= 2EX_A^2 - 2EX_A Y_A = 2(1 - \varrho)/\lambda^2. \end{aligned}$$

Gdyby natomiast prawdziwe było również założenie (I 1a)*, wtedy

$$(3) \quad E(X_A - Y_A)^2 = \frac{2}{\kappa^2},$$

gdzie $\kappa = 1/E|X_A - Y_A|$. Równość (3) wynika stąd, że przy (I 1a)* zmienna $|X_A - Y_A|$ ma rozkład wykładniczy z pewnym parametrem κ . Zatem

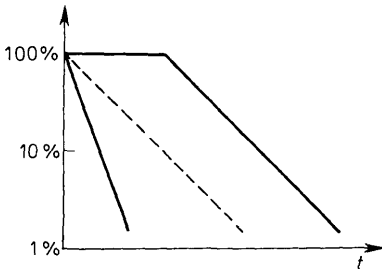
$$\begin{aligned} \frac{2}{\kappa^2} &= \frac{2(1 - \varrho)}{\lambda^2}, \\ (4) \quad \kappa &= \lambda/\sqrt{1 - \varrho}. \end{aligned}$$

Dla $\varrho > 0$ jest więc $\kappa > \lambda$, co oznacza, że krzywe alfa i beta nie są równoległe.

$$\left\{ \begin{array}{l} \text{III 2, III 3', II 1} \\ \text{(I 1a)*, I 2} \end{array} \right\} \Rightarrow \{ \sim \text{(I 1b)*} \}.$$

Przy założeniu I 2 współczynniki kierunkowe κ i λ mogą się dosyć wyraźnie różnić. Na rysunku 4 przedstawiono krzywe alfa i beta przy $\varrho = 0.85$, $\lambda = 1$ i $\kappa = \lambda/\sqrt{1 - \varrho} = 2.58$. Takie różnice między krzywymi nie mogą być uznane za „przybliżoną równoległość”, czyli założenie I 1b nie jest spełnione.

Jeśli w powyższym wywodzie zastąpimy założenie III 3' założeniami III 3, III 1 i II 2, to sytuacja nie ulegnie istotniejszym zmianom. Wynika to z tego, że wobec założenia II 2 para zmiennych (X_B, Y_B) nie ma większego wpływu na

Rys. 4. Krzywe alfa i beta dla $\lambda = 1$ i $\rho = 0.85$

założenia I 1 i I 2. Krzywa beta pozostaje taka sama, jak przy III 3'. Krzywa alfa przestaje oczywiście być linią łamaną, ale przy niewielkich odchyleniach można mówić o przybliżonej liniowości krzywej alfa dla $t > ET_B$. Jeśli przyjmiemy, że założenie I 1a jest spełnione, to założenie I 1b trzeba odrzucić z tych samych przyczyn, na podstawie których odrzucono (I 1b)* przy III 3'.

Układ założeń I 1-2, II 1-2, III 1-3 jest więc sprzeczny. Trzeba zatem usunąć przynajmniej jedno z założeń grup I i II lub odstąpić od modelu TP (założenie III 2).

Rozważmy sytuację powstałą po usunięciu postulatu I 1b, który jest bardzo nieprecyzyjnie sformułowany i nie wydaje się dostatecznie uzasadniony przez empiryczne krzywe alfa i beta (por. rys. 2). Usuwając I 1b, będziemy jednocześnie uwzględniać postulat I 3, który nie był dotąd brany pod uwagę. Jak poprzednio, rozpatrzmy model TP przy założeniu III 3', gdyż — zgodnie z poprzednią dyskusją — przy II 2 zastąpienie III 3' założeniami III 3 i III 1 nie ma dużego wpływu na kształt krzywej alfa i regresji X względem Y .

Przy III 3' założenia III 2, II 1 i I 3 oznaczają, że

Założenie III 5: para (X_A, Y_B) ma symetryczny rozkład wykładniczy z liniową regresją.

Przedstawimy dwie omawiane w literaturze rodziny dwuwymiarowych rozkładów dodatnio skorelowanych i spełniających III 5.

Pierwsza rodzina składa się z rozkładów Morana. Para (X_A, Y_A) ma rozkład Morana z parametrami (ρ, λ) , jeśli

$$X_A = \frac{1}{2\lambda}(\xi_1^2 + \eta_1^2), \quad Y_A = \frac{1}{2\lambda}(\xi_2^2 + \eta_2^2),$$

gdzie pary (ξ_1, ξ_2) i (η_1, η_2) są niezależne i każda z nich ma dwuwymiarowy rozkład normalny $N_2(0, 0, 1, 1, \sqrt{\rho})$. Jak łatwo zauważyć, $\text{corr}(X_A, Y_A) = \rho$ oraz $E(X_A | Y_A = y) = \rho y + (1 - \rho)\lambda^{-1}$. Wykażemy, że zmienna $|X_A - Y_A|$ ma rozkład wykładniczy z parametrem κ .

Niech (X, Y) ma rozkład Morana z parametrami $(\rho, 1)$. Funkcja charakterystyczna pary (X, Y) ma postać [7, str. 268]

$$\varphi_{t_1, t_2} = Ee^{i(t_1 X + t_2 Y)} = ((1 - it_1)(1 - it_2) + \rho t_1 t_2)^{-1},$$

a funkcja charakterystyczna zmiennej $X - Y$ jest dana wzorem

$$(5) \quad \varphi_1(t) = Ee^{it(X-Y)} = (1 + (1-\varrho)t^2)^{-1},$$

a więc φ_1 jest funkcją charakterystyczną dwustronnego rozkładu wykładniczego z parametrem $1/\sqrt{1-\varrho}$. Wynika stąd, że $|X - Y|$ ma rozkład wykładniczy, gdy (X, Y) ma rozkład Morana z dowolnymi parametrami ϱ, λ . Jeśli więc (X_A, Y_A) ma taki rozkład i $X_B = Y_B = t_0$, to spełnione jest założenie (I 1a)*. Zatem

$$\left\{ \begin{array}{l} ((X_A, Y_A) \text{ ma rozkład Morana} \\ \text{z dostatecznie dużym } \varrho) \wedge \text{III } 3' \end{array} \right\} \Rightarrow \{(\text{I } 1a)^*, \text{I } 2, \text{I } 3, \sim \text{I } 1b\}.$$

Z tej implikacji i z poprzednich rozważań o niewielkim wpływie pary (X_B, Y_B) wynika, że układ założeń I 1a, 2, 3, III 1-2, III 1-3 nie jest sprzeczny.

Rozpatrzmy teraz drugą rodzinę dwuwymiarowych rozkładów wykładniczych z dodatnią korelacją, spełniających III 5. Niech (X_A, Y_A) będzie zdefiniowana następująco:

$$(6) \quad X_A = \frac{1}{\lambda}(\xi_0 + \xi_1), \quad Y_A = \frac{1}{\lambda}(\xi_0 + \xi_2),$$

gdzie ξ_0, ξ_1, ξ_2 są niezależnymi zmiennymi losowymi, ξ_0 ma rozkład $\Gamma(1, \varrho)$ (rozkład gamma z parametrem skali równym 1 i parametrem kształtu równym ϱ), a ξ_1 i ξ_2 mają rozkłady $\Gamma(1, 1-\varrho)$. Wtedy

$$\text{corr}(X_A, Y_A) = \varrho, \quad E(X_A | Y_A = y) = \varrho y.$$

Funkcja charakterystyczna pary (X_A, Y_A) jest następująca [7, str. 224]:

$$Ee^{it_1 X_A + it_2 Y_A} = \left(1 - i \frac{(t_1 + t_2)}{\lambda}\right)^{-\varrho} \left(1 - \frac{it_1}{\lambda}\right)^{-(1-\varrho)} \left(1 - \frac{it_2}{\lambda}\right)^{-(1-\varrho)}.$$

Wtedy

$$Ee^{it(X_A - Y_A)} = (1 + (t/\lambda)^2)^{\varrho-1}.$$

Zgodnie z (5) nie jest to funkcja charakterystyczna dwustronnego rozkładu wykładniczego, ale przy $\varrho \rightarrow 0$ dąży do funkcji charakterystycznej tego rozkładu. Zatem przy $\varrho \rightarrow 0$ rozkład zmiennej $|X_A - Y_A|$ jest zbieżny do wykładniczego. Wynika stąd, że założenie I 1a może być spełnione tylko przy małej korelacji między X_A i Y_A , a wtedy nie jest spełnione założenie I 2. Mamy zatem

$$\left\{ \begin{array}{l} ((X_A, Y_A) \text{ jest elementem rodziny (6)} \\ \text{z dostatecznie małym } \varrho) \wedge \text{III } 3' \end{array} \right\} \Rightarrow \{\text{I } 1a, \text{I } 3, \sim \text{I } 2, \sim \text{I } 1b\},$$

$$\left\{ \begin{array}{l} ((X_A, Y_A) \text{ jest elementem rodziny (6)} \\ \text{z dostatecznie dużym } \varrho) \wedge \text{III } 3' \end{array} \right\} \Rightarrow \{\text{I } 2, \text{I } 3, \sim \text{I } 1a, \sim \text{I } 1b\}.$$

Z tych implikacji widać, że przy założeniach I 1a, I 2-3, II, III rozpatrywana rodzina rozkładów nie nadaje się do modelowania rozkładu pary (X_A, Y_A) .

5. Uwagi końcowe. Rozważania przedstawione w § 4 stanowią pewien etap prac nad weryfikacją modelu TP, które są obecnie prowadzone przez autorkę niniejszej pracy. Z rozważań tych wynika, że nawet czysto formalna dyskusja trzech grup założeń dotyczących modelu dla czwórki zmiennych (X_A, Y_A, X_B, Y_B) stawia pod znakiem zapytania szeroko znany i uznawany przez biologów model TP. Użyteczność dyskusji jest jednak stosunkowo mała ze względu na nieprecyzyjne formułowanie założeń grupy I dotyczących obserwowalnych danych, a zwłaszcza założenia I 1. Dochodzimy w ten sposób do jądra problemu, jakim jest identyfikacja modelu na podstawie materiału doświadczalnego. Narzędzia statystyczne służące do tego celu nie są zbyt przekonujące. Jak wiadomo, testy istotności pozwalają tylko odrzucać weryfikowany model lub zawieszać decyzję, a przy tym ich własności są wyraźnie uzależnione od wielkości próby. Testy takie są często oparte na estymatorach parametrów weryfikowanego modelu. W praktyce statystycznej zdarza się, że jeśli test istotności nie dał podstaw do odrzucenia modelu, przyjmuje się ów model za zweryfikowany, a wyniki estymacji traktuje jako estymowane wartości jego parametrów. Jeszcze bardziej wątpliwa jest identyfikacja omówionych tutaj modeli, polegająca na subiektywnym ocenianiu własności teoretycznych krzywych alfa i beta na podstawie ich empirycznych odpowiedników, a taka identyfikacja modeli zajmuje w literaturze biologicznej bardzo dużo miejsca.

Przy formułowaniu modeli trzeba koniecznie rozpatrywać rozkłady całej czwórki (X_A, X_B, Y_A, Y_B) , starannie specyfikując założenia grup I i II.

Sądzymy, że wśród założeń szczegółowych należy dopuścić do rozważań prawostronnie skośne rozkłady czasu T_A inne niż wykładnicze. Należy do nich na przykład rozkład gamma, który może być użyteczny wtedy, gdy w fazie G_1 wyróżnia się dwa lub więcej zjawisk decydujących o rozpoczęciu syntezy DNA. Istnieje rodzina dodatnio skorelowanych dwuwymiarowych rozkładów gamma z liniową regresją opisaną w [7], którą warto byłoby zbadać w modelu siostrzanym cyklu komórkowego.

Parametryczne modele siostrzane cyklu komórkowego są konstruowane głównie w tym celu, żeby dopomóc przy ocenie wpływu wybranego czynnika oddziałującego na populację komórek. Ocena taka ma służyć do badania mechanizmów rządzących cyklem komórkowym. Przy założeniu, że cykl składa się z przedziału A i przedziału B i że obserwujemy czas T , oceniane są zmiany zachodzące pod wpływem czynnika w rozkładzie pary (T_A, T_B) . Oznaczmy symbolami T_A^+ i T_B^+ czasy przebywania komórki w przedziałach A i B w populacji komórek poddanych wpływowi czynnika. Po przyjęciu założeń o łącznych rozkładach pary (T_A, T_B) i pary (T_A^+, T_B^+) , należy ocenić wpływ czynnika, porównując wybrane parametry tych rozkładów, a następnie zapro-

jektować doświadczenia umożliwiające estymację obranych parametrów przy istniejących ograniczeniach obserwowalności.

Wybór odpowiednich parametrów odgrywa ogromną rolę. W praktyce oprócz wpływu badanego czynnika zaznacza się również wpływ wielu innych zakłócających czynników, który trudno jest wyeliminować. Jeżeli nie można tego zmienić przez odpowiednie zaplanowanie doświadczenia, trzeba znaleźć takie parametry rozkładu obserwowalnych danych, które nie są wrażliwe na wpływ niekontrolowanych czynników. Zagadnienie to przedstawiono nieco szerzej w [4].

Bibliografia

- [1] Laroy N. Castor, *A G_1 rate model accounts for cell-cycle kinetics attributed to "transition probability"*, Nature 287 (1980), 857-859.
- [2] D. S. Czernawski, J. M. Romanowski, N. W. Stiepanowa, *Biofizyka matematyczna*, Nauka, Moskwa 1984 (w języku rosyjskim).
- [3] J. Ćwik, M. Gołembiewska, T. Kowalczyk, E. Pleszczyńska, *Conceptual and statistical problems of sister dependence*, Biometrika 69 (1982), 513-20.
- [4] J. Ćwik, M. Gołembiewska, *Badanie komórek siostrzanych w: T. Bromek, E. Pleszczyńska (red.), Wnioskowanie statystyczne w teorii i praktyce*, PWN, Warszawa 1988.
- [5] M. Gołembiewska-Skoczylas, S. Radzikowski, *Relation between macronuclear DNA and total protein content and generation time in the *Chilodonella steini* (Ciliata) sister cells*, Protozoologia 129 (1985), 44-51.
- [6] P. J. Green, *A "random transition" in the cell cycle?* Nature 285 (180), 116.
- [7] N. L. Johnson, S. Kotz, *Distributions in statistics: Continuous multivariate distributions*, J. Wiley, Inc., New York 1972.
- [8] T. Kowalczyk, E. Pleszczyńska, *Monotonic dependence functions of bivariate distributions*, Ann. Statist. 5 (1977), 1221-1227.
- [9] A. Lasota, M. C. Mackey, *Globally asymptotic properties of proliferating cell populations*, J. Math. Biology 19 (1984), 43-62.
- [10] E. Lehmann, *Some concepts of dependence*, Ann. Math. Statist. 37 (1966), 1137-1153.
- [11] K. V. Mardia, *Families of Bivariate Distributions*, Griffin, London 1970.
- [12] P. D. Minor, J. A. Smith, *Explanation of degree of correlation of sibling generation times in animal cells*, Nature 248 (1974), 241-243.
- [13] R. Shields, *Transition probability and the origin of variation in the cell cycle*, Nature 267 (1977), 704-707.
- [14] R. Shields, *Further evidence for a random transition in the cell cycle*, Nature 273 (1978), 755-758.
- [15] J. A. Smith, L. Martin, *Do cells cycle?*, Proc. Nat. Acad. Sci. USA 70 (1973), 1263-1267.
- [16] J. A. Smith, D. J. R. Laurence, P. S. Rudland, *Limitations of cell kinetics in distinguishing cell cycle models*, Nature 293 (1981), 648-650.
- [17] E. J. J. van Zoelen, P. T. van der Saag, S. W. de Laat, *Sources of variation in intermitotic time and their relation to cell cycle kinetics*, J. Theor. Biol. 93 (1981), 1037-1045.