ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria III: MATEMATYKA STOSOWANA XX (1982)

J. Holzheimer (Wrocław)

Estymacja parametrów procesu Galtona–Watsona

(*Praca przyjęta do druku 18.04.1980*)

0. Wstęp. Praca stanowi próbę przeglądu estymatorów parametrów najprostszego procesu gałązkowego — jednowymiarowego procesu Galtona–Watsona. Część pierwsza pracy zawiera opis matematyczny procesu oraz jego podstawowe własności. W części drugiej rozpatrywane są znane estymatory średniej w rozkładzie prawdopodobieństwa kreacji; część trzecia jest poświęcona estymacji prawdopodobieństw kreacji, a w części czwartej rozważa się problem estymacji wariancji.

Ze względu na odmienne zachowanie się procesu w zależności od średniej w rozkładzie prawdopodobieństwa kreacji zagadnienie estymacji parametrów rozpatruje się oddzielnie dla podkrytycznego, krytycznego i nadkrytycznego procesu Galtona–Watsona.

1. Proces Galtona–Watsona. Niech $Z = (Z_n;\ n = 0, 1, \ldots)$, $Z_0 = 1$, będzie ciągiem zmiennych losowych o wartościach całkowitych nieujemnych, stanowiącym proces Galtona–Watsona, tzn.

$$Z_{n+1} = \zeta_1^{(n)} + \zeta_2^{(n)} + \ldots + \zeta_{Z_n}^{(n)}, \qquad n \geqslant 0,$$

$$P(Z_1 = k) = p_k, \qquad p_k \geqslant 0, \qquad \sum_{k=0}^{\infty} p_k = 1,$$

gdzie $\zeta_i^{(n)}$, $i = 1, 2, \ldots, Z_n$, są zmiennymi losowymi wzajemnie niezależnymi, o rozkładzie Z_1, niezależnymi od Z_n. $Z_{n+1} = 0$, gdy $Z_n = 0$.

W dalszych częściach pracy będziemy się często odwoływali do popularnej interpretacji procesu Galtona–Watsona. Rozważmy populację indywiduów zdolnych do rozszczepiania się na indywidua tego samego typu. Załóżmy, że populacja indywiduów w chwili 0 składa się z Z_0 indywiduów tworzących początkowe (zerowe) pokolenie. Każde indywiduum danej populacji jest tego samego typu oraz rozszczepia się niezależnie od pozostałych indywiduów zgodnie z rozkładem prawdopodobieństwa $(p_k;\ k = 0, 1, \ldots)$ nazywanym *rozkładem prawdopodobieństwa kreacji*. Potomkowie indywiduów pokolenia początkowego tworzą pierwsze pokolenie procesu Galtona–Watsona. Indywidua każdego pokolenia rozszczepiają się niezależnie od liczebności poprzednich pokoleń.

Przyjmijmy następujące oznaczenia dla wartości oczekiwanej i wariancji liczby bezpośrednich potomków danego indywiduum:

$$m = EZ_1, \quad \sigma^2 = \mathrm{Var}(Z_1).$$

Dalej będziemy zakładali, że $0 < \sigma^2 < +\infty$.

Proces Galtona–Watsona Z nazywamy *podkrytycznym*, gdy $m < 1$, *krytycznym*, gdy $m = 1$, *nadkrytycznym*, gdy $m > 1$. Rozróżnienie tych trzech przypadków jest istotne przy rozważaniu asymptotyki procesu. Jeśli $q = P(\lim_{n\to+\infty} Z_n = 0)$ oznacza liczbę zwaną *prawdopodobieństwem wymarcia procesu*, to $q = 1$, gdy $m \leqslant 1$, natomiast $q < 1$, gdy $m > 1$.

Zachowanie się procesu Z zależy od parametru m, a więc jednym z podstawowych zadań statystyki procesów gałązkowych jest określenie dobrego (w pewnym sensie) estymatora tego parametru.

W przypadku nadkrytycznym nasz proces z prawdopodobieństwem 1 (z P.1) osiąga 0 dla dostatecznie dużych n lub wzrasta do $+\infty$, gdy $n \to +\infty$. Ta własność niestabilności procesu wydaje się być sprzeczna z zachowaniem się populacji biologicznych, które często dążą do stanu równowagi probabilistycznej. Z tego powodu rozważany model bez modyfikacji raczej nie nadaje się do opisu zjawisk występujących w populacjach biologicznych. Niemniej jednak statystyka tego procesu jest interesującym zagadnieniem, a przy tym może być punktem wyjścia do rozważania modeli zmodyfikowanych, jak na przykład procesu Galtona–Watsona z imigracją i procesu gałązkowego w losowym środowisku.

Rozważmy pytanie, przy jakich założeniach istnieje taki ciąg $C = (C_n; n = 0, 1, \ldots)$ stałych, że zachodzi zbieżność

$$\frac{Z_n}{C_n} \xrightarrow{P} W,$$

gdzie symbol $\xrightarrow{P}$ oznacza zbieżność wg prawdopodobieństwa, a W niezdegenerowaną zmienną losową. C. C. Heyde rozwiązał ten problem pokazując (zob. [1] lub [11]), że zawsze istnieje taki ciąg C, $C_n \to +\infty$; $C_{n+1}/C_n \to m$, gdy $n \to +\infty$, że zmienne losowe $W_n = Z_n/C_n$ dążą z P.1 do zmiennej losowej W z $P(W > 0) = 1-q$. Jeżeli $EZ_1 \log Z_1 < +\infty$, to możemy wziąć $C_n = m^n$, w przeciwnym razie, jeżeli $EZ_1 \log Z_1 = +\infty$, to $P(W = 0) = 1$, czyli zmienna losowa W jest zdegenerowana.

Wprowadźmy oznaczenie $S_n = Z_0 + Z_1 + \ldots + Z_n$. C. C. Heyde pokazał również, że

$$\frac{S_n}{1+m+\ldots+m^n} \to W \quad \text{z P.1.}$$

A. Badalbajew (zob. [2]) rozważał podwójnie indeksowany ciąg zmiennych losowych $W_n^{(r)} = Z_n^{(r)}/rm^n$, gdzie $(Z_n^{(r)}; n = 0, 1, \ldots)$ jest procesem Galtona–Watsona, dla którego $Z_0^{(r)} = r$. Udowodnił on następujące

TWIERDZENIE 1.1. (I) *Jeżeli $m < 1$ oraz $r, n \to +\infty$ w taki sposób, że $rm^n \to +\infty$, to*

$$\frac{Z_n^{(r)}}{rm^n} \xrightarrow{P} 1;$$

(II) *jeżeli $m = 1$ oraz $r, n \to +\infty$ w taki sposób, że $r/n \to +\infty$, to*

$$\frac{Z_n^{(r)}}{r} \xrightarrow{P} 1;$$

(III) *jeżeli $m > 1$ oraz $r \to +\infty$, to*

$$\frac{Z_n^{(r)}}{rm^n} \xrightarrow{P} 1 \quad \textit{dla każdego } n.$$

N. M. Yanev (zob. [16]) rozważył również zbieżność zmiennych losowych $Z_n^{(r)}/r$ w przypadkach, gdy $r/n \to 0$ oraz gdy $n/r \to K$, $0 < K < +\infty$.

2. Estymator Lotki–Nagajewa. Rozważmy estamator średniej m oparty na próbie $(Z_0 = 1, Z_1, \ldots, Z_{n+1})$, zdefiniowany w następujący sposób:

$$\overline{m} = \begin{cases} \dfrac{Z_{n+1}}{Z_n}, & Z_n > 0, \\ 1 & \text{w przeciwnym razie}. \end{cases}$$

Estymator ten był badany przez A. J. Lotkę oraz A. V. Nagajewa (zob. [14]). Własności asymptotyczne tego estymatora badał W. Bühler (zob. [3]). Pokazał on, że w przypadku nadkrytycznym

$$P\big(\sigma^{-1}(Z_n)^{1/2}(\overline{m}-m) \leqslant x \,|\, Z_n > 0\big) \to \Phi(x),$$

gdzie $\Phi(x)$ jest dystrybuantą rozkładu normalnego ze średnią 0 i wariancją 1.

Zbiór $A = (Z_n > 0, n = 0, 1, \ldots)$ nazywamy *zbiorem niepochłaniania procesu* Z. Ponieważ $P(A) = 1-q$, więc miara probabilistyczna $P_A(\cdot) = P(\cdot\,|A)$ ma sens jedynie w przypadku nadkrytycznym. J. P. Dion uogólnił wynik W. Bühlera rozważając dowolne miary probabilistyczne Q absolutnie ciągłe względem miary P_A ($Q \ll P_A$). Wprowadzenie miary Q jest próbą osłabienia założenia o niezależności rozszczepiania się indywiduów tej samej generacji. Oczywiście, założenie $Q \ll P_A$ jest założeniem bardzo mocnym, ale pozwala ono rozważać populacje biologiczne, w których występuje zależność między rozszczepianiem indywiduów tej samej generacji. Rezultat J. P. Diona jest następujący:

TWIERDZENIE 2.1. *Dla nadkrytycznego procesu Galtona–Watsona Z zachodzi*

$$Q\big(\sigma^{-1}(Z_n)^{1/2}(\overline{m}-m) \leqslant x\big) \to \Phi(x),$$

$$Q\big(\sigma^{-1}m^{n/2}(\overline{m}-m) \leqslant x\big) \to \int_0^\infty \Phi\big(x\sqrt{y}\big)\,dS(y),$$

gdzie $S(y) = P_A(W < y)$, a W jest zmienną losową będącą granicą z P.1 *zmiennych losowych Z_n/m^n.*

J. P. Dion znalazł również przedział ufności dla m.

Można ponadto wykazać, że w przypadku nadkrytycznym $\overline{m} \to m$ z P.1. na zbiorze A. Wynika to z rezultatu C. C. Heyde (zob. [11]). K. S. Crump i R. B. Howe wykazali, że $E(\overline{m}|A) \to m$, gdy $n \to +\infty$. Oznacza to, że estymator $\overline{m}$ jest asymptotycznie nieobciążony na zbiorze A. Podane własności estymatora Lotki–Nagajewa dotyczą jedynie nadkrytycznego procesu Galtona–Watsona. A. Nagajew (zob. [14]) rozważał również asymptotykę $\overline{m}$ w przypadku podkrytycznym i krytycznym pod warunkiem $Z_n > 0$.

Jest on określony jedynie na zbiorze niepochłaniania procesu, a więc jest niewygodny do praktycznych celów. Jak już zauważyliśmy poprzednio, można badać asymptotykę procesu Galtona–Watsona przy założeniu, że liczba indywiduów pokolenia zerowego r dąży do nieskończoności lub gdy zarówno r, jak i n dążą do nieskończoności. Takie podejście pozwala nie ograniczać się do zbioru $(Z_n > 0)$.

A. Badalbajew w [2] rozważał estymator m postaci

$$\overline{m}_r = \frac{Z_{n+1}^{(r)}}{1+Z_n^{(r)}},$$

przy czym wykazał, że jest on mocno zgodny, tzn. $\overline{m}_r \to m$ z P.1., gdy $r \to +\infty$, oraz badał asymptotykę tego estymatora.

TWIERDZENIE 2.2. (I) *Jeżeli* $m < 1$ *oraz* $r, n \to +\infty$ *w taki sposób, że* $rm^n \to +\infty$, *to*

$$P\big(\sigma^{-1}(rm^n)^{1/2}(\overline{m}_r - m) \leqslant x\big) \to \Phi(x);$$

(II) *jeżeli* $m = 1$ *oraz* $r, n \to +\infty$ *w taki sposób, że* $r/n \to +\infty$, *to*

$$P\big(\sigma^{-1}(r)^{1/2}(\overline{m}_r - m) \leqslant x\big) \to \Phi(x);$$

(III) *jeżeli* $m > 1$ *oraz* $r \to +\infty$, *to*

$$P\big(\sigma^{-1}(rm^n)^{1/2}(\overline{m}_r - m) \leqslant x\big) \to \Phi(x)$$

dla każdego n.

Mocna zgodność estymatora $\overline{m}$ wynika z faktu, że $Z_n^{(r)}$ daje się przedstawić w postaci

$$Z_n^{(r)} = Z_{n,1} + Z_{n,2} + \ldots + Z_{n,r},$$

gdzie $Z_{n,i}$ jest liczbą indywiduów n-tego pokolenia, będących potomkami i-tego indywiduum pokolenia zerowego. Zmienne losowe $(Z_{n,i};\ i = 1, 2, \ldots, r)$ są niezależne o rozkładzie takim jak Z_n.

Twierdzenia 2.2 dowodzi się standardowymi metodami (zob. [14] lub [16]). Można również badać asymptotykę estymatora $\overline{m}_r$, gdy $r \to +\infty$.

2.1. *Estymator Harrisa.* T. Harris w pracy [10] udowodnił, że

$$\hat{m} = \frac{Z_1 + Z_2 + \ldots + Z_{n+1}}{Z_0 + Z_1 + \ldots + Z_n}$$

jest estymatorem największej wiarogodności parametru m opartym na próbie (Z_{jk}; $j = 0, 1, \ldots, n+1, k = 0, 1, \ldots$), gdzie Z_{jk} oznacza liczbę indywiduów j-ej generacji mających k bezpośrednich potomków. Tamże Harris wykazał, że $\hat{m} \to m$ z P.1 na zbiorze A. P. D. Feigin w pracy [9] udowodnił, że estymator ten jest również estymatorem największej wiarogodności opartym na próbie ($Z_0 = 1, Z_1, \ldots, Z_{n+1}$). J. P. Dion udowodnił ponadto asymptotyczną normalność $\hat{m}$ oraz podał przedział ufności dla m.

TWIERDZENIE 2.3. *Dla nadkrytycznego procesu Galtona–Watsona Z i dowolnej miary probabilistycznej $Q \ll P_A$ mamy*

$$Q\left(\sigma^{-1}(S_n)^{1/2}(\hat{m}-m) \leqslant x\right) \to \Phi(x),$$

$$Q\left(\sigma^{-1}(1+m+\ldots+m^n)^{1/2}(\hat{m}-m) \leqslant x\right) \to \int_0^\infty \Phi(x\sqrt{y})\,dS(y).$$

U w a g a. J. Scott w pracy [15] podał pewne centralne twierdzenie graniczne dla martyngałów, z którego jako wniosek otrzymuje przedstawione wyżej własności asymptotyczne estymatorów Lotki–Nagajewa i Harrisa.

Dowód twierdzenia 2.3 wynika z centralnego twierdzenia granicznego dla sum niezależnych zmiennych losowych o losowej liczbie składników (zob. [5], tw. 17.2), ze zbieżności

$$\frac{S_n}{1+m+\ldots+m^n} \to W \quad \text{z} \quad \text{P.1}$$

oraz z równości

$$\sigma^{-1}(S_n)^{1/2}(\hat{m}-m) = \sigma^{-1}(S_n)^{-1/2}\sum_{i=1}^{S_n}(\zeta_i - m),$$

gdzie (ζ_i; $i = 1, 2, \ldots$) jest ciągiem niezależnych zmiennych losowych o rozkładzie Z_1. Rozważanie estymatora $\hat{m}$ na zbiorze A ma sens jedynie wtedy, gdy $m > 1$. W przypadku podkrytycznym i krytycznym interesujące byłoby rozważenie estymatora $\hat{m}$ pod warunkiem $Z_n > 0$. N. M. Yanev w pracy [16] badał asymptotykę estymatora $\hat{m}$ w przypadku, gdy $r \to +\infty$ dla ustalonego n oraz gdy $r, n \to +\infty$.

TWIERDZENIE 2.4. *Jeżeli $r \to +\infty$ dla ustalonego n, to*

(I) $\hat{m} \to m$ *z* P.1, $E\hat{m} \to m$, *oraz*

(II) $\hat{m}$ *jest asymptotycznie normalny, tzn.*

$$P\left(\sigma^{-1}\left(r(1+m+\ldots+m^n)\right)^{1/2}(\hat{m}-m) \leqslant x\right) \to \Phi(x).$$

TWIERDZENIE 2.5. *Jeżeli $r, n \to +\infty$, to*

(I) $\hat{m} \xrightarrow{P} m$, $E\hat{m} \to m$,

(II) *jeżeli $m < 1$, to*

$$P\left(\sigma^{-1}\left(\frac{r}{1-m}\right)^{1/2}(\hat{m}-m) \leqslant x\right) \to \Phi(x),$$

(III) *jeżeli* $m = 1$ *oraz* $r/n \to +\infty$, *to*

$$P\big(\sigma^{-1}(rn)^{1/2}(\hat{m}-m) \leqslant x\big) \to \Phi(x),$$

(IV) *jeżeli* $m > 1$ *oraz* $E(Z_1^4) < +\infty$, *to*

$$P\left(\sigma^{-1}\left(\frac{r(m^{n+1}-1)}{m-1}\right)^{1/2}(\hat{m}-m) \leqslant x\right) \to \Phi(x).$$

2.2. *Inne estymatory średniej* m. C. C. Heyde i J. Leslie w pracy [13] zaproponowali estymator średniej m oparty jedynie na liczebności n-tej generacji Z_n. Estymator ten ma postać

$$m^* = (Z_n)^{1/n}.$$

W przypadku nadkrytycznym estymator ten jest mocno zgodny na zbiorze A oraz zachodzi zbieżność

$$n(m^*-m) \to m \log W \quad \text{z } \mathrm{P}_A.1.$$

3. Estymacja prawdopodobieństwa kreacji. Rozważając próbę $(Z_{jk}; j = 0, 1, \ldots, n+1, k = 0, 1, \ldots)$, T. Harris otrzymał estymatory największej wiarogodności dla prawdopodobieństw kreacji następującej postaci

$$\hat{p}_k = \frac{Z_{0k}+Z_{1k}+ \ldots +Z_{nk}}{S_n}, \qquad k = 0, 1, \ldots$$

J. Dion w pracy [7] udowodnił asymptotyczną normalność tych estymatorów.

TWIERDZENIE 3.1. *Dla nadkrytycznego procesu Galtona–Watsona* Z *oraz dla dowolnej miary probabilistycznej* $Q \ll P_A$ *zachodzi*

$$Q\left(\left(\frac{S_n}{p_k(1-p_k)}\right)^{1/2}(\hat{p}_k-p_k) \leqslant x\right) \to \Phi(x),$$

$$Q\left(\left(\frac{1+m+ \ldots +m^n}{p_k(1-p_k)}\right)^{1/2}(\hat{p}_k-p_k) \leqslant x\right) \to \int_0^\infty \Phi(x\sqrt{y})\,dS(y).$$

Uwaga. Dion znalazł również przedział ufności dla danego p_k. Ponieważ nie jest znany rozkład łączny dla estymatorów prawdopodobieństw p_k, $k = 0, 1, \ldots$, więc podanie łącznego przedziału ufności, tzn. dla wszystkich prawdopodobieństw p_k, nie może być brane pod uwagę.

Podobnie jak w przypadku estymatorów $\overline{m}_r$ i $\hat{m}$, można badać asymptotykę estymatorów $\hat{p}_k$, gdy $r \to +\infty$ oraz $r, n \to +\infty$. Prawdziwe są następujące twierdzenia.

TWIERDZENIE 3.2. *Jeżeli* $r \to +\infty$ *dla ustalonego* n, *to*

(I) $\hat{p}_k \to p_k$ *z* P.1 *oraz* $E\hat{p}_k \to p_k$, $k = 0, 1, \ldots$,

(II) $P\left(\left(\frac{r(1+m+ \ldots m^n)}{p_k(1-p_k)}\right)^{1/2}(\hat{p}_k-p_k) \leqslant x\right) \to \Phi(x).$

TWIERDZENIE 3.3. *Jeżeli* $r, n \to +\infty$, *to*:

(I) $\hat{p}_k \xrightarrow{P} p_k$, $E\hat{p}_k \to p_k$, $k = 0, 1, \ldots$,

(II) *jeżeli* $m < 1$, *to*

$$P\left(\left(\frac{r}{p_k(1-p_k)(1-m)}\right)^{1/2}(\hat{p}_k - p_k) \leqslant x\right) \to \Phi(x),$$

(III) *jeżeli* $m = 1$ *oraz* $r/n \to +\infty$, *to*

$$P\left(\left(\frac{rn}{p_k(1-p_k)}\right)^{1/2}(\hat{p}_k - p_k) \leqslant x\right) \to \Phi(x),$$

(IV) *jeżeli* $m > 1$ *oraz* $E(Z_1^4) < +\infty$, *to*

$$P\left(\left(\frac{r(m^{n+1}-1)}{p_k(1-p_k)(m-1)}\right)^{1/2}(\hat{p}_k - p_k) \leqslant x\right) \to \Phi(x).$$

Dowody tych twierdzeń są standardowe; wymagają jedynie pewnych modyfikacji dowodu N. M. Yaneva (zob. [16]) dla $\hat{m}$. Wykorzystuje się m.in. fakt, że

$$Z_{0k}^{(r)} + Z_{1k}^{(r)} + \ldots + Z_{nk}^{(r)} = \sum_{j=1}^{S_n^{(r)}} T_j^{(k)},$$

gdzie $(T_j^{(k)}; j = 1, 2, \ldots)$, $k = 0, 1, \ldots$, jest ciągiem zmiennych losowych zdefiniowanych w następujący sposób:

$$T_j^{(k)} = \begin{cases} 1, & \text{gdy } \zeta_j = k, \\ 0 & \text{w przeciwnym razie,} \end{cases}$$

a $S_n^{(r)} = Z_0^{(r)} + Z_1^{(r)} + \ldots + Z_n^{(r)}$; $(\zeta_j; j = 1, 2, \ldots)$ jest ciągiem niezależnych zmiennych losowych o rozkładzie Z_1.

Można rozważyć estymatory prawdopodobieństw kreacji oparte na próbie $(Z_{nk}; k = 0, 1, \ldots)$ postaci

$$\bar{p}_k = \frac{Z_{nk}}{Z_n}, \qquad k = 0, 1, \ldots.$$

Estymatory te są asymptotycznie normalne.

TWIERDZENIE 3.4. *Dla nadkrytycznego procesu Galtona–Watsona* Z *oraz dla dowolnej miary probabilistycznej* $Q \ll P_A$ *mamy*

$$Q\left(\left(\frac{Z_n}{p_k(1-p_k)}\right)^{1/2}(\bar{p}_k - p_k) \leqslant x\right) \to \Phi(x),$$

$$Q\left(\left(\frac{m^n}{p_k(1-p_k)}\right)^{1/2}(\bar{p}_k - p_k) \leqslant x\right) \to \int_0^\infty \Phi(x\sqrt{y})\, dS(y).$$

W dowodzie korzystamy z faktu, że Z_{nk} daje się przedstawić w postaci $Z_{nk} = \sum_{j=1}^{Z_n} T_j^{(k)}$. Teza twierdzenia wynika z równości

$$\frac{(Z_n)^{1/2}}{(\operatorname{Var}(T_j^{(k)}))^{1/2}}(\bar{p}_k - p_k) = \frac{Z_n^{-1/2}}{(\operatorname{Var}(T_j^{(k)}))^{1/2}} \sum_{j=1}^{Z_n} (T_j^{(k)} - p_k)$$

oraz z centralnego twierdzenia granicznego dla sum niezależnych zmiennych losowych o losowej liczbie składników (zob. [5], tw. 12.2).

Rozważmy próbę ($Z_{jk}^{(r)}$; $j = 0, 1, \ldots, n$, $k = 0, 1, \ldots$), gdzie $Z_{jk}^{(r)}$ jest liczbą indywiduów j-tego pokolenia mających k bezpośrednich potomków, przy założeniu, że $Z_0^{(r)} = r$. Można zaproponować estymatory prawdopodobieństw kreacji wykorzystując znajomość n-tego pokolenia procesu

$$\bar{p}_k^{(r)} = \frac{Z_{nk}^{(r)}}{1 + Z_n^{(r)}}, \qquad k = 0, 1, \ldots$$

Estymatory te są mocno zgodne, tzn. $\bar{p}_k^{(r)} \to p_k$ z P.1, gdy $r \to +\infty$ dla $k = 0, 1, \ldots$ Ponadto zachodzi

TWIERDZENIE 3.5.

(I) *Jeżeli* $m < 1$ *oraz* $r, n \to +\infty$ *w taki sposób, że* $rm^n \to +\infty$, *to*

$$P\left(\left(\frac{rm^n}{p_k(1-p_k)}\right)^{1/2} (\bar{p}_k^{(r)} - p_k) \leqslant x\right) \to \Phi(x),$$

(II) *jeżeli* $m = 1$ *oraz* $r, n \to +\infty$ *w taki sposób, że* $r/n \to +\infty$, *to*

$$P\left(\left(\frac{r}{p_k(1-p_k)}\right)^{1/2} (\bar{p}_k^{(r)} - p_k) \leqslant x\right) \to \Phi(x),$$

(III) *jeżeli* $m > 1$ *oraz* $r \to +\infty$, *to*

$$P\left(\left(\frac{rm^n}{p_k(1-p_k)}\right)^{1/2} (\bar{p}_k^{(r)} - p_k) \leqslant x\right) \to \Phi(x),$$

dla każdego n.

W dowodzie korzystamy z równości $Z_{nk}^{(r)} = \sum_{j=1}^{Z_n^{(r)}} T_j^{(k)}$, gdzie $(T_j^{(k)})$ jest ciągiem zmiennych losowych zdefiniowanym jak w dowodzie twierdzenia 3.2. Mocna zgodność estymatora wynika z równości

$$Z_{nk}^{(r)} = Z_{nk,1} + Z_{nk,2} + \ldots + Z_{nk,r},$$

gdzie ($Z_{nk,i}$; $i = 1, 2, \ldots, r$) jest ciągiem niezależnych zmiennych losowych o rozkładzie takim, jak Z_{nk}, oraz stąd, że

$$E(Z_{nk}) = E\left\{E\left(\sum_{j=1}^{Z_n} T_j^{(k)} | Z_n\right)\right\} = p_k m^n.$$

4. Estymacja wariancji.

4.1. *Estymacja wariancji przy znanej średniej.* Niech

$$\tau_k = \left(\frac{Z_{k+1}}{Z_k} - m\right)^2 Z_k, \qquad k = 0, 1, \ldots$$

J. P. Dion w [8] zaproponował estymator wariancji σ^2 postaci

$$\bar{\sigma}_n^2 = \frac{1}{n+1} \sum_{k=0}^{n} \tau_k,$$

oraz podał jego podstawowe własności.

TWIERDZENIE 4.1. *Niech Z będzie nadkrytycznym procesem Galtona–Watsona, dla którego $p_0 = 0$ oraz $E(Z_1^4) < +\infty$. Wówczas*

$$E(\bar{\sigma}_n^2) = \sigma^2,$$

$$\operatorname{Var}(\bar{\sigma}_n^2) = (n+1)^{-1}\Big[2\sigma^4 + (n+1)^{-1}\big(\operatorname{Var}(Z_1 - m)^2 - 2\sigma^4\big) \sum_{k=0}^{n} E(Z_k^{-1})\Big].$$

Estymator $\bar{\sigma}_n^2$ jest więc nieobciążony, a jego słaba zgodność wynika z wyrażenia na wariancję i z nierówności Czebyszewa. Dion udowodnił również asymptotyczną normalność tego estymatora.

TWIERDZENIE 4.2. *Niech Z będzie nadkrytycznym procesem Galtona–Watsona, dla którego $p_0 = 0$ oraz $E(Z_1^6) < +\infty$. Wówczas*

$$P\left(\bar{\sigma}^2 \left(\frac{n+1}{2}\right)^{1/2} (\bar{\sigma}_n^2 - \sigma^2) \leqslant x\right) \to \Phi(x).$$

4.2. *Estymacja wariancji przy nieznanej średniej.* Dla estymacji wariancji przy nieznanej średniej Dion w pracy [8] zaproponował estymator postaci

$$\tilde{\sigma}_n^2 = \frac{1}{n+1} \sum_{k=0}^{n} Z_k \left(\frac{Z_{k+1}}{Z_k} - \hat{m}\right)^2.$$

Warto zauważyć, że C. C. Heyde w pracy [12] rozważał estymator tej postaci, ale w miejscu estymatora Harrisa $\hat{m}$ występował estymator Lotki–Nagajewa $\overline{m}$. Słaba zgodność estymatora $\tilde{\sigma}_n^2$ wynika z następującego twierdzenia.

TWIERDZENIE 4.3. *Niech Z będzie nadkrytycznym procesem Galtona–Watsona, dla którego $p_0 = 0$, $E(Z_1^4) < +\infty$. Wówczas*

$$|\bar{\sigma}_n^2 - \tilde{\sigma}_n^2| (1+n)^{1-\varepsilon} \xrightarrow{P} 0 \qquad \text{dla każdego } \varepsilon > 0.$$

Można ten wynik rozszerzyć na przypadek dowolnej miary probabilistycznej $Q \ll P$. Dowód wynika z pracy Diona ([7], tw. 3). Przedstawione estymatory wariancji są rozważane jedynie w przypadku $p_0 = 0$. Okazuje się przy tym, że nie można rozszerzyć własności tych estymatorów na zbiór niepochłaniania A metodami, które były stosowane przy rozważaniu estymatorów $\hat{m}$ i $\overline{m}$. Ciekawe byłoby rozważenie asymptotyki tych estymatorów na zbiorze A w ogólnej sytuacji, gdy $p_0 > 0$.

Dion w pracy [8] wykazał, że estymator największej wiarogodności dla wariancji σ^2 oparty na próbie $(Z_{jk};\ j = 0, 1, \ldots, n+1,\ k = 0, 1, \ldots)$ ma postać

$$\hat{\sigma}_n^2 = \sum_{k=0}^{\infty} (k-\hat{m})^2 \hat{p}_k.$$

Asymptotyczna normalność estymatora $\hat{\sigma}_n^2$ wynika z następującego twierdzenia.

TWIERDZENIE 4.4. *Jeżeli Z jest nadkrytycznym procesem Galtona–Watsona oraz $E(Z_1^4) < +\infty$, to dla miary probabilistycznej $Q \ll P_A$ zachodzi*

$$Q\left(\left(\frac{S_n}{\operatorname{Var}(Z_1-m)^2}\right)^{1/2} (\hat{\sigma}_n^2-\sigma^2) \leqslant x\right) \to \Phi(x),$$

$$Q\left(\left(\frac{1+m+\ldots+m^n}{\operatorname{Var}(Z_1-m)^2}\right)^{1/2} (\hat{\sigma}_n^2-\sigma^2) \leqslant x\right) \to \int_0^{\infty} \Phi(x\sqrt{y})\,dS(y).$$

Wykorzystując estymatory $\overline{p}_k$ można zaproponować estymator dla wariancji z nieznaną średnią, mianowicie

$$\underline{\sigma}_n^2 = \sum_{k=0}^{\infty} (k-\overline{m})^2 \overline{p}_k.$$

Zachodzi wówczas następujące

TWIERDZENIE 4.5. *Przy założeniach twierdzenia* 4.4 *zachodzą zbieżności*

$$Q\left(\left(\frac{Z_n}{\operatorname{Var}(Z_1-m)^2}\right)^{1/2} (\underline{\sigma}_n^2-\sigma^2) \leqslant x\right) \to \Phi(x),$$

$$Q\left(\left(\frac{m^n}{\operatorname{Var}(Z_1-m)^2}\right)^{1/2} (\underline{\sigma}_n^2-\sigma^2) \leqslant x\right) \to \int_0^{\infty} \Phi(x\sqrt{y})\,dS(y).$$

Dowód jest analogiczny do dowodu twierdzenia 4.4.

Można badać asymptotykę estymatora największej wiarogodności dla wariancji $\hat{\sigma}_n^2$, gdy $r, n \to +\infty$. Zachodzi następujące

TWIERDZENIE 4.6. *Jeżeli $E(Z_1^4) < +\infty$ oraz $r, n \to +\infty$, to*

(I) *jeżeli $m < 1$, to*

$$P\left(\left(\frac{r}{\operatorname{Var}(Z_1-m)^2(1-m)}\right)^{1/2} (\hat{\sigma}_n^2-\sigma^2) \leqslant x\right) \to \Phi(x),$$

(II) *jeżeli* $m = 1$ *oraz* $r/n \to +\infty$, *to*

$$P\left(\left(\frac{rn}{\mathrm{Var}(Z_1-m)^2}\right)^{1/2}(\hat{\sigma}_n^2-\sigma^2) \leqslant x\right) \to \Phi(x),$$

(III) *jeżeli* $m > 1$, *to*

$$P\left(\left(\frac{r(m^{n+1}-1)}{\mathrm{Var}(Z_1-m)^2(m-1)}\right)^{1/2}(\hat{\sigma}^2-\sigma^2) \leqslant x\right) \to \Phi(x).$$

D o w ó d. Za pomocą bezpośrednich rachunków można sprawdzić, że

$$\left|\hat{\sigma}_n^2-\sum_{k=0}^{\infty}(k-m)^2\hat{p}_k\right|(S_n^{(r)})^{1/2}\xrightarrow{P} 0.$$

Stąd wynika, że wystarczy wykazać prawdziwość tezy dla estymatora ${}^*\sigma_n^2 = \sum_{k=0}^{\infty}(k-m)^2\hat{p}_k$. Estymator ${}^*\sigma_n^2$ daje się przedstawić w postaci

$${}^*\sigma_n^2 = \frac{1}{S_n^{(r)}}\sum_{i=1}^{S_n^{(r)}}(\zeta_i-m)^2,$$

gdzie $S_n^{(r)} = Z_0^{(r)}+Z_1^{(r)}+\ldots+Z_n^{(r)}$, a $(\zeta_i;\ i = 1, 2, \ldots)$ jest ciągiem niezależnych zmiennych losowych o rozkładzie Z_1. Dalej dowód przeprowadza się standardowymi metodami (zob. [16]).

Prace cytowane

[1] K. B. A t h r e y a, P. E. N e y, *Branching processes*, Springer, Berlin 1972.

[2] A. B a d a l b a j e w, *Some properties of an estimate of the mean of a branching process*, Fan, Task. G. U. Taszkent 1977.

[3] W. J. B ü h l e r, *Ein zentraler Grenzwertsatz für Verzweigungsprozesse*, Z. Wahrscheinlichtkeitstheorie verw. Geb. 11 (1969), 139–141.

[4] P. B i l l i n g s l e y, *Statistical inference for Markov processes*, Chicago, 1961.

[5] —, *Convergence of probability measures*, Wiley, New York 1968.

[6] K. S. C r u m p, R. B. H o w e, *Nonparametric estimation of the age of a Galton–Watson branching process*, Biometrika 59 (1972), 533–538.

[7] J. P. D i o n, *Estimation of the mean and the initial probabilities of a branching process*, J. Appl. Prob. 11 (1974), 687–694.

[8] —, *Estimation of the variance of a branching process*, Ann. Statist. 3 (1975), 1183–1187.

[9] P. D. F e i g i n, *A note on maximum likelihood estimation for simple branching process*, Austr. J. Statist. 19 (1977), 152–154.

[10] T. E. H a r r i s, *Branching processes*, Ann. Math. Statist. 19 (1948), str. 474–494.

[11] C. C. H e y d e, *Extension of a result of Seneta for the supercritical Galton–Watson process*, ibidem 41 (1970), 739–742.

[12] —, *On estimating the variance of the offspring distribution in a simple branching process*, Adv. Appl. Prob. 3 (1974), 421–433.

[13] C. C. Heyde, J. R. Leslie, *Improved classical limit analogues for Galton–Watson process with or without immigration*, Bull. Austr. Math. Soc. 5 (1971), 145–155.

[14] A. V. Nagajew, *On estimating the expected number of direct descendants of a particle in a branching process*, Theor. Prob. Appl. 12 (1967), 312–320.

[15] J. Scott, *A central limit theorem for martingales and an application to branching process*, Stoch. Proc. Appl. Vol. 6, No 3 (1978), 89–94.

[16] N. M. Yanev, *On the statistics of branching processes*, Theor. Prob. Appl. 20 (1975), 505–515.