ЕКОНОМІКО-МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ

УДК: 519.86:330.46

НЕПАРАМЕТРИЧНА ОЦІНКА МОДЕЛІ ПРОЦЕСУ ЖИТТЯ (ІСНУВАННЯ) НА ОСНОВІ ПОПЕРЕЧНИХ ДАНИХ

НАКВАСЮК Н., аспірант Прикарпатського національного університету ім. Василя Стефаника

Оцінка моделі життя (існування) ґрунтується на використанні даних тільки однієї сукупності. Однак цей підхід, хоч і має відноситься до одного процесу, інколи, за змінних умов, потребує дослідження різних процесів. Це порушує такий постулат, як однорідність умов, що важливий при дослідженні багатьох процесів.

На противагу аналізу сукупності, поперечний аналіз дає можливість спостерігати умови впливу різноманітних економічних, суспільних, епідеміологічних, кліматичних чинників, що діють у певному календарному році (за виключенням короткотривалих процесів).

1. Дослідження і поперечні дані

При проектуванні поперечних досліджень, потрібно розпочати з визначення групи дослідження, тобто ідентифікованих осіб (чи інших об'єктів), на прикладі яких досліджується процес. Прикладом такої групи може бути все населення якоїсь держави чи виділеної адміністративної одиниці, наприклад, міста чи області.

При поперечному дослідженні потрібно чітко обирати *період спостереження*. На початку цього періоду є багато одиниць, які вже належать до групи дослідження. Інші одиниці долучаються до групи в ході досліджуваного періоду, окремі одиниці можуть уникнути спостереження, не зазнавши події. Такий рух *до-і-із* в ході періоду спостереження називається *міграцією*, вона типова для так званих "відкритих популяцій".

Поперечне дослідження надає інформацію про чисельність групи на початку періоду спостереження, а також про кількість об'єктів, що зазнали події в цей період, незважаючи на те, в який момент спостереження вони увійшли до групи. Ці дані часто називають *поперечними, поточними, обліковими даними* або *даними перепису*. Вони віддзеркалюють точні короткі фази процесу. Через відповідну класифікацію та маніпулювання даними, тобто кількістю об’єктів, що зазнали події, можна здійснити оцінку потрібних величин. Вагомою перевагою поперечного дослідження є можливість опертися, при оцінці, на попередній досвід групи, особливо, коли час до події є довгий. Потрібно зауважити, що дані поперечних досліджень можуть бути так перетворені, що уподібнюються до даних сукупності, хоча з самої природи спостереження випливає, що вони відносяться до різних сукупностей.

Дані перепису охоплюють *розклад актуального часу тривання* (наприклад, вік, стаж роботи) в окресленому моменті (наприклад, на початок календарного року) всіх актуально існуючих одиниць досліджуваної популяції та *розклад скінченого тривання* цих одиниць, що покинули групу в окресленому відрізку календарного часу (наприклад, вік у момент смерті, стаж роботи на час звільнення). Типове розміщення даних перепису про процес існування представлено у *таблиці*.

Відрізки часу тривання	Діапазон відрізка	Кількість об’єктів, що зазнали події	Кількість об’єктів, під час тривання		
			початок періоду	кінець періоду	середина періоду
$(x_r, x_{r+1}]$	c_r	D_r	P_r^z	P_r^{z+1}	P_r
x_0–x_1 x_1–x_2 … x_w	c_0 c_1 … c_w	D_0 D_1 … D_w	P_0^z P_1^z … P_w^z	P_0^{z+1} P_1^{z+1} … P_w^{z+1}	P_0 P_1 … P_w
Разом	C	D	P^z	P^{z+1}	P

Прийняті позначення: D – кількість об’єктів, які зазнають події в певний період календарного часу; P – кількість об’єктів, що існують в окреслений момент календарного часу. Незважаючи на це, спостереження стосуються одного уточненого календарного періоду, найчастіше року, з чітко окресленим початковим та кінцевим моментом. Для них прийнято позначення: z (верхній індекс) для окреслення початку періоду та $z+1$ – на позначення кінця періоду, який одночасно виступає початком наступного періоду. Діапазон інтервалу часу тривання – c. Якщо інтервали часу тривання є одиничними (наприклад, річні), то позначатимемо їх як $(x, x+1)$ і символ x буде виступати у ролі індексу. Якщо інтервали часу тривання могли б мати різний діапазон, то найкраще щоб ширші інтервали були цілими величинами основного (одиничного) інтервалу. Для опрацювання емпіричних даних, інтервали трактуємо як правосторонньо закриті.

В окремих випадках типове розміщення даних перепису, подане у таблиці, може бути недостатнім. Часто потрібні детальні дані про кількість об'єктів, які зазнають події (які покидають групу дослідження) в даному календарному періоді. Найкраще поділяти пов'язану кількість таких фактів (D_x), які виступають в одиничному періоді часу тривання на ті, які є частиною об'єктів, що належать до даної групи часу тривання на початковому етапі періоду дослідження (${}_{\delta}D_x^z$), а також на ті, що залишилися, які будуть частиною об'єктів, що увійшли до цієї групи часу тривання в ході періоду дослідження (${}_{\alpha}D_x^z$). Сутність цієї деталізації можна представити у графічній формі на секторі діаграми Лексіса, що ілюструє одночасно відношення між даними сукупності та даними перепису [1].

Кінцевою метою цієї процедури є поєднання частин об'єктів, які зазнали події у двох інтервалах часу тривання, щоб отримати величини таких фактів у даному календарному році, але які відносяться до однієї сукупності:

$$ {}_{\delta}D_x^z + {}_{\alpha}D_{x+1}^z, \quad для\ x = 0,1,...,w. \tag{1} $$

Така процедура наближає систему даних перепису до системи даних сукупності, будь-які отримані в такий спосіб суми не є рівні величинам d_i виходів з сукупності в окреслених інтервалах часу тривання (t_i, t_{i+1}). Величини d_i в аналізі сукупності стосуються двох досліджуваних періодів, а саме:

$$ d_i = d_x = {}_{\alpha}D_x^z + {}_{\delta}D_x^{z+1}, \quad для\ x = 0,1,...,w. \tag{2} $$

Обидві суми (1) і (2) відносяться до однієї сукупності й використовуються для побудови загально корисних співчинників.

Потрібно зауважити, що класифікація фактів настання події об'єктами відповідно до сукупності дозволяє окреслити кількість об'єктів, які в даному періоді дослідження досягають часу тривання x, що неможливо визначити на основі типової системи даних перепису, які представляє таблиця. Розділивши факти на частини ${}_{\alpha}D$ і ${}_{\delta}D$, достатньо буде утворити суми $P_x^{z+1} + {}_{\alpha}D_x^z$ чи різницю $P_x^z - {}_{\delta}D_x^z$ (x=0, 1, ..., w) при цьому $P_x^z - {}_{\delta}D_x^z = P_{x+1}^{z+1} + {}_{\alpha}D_{x+1}^z$. Отримані суми чи різниці відповідають в аналізі сукупності величинам $l_i = l_x$ об'єктів, що доживають в окресленому періоді тривання (на початку інтервалу часу тривання). В обговорюваному випадку розмежування об'єктів, що зазнали події ${}_{\alpha}D$ і ${}_{\delta}D$ практично не становить труднощів. Вистачить вихідних даних перепису у відношенні як у таблиці для річних інтервалів часу тривання і додатково потрібна інформація про вихідну кількість когорти даного року l_0. Потрібні величини ${}_{\alpha}D$ і ${}_{\delta}D$ отримаємо з вирахування, наприклад, ${}_{\delta}D_0^z = l_0 - P_0^{z+1}$; ${}_{\delta}D_0^z = D_0^z - {}_{\alpha}D_0^z$; ${}_{\delta}D_1^z = (P_0^z - P_1^{z+1}) - {}_{\delta}D_0^z$ і т.д.

Спираючись на дані перепису та їх деталі, можна визначити певні співчинники, що мають велике значення при дискретній оцінці функцій теорії ймовірності процесу існування:

• частковий коефіцієнт подій:

$$m_x = \frac{D_x^z}{\frac{1}{2}(P_x^z + P_x^{z+1})} = \frac{D_x^z}{P_x}; \tag{3}$$

• перехідний коефіцієнт подій:

$$w_x = \frac{{}_{\delta}D_x^z + {}_{\alpha}D_{x+1}^z}{P_x^z} = \frac{P_x^z - P_{x+1}^{z+1}}{P_x^z}; \tag{4}$$

• коефіцієнт подій сукупності:

$$z_x = \frac{{}_{\delta}D_x^z + {}_{\alpha}D_{x+1}^z}{\frac{1}{2}(P_x^z + P_{x+1}^{z+1})} = \frac{P_x^z - P_{x+1}^{z+1}}{\frac{1}{2}(P_x^z + P_{x+1}^{z+1})}; \tag{5}$$

• умовна ймовірність настання події:

$$q_x = \frac{{}_{\delta}D_x^z + {}_{\alpha}D_{x+1}^z}{l_x^z} = \frac{{}_{\delta}D_x^z + {}_{\alpha}D_{x+1}^z}{P_{x-1}^z - {}_{\delta}D_{x-1}^z}. \tag{6}$$

Подані приклади можна узагальнити для довільних інтервалів часу тривання (x_r, x_{r+1}). У демографічному підході до оцінки моделі існування звичайна кількість смертей і живих осіб розміщується в річних інтервалах віку, але інколи також в інтервалах більших, ніж рік, зазвичай 5-ти річні.

2. ЧАСТКОВИЙ КОЕФІЦІЄНТ ПОДІЙ ЯК ОЦІНКА ІНТЕНСИВНОСТІ ПРОЦЕСУ

Частковий коефіцієнт подій (3), узагальнений для довільного інтервалу часу тривання $c_r = x_{r+1} - x_r$, має вигляд: $m_r = \frac{D_r}{P_r}$, для $r = 0, 1, ..., w$, оминаючи верхній індекс, що означає календарний період, якого власне і стосується коефіцієнт.

При умові, що кількість об’єктів, які існують протягом цілого досліджуваного періоду в даному інтервалі часу тривання (x_r, x_{r+1}) є сталою і становить P_r, а також, що кожен об’єкт у такій самій мірі піддається настанню події, частковий коефіцієнт подій m_r є оцінкою найбільшої ймовірності сталої інтенсивності процесу λ_r. Кількість фактів настання події D_r у досліджуваному періоді є випадковою величиною, крім того маємо справу зі схемою зворотної вибірки D_r серед P_r об’єктів, що при прийнятих умовах змінної D_r, мають біноміальний розклад:

$$D_r \sim bin\{n = P_r; q_r = 1 - \exp(-\lambda_r)\},$$

де q_r – ймовірність настання події (що виражає згадувану вище сталу схильність) у досліджуваному періоді.

Якщо ймовірність q_r є малою, то при $P_r \to \infty$ (практично для досить великих P_r) біноміальне рівняння збігається з рівнянням Пуассона з параметром $P_r\{1 - \exp(-\lambda_r)\} \approx P_r \lambda_r$. Разом з тим кількість фактів настання події D_r описується рівнянням Пуассона:

$$Poisson(D_r; P_r\lambda_r) = \frac{(P_r\lambda_r)^{D_r} \exp(-P_r\lambda_r)}{D_r!}.$$

Щоб визначити оператор невеликої величини λ_r, скористаємося функцією достовірності, яка для одного спостереження D_r має такий же вигляд, як наведена вище функція, але є вже функцією параметру λ_r. Шуканий оператор є розв’язком рівняння:

$$\frac{d \ln L}{d\lambda_r} = \frac{d}{d\lambda_r}[D_r \ln P_r + D_r \ln \lambda_r - P_r\lambda_r - \ln(D_r!)] = 0, \tag{10}$$

і має вигляд:

$$\hat{\lambda}_r = \frac{D_r}{P_r} = m_r. \tag{11}$$

Дисперсію оператора m сталої інтенсивності процесу знаходимо з другої похідної логарифму функції достовірності $1/D^2(m_r) = -d^2 1nL/d\lambda_r$, що дає:

$$D^2(m_r) = \frac{\lambda_r}{P_r}, \tag{12}$$

і відповідає дисперсії, отриманій просто з рівняння Пуассона. Очікування цієї дисперсії отримаємо з рівняння:

$$S^2(m_r) = \frac{m_r}{P_r} = \frac{m_r^2}{D_r} . \tag{13}$$

Концепція часткового коефіцієнта подій, трохи відома у демографічних застосуваннях, не є досить чіткою та зрозумілою. Згідно з обов'язковою формулою, потрібно встановити кількість об'єктів, що зазнають події у даному інтервалі часу тривання та віднести її до середньої кількості об'єктів, що живуть у цьому інтервалі. Частковий коефіцієнт має відповідати на питання, яка кількість об'єктів, що належать події представлені відносно можна очікувати в даному інтервалі часу тривання протягом досліджуваного періоду, (наприклад року). Натомість, легко зауважити, що не ті самі об'єкти піддаються "ризику настання" події в досліджуваному періоді. У кожну хвилину нашого спостереження виходять певні об'єкти при перетині верхньої межі даного інтервалу часу тривання, і входять інші об'єкти, досягаючи час тривання, охоплений дослідженням. Можна це показати на діаграмі Лексіса (*рис. 1*), де з метою кращого розуміння прийнято рівні інтервали часу тривання періодичні спостереження (річні).

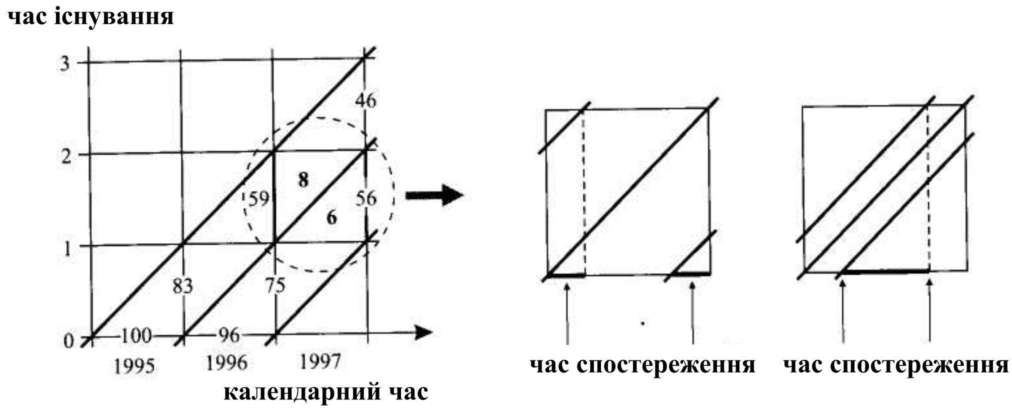

Рис. 1. Частковий коефіцієнт подій – суть і сумніви

Найстарші об'єкти з сукупності 1995 і наймолодші з сукупності 1996 лише короткий час залишаються на полі спостереження, яким є календарний рік 1997. Найдовше велося спостереження за тими об'єктами, що появилися на трохи – перед чи трохи після 1 січня 1996 р. Таке перебування нових об'єктів на полі спостереження спричиняє неясність у визначенні середньої кількості об'єктів. Потрібно зауважити, що жодного об'єкта, який виступає на початку даного року у даному інтервалі часу тривання, немає серед тих, які вирішують кількісний стан на тому самому інтервалі на початку наступного року. Це абсолютно інші об'єкти і належать до інших сукупностей. Така неясність є не дуже помітною, коли оперуємо багаторічними (чи великими періодами) інтервалами часу тривання.

Не всі виходи з обох сукупностей у даному періоді спостереження містяться в чисельності коефіцієнта. Чисельність становлять лише ті об'єкти, що зазнають події в даному періоді спостереження в окресленому інтервалі часу тривання; інші об'єкти, що в досліджуваному періоді зазнають події, належать до вищого чи нижчого класу часу тривання.

Однак ці неясності не дискваліфікують описуваного часткового коефіцієнта подій як цінної інформації росту процесу існування в даному класі часу тривання. Виявлено, що коефіцієнт m_r для даних перепису є оцінкою сталої інтенсивності процесу λ_r в даному інтервалі часу тривання. Можна сказати, що він відповідає оператору $\hat{\lambda}_i^{(2)}(i=r)$ для даних сукупності. Порівняння двох прикладів показує, наближення очевидне, особливо з погляду на те, що всі об'єкти, які змінюються під час спостереження, піддаються ризику настання події в даному класі часу тривання. Чисельники виразів для m_r та $\hat{\lambda}_i^{(2)}$ представляють ту саму величину, хоча відповідно до іншого типу даних. Якщо говорити про знаменники, то середню кількість об'єктів у періоді спостереження (чи краще – кількість об'єктів на середині етапу періоду спостереження) можна ідентифікувати за кількістю одиниць часу (наприклад, років), які в ході періоду спостереження досліджуваної одиниці групи об'єктів прожили в інтервалі часу тривання (x_r, x_{r+1}).

Використання m_r як оператора найбільшої ймовірності процесу вимагає досить ригористичних положень. Крім того, реальність відходить зазвичай від принципів моделі, тому потрібно сподіватися, що оператор може не мати властивостей, які випливають з методу. Особливо у ситуації, коли $P_r \neq const$, а різниця між P_r^z і P_r^{z+1} є значною, потрібно бути обережним із висновками, особливо при оцінці похибки стандартного очікування $S(m_r)$.

3. ВИКОРИСТАННЯ ЧАСТКОВИХ КОЕФІЦІЄНТІВ ПОДІЙ ДЛЯ ОЦІНКИ ФУНКЦІЇ ІСНУВАННЯ

Узятий із даних перепису оператор m_r сталої інтенсивності процесу λ_r можна використовувати для оцінки інших функцій, які описують процес існування. Зупинимося на оцінці *функції існування* і представимо два найбільш відомі й загальні підходи, що найчастіше використовуються.

Метод кривої стабільності. Залежність функції існування $S(t)$ від інтенсивності процесу $\lambda(t)$ записується таким виразом:

$$S(t) = \exp\left\{ -\int_0^t \lambda(u)du \right\}. \tag{14}$$

Застосовуючи східцеве наближення λ_r функції інтенсивності та скориставшись її оператором m_r, отримаємо з рівняння (14) такий оператор функції існування:

$$\hat{S}_r = \exp\left\{-\sum_{j=0}^{r-1} m_j c_j\right\}, \text{ для } r = 0,1,...,w, \tag{15}$$

де c_j – ширина інтервалу часу тривання (x_j, x_{j+1}).

Описуючи зміну цього оператора, слід зауважити, що він є нелінійною функцією випадкових змінних m_j:

$$\hat{S}_r = \exp(-m_0c_0 - m_1c_1 - ... - m_{r-1}c_{r-1}) . \tag{16}$$

Припускаючи, що змінні m_j і m_k є незкорельованими, визначимо відповідне наближення виразу:

$$D^2(\hat{S}_r) \approx \sum_{j=0}^{r-1}\left(\left.\frac{\partial \hat{S}_r}{\partial m_j}\right|_{\lambda_j}\right)^2 D^2(m_j) , \tag{17}$$

де $(\partial \hat{S}_r / \partial m_j)_{\lambda_j}$ означає, що часткова похідна $\partial \hat{S}_r / \partial m_j$ обчислюється при $m_j = \lambda_j$ [2; 3].

Якщо тепер у виразі (17) врахувати, що $(\partial \hat{S}_r / \partial m_j)_{\lambda_j} = -c_j S_r$, а далі замінити λ_j на її очікування та використати вираз (12) для зміни оператора m_j, то отримаємо вираз для похибки стандартно очікування $\hat{S}_r$:

$$S(\hat{S}_r) = \hat{S}_r \sqrt{\sum_{j=0}^{r-1} \frac{c_j^2 m_j}{P_j}} . \tag{18}$$

Очікування функції $S(x)$ може бути застосовано для будь-якої величини $x \in (x_r, x_{r+1})$ і тоді оператор можна записати як:

$$\hat{S}(x) = \exp\left\{-\sum_{j=0}^{r-1} m_j c_j - (x - x_r) m_r\right\} . \tag{19}$$

Статистичний метод. Виразимо функцію існування S_r у вигляді добутку умовних ймовірностей існування (знаючи, що $S_r = p_{0r}$):

$$S_r = p_{01}p_{12}...p_{r-1,r} = \prod_{j=0}^{r-1} p_{j,j+1}\,, \tag{20}$$

де $p_{j,j+1}$ відноситься до довільних інтервалів (x_j, x_{j+1}), які, звичайно, згруповані.

Скориставшись наближенням $m_j \approx \hat{\lambda}_j^{(2)}$, представимо відомий з аналізу сукупний оператор інтенсивності процесу $\hat{\lambda}_j^{(2)} = d_j/(l_{j+1}c_j + d_j c_j a_j)$ як:

$$\hat{\lambda}_j^{(2)} = m_j = \frac{1-\hat{p}_{j,j+1}}{[\hat{p}_{j,j+1}(1-a_j)+a_j]c_j}\,. \tag{21}$$

Розв'язуючи вище подане рівняння стосовно $\hat{p}_{j,j+1}$, знайдемо такий вигляд оператора ймовірності існування:

$$\hat{p}_{j,j+1} = \frac{1-a_j c_j m_j}{1+(1-a_j)c_j m_j}\,, \tag{22}$$

і використовуючи ще вираз (20), отримаємо:

$$\hat{S}_r^{(2)} = \prod_{j=0}^{r-1} \frac{1-a_j c_j m_j}{1+(1-a_j)c_j m_j}\,, \text{ для } r = 0,1,...,w. \tag{23}$$

Для a_j=1/2, маємо:

$$\hat{S}_r^{(2)} = \prod_{j=0}^{r-1} \frac{2-c_j m_j}{2+c_j m_j}\,. \tag{24}$$

Наближений вираз для дисперсії оператора отримаємо з виразу (17), який застосовується в двох випадках, перший для функції $\hat{S}_r = \varphi(p_j)$, а другий для $\hat{p}_{j,j+1} = \psi(m_j)$:

$$D^2(\hat{S}_r^{(2)}) \approx \sum_{j=0}^{r-1} \left(\left.\frac{\partial \hat{S}_r}{\partial \hat{p}_j}\right|_{p_j} \right)^2 \left(\left.\frac{\partial \hat{p}_j}{\partial m_j}\right|_{\lambda_j} \right)^2 D^2(m_j)\,, \tag{25}$$

Виконуючи зазначену дію і впроваджуючи очікувані m_j та $\hat{S}_r^{(2)}$, отримаємо вираз для оцінки дисперсії оператора, й далі його стандартну похибку [3]:

$$S(\hat{S}_r^{(2)}) = \hat{S}_r \left\{ \sum_{j=0}^{r-1} \frac{m_j c_j^2}{P_j [1 + (1 - a_j) c_j m_j]^2 [1 - a_j c_j m_j]^2} \right\}^{\frac{1}{2}}. \qquad (26)$$

У випадку для a_j=1/2, стандартну похибку порахуємо за виразом:

$$S(\hat{S}_r^{(2)}) = \hat{S}_r \left\{ \sum_{j=0}^{r-1} \frac{16 m_j c_j^2}{P_j (4 - c_j^2 m_j^2)^2} \right\}^{\frac{1}{2}}. \qquad (27)$$

Оператори $\hat{S}_r$ і $\hat{S}_r^{(2)}$ дають дуже наближену оцінку, тільки очікування відповідно $\hat{S}_r^{(2)}$ може відноситися до виключених точок r=1, 2, …, w.

4. ВИКОРИСТАННЯ ЧАСТКОВИХ КОЕФІЦІЄНТІВ ПОДІЙ В ОЦІНЦІ ІНШИХ ФУНКЦІЙ ПРОЦЕСУ ІСНУВАННЯ

Інші функції ймовірності, що описують процес існування, можна оцінити за допомогою отриманих операторів $\hat{S}_r$ або $\hat{S}_r^{(2)}$ функції існування, використовуючи основні залежності. Оператором функції густини ймовірності є представлення:

$$\hat{f}_r = \frac{\hat{S}_r - \hat{S}_{r+1}}{c_r}. \qquad (28)$$

Умовну ймовірність настання події визначимо за допомогою рівняння:

$$\hat{q}_r = \frac{\hat{S}_r - \hat{S}_{r+1}}{c_r \hat{S}_r}, \qquad (29)$$

хоча можна також оцінювати безпосередньо, без участі функції існування, виражаючи як функцію часткового коефіцієнта подій, наприклад:

$$\hat{q}_r = \frac{1 - \exp(-m_r c_r)}{c_r}, \qquad (30)$$

або:

$$\hat{q}_r = \frac{2m_r}{2 + m_r c_r}, \tag{31}$$

при цьому вираз (31) – скорочений і спрощений вираз (30). В обох випадках оцінку $\hat{q}_r$ потрібно відносити до першого одиничного підінтервалу $(x_r, x_r{+}1)$ в інтервалі (x_r, x_{r+1}). Найбільш підходящий приклад оператора умовної ймовірності настання події є (31), що показує співвідношення між $\hat{q}_r$ та $\hat{\lambda}_r \approx m_r$. Він відповідає концепційному представленню (24) оператора функції існування. Похибку стандартної оцінки $\hat{q}_r$ можна вирахувати з виразу:

$$S(\hat{q}_r) \approx \left\{ \frac{16 m_r}{P_r (2 + c_r m_r)^4} \right\}^{\frac{1}{2}}. \tag{32}$$

Для оцінки середнього подальшого часу тривання можна скористатися оператором у вигляді:

$$\hat{e}_r = \frac{1}{\hat{S}_r} \left\{ \sum_{j=r}^{w-1} \frac{\hat{S}_j - \hat{S}_{j+1}}{m_j} + \frac{1}{2} c_w \hat{S}_w \right\}, \text{ для } r = 0, 1, ..., w, \tag{33}$$

в якому потрібно використати одну з двох можливих оцінок функції існування (15) або (24). Отже, матимемо дві оцінки з незначною різницею.

5. Замітки щодо використання інших коефіцієнтів в оцінці моделі існування

У першому розділі наведено чотири коефіцієнти для даних перепису, серед яких найчастіше вживається частковий коефіцієнт. Виходячи з мети аналізу цього достатньо, тим паче, що коефіцієнт не вимагає розмежування випадків настання події D_x, що наступають в даному інтервалі часу тривання на ${}_{\delta}D_x^z$ та ${}_{\alpha}D_x^z$. Досить скрупульозною є вимога, щоб частковий коефіцієнт можна визнати як ефективний оператор сталої інтенсивності процесу, в наближенні здійснення для великих груп спостережуваних одиниць (наприклад, великих популяцій). Не завжди досліджувані групи є настільки великі, щоб були протягом всього часу абсолютно незмінні кількості одиниць в окремих інтервалах часу тривання. Тому, для раніше обговорених неясностей, на які звернув увагу Прессат [4], краще використовувати інші величини:

- коефіцієнт перехідний, як оператор ймовірності настання події;
- коефіцієнт сукупності, як оператор інтенсивності процесу;

- коефіцієнт, який має конструкцію, що відповідає частковому визначенню ймовірності, тобто є природнім оператором умовної ймовірності настання події.

Перехідний коефіцієнт можемо записати у двох альтернативних, ідентично рівних виглядах (співвідношення (4)):

$$w_x = \frac{{}_\delta D_x^z + {}_\delta D_{x+1}^z}{P_x^z}, \tag{34}$$

$$w_x^* = \frac{P_x^z - P_{x+1}^{z+1}}{P_x^z}, \tag{35}$$

при цьому рівняння (35) відноситься до ситуації, коли час тривання згрупований на річні інтервали при річному періоді спостереження.

Перехідний коефіцієнт ґрунтується на класичному визначенні ймовірності і як такий, що виражає можливість настання об'єктом події протягом періоду спостереження (календарного року) за умови, що на початку періоду дослідження об'єкт був у даному класі часу тривання. Крім того, він відноситься до чітко виділеної сукупності, і не є однак умовною ймовірністю настання події (q_i). З метою уникнення непорозуміння, обидві величини показано на діаграмі Лексіса (*рис. 2*).

$$w_{1,2} = \frac{8+5}{59} = 0{,}220 \qquad q_{1,2} = \frac{8+8}{67} = 0{,}239$$

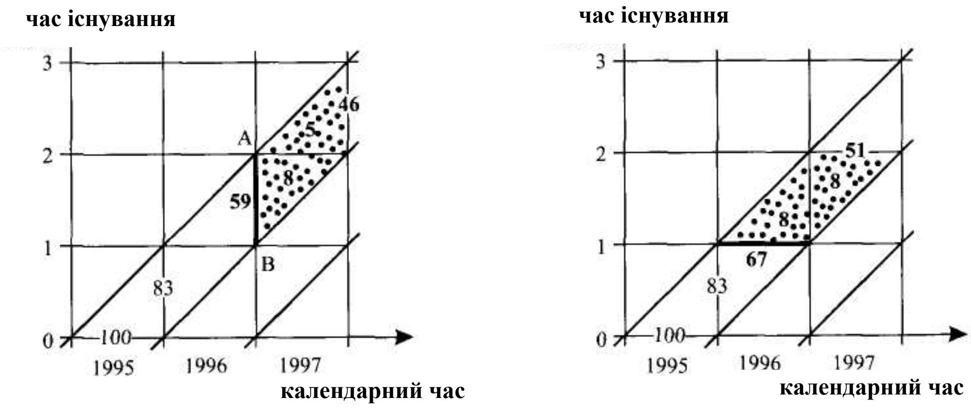

Рис. 2. Перехідний коефіцієнт та умовна ймовірність

Перехідний коефіцієнт w_r трактуватимемо як оператор ймовірності певної події в одиниці календарного часу в групі одиниць названих на початку цього періоду [1]. Впровадження індексу r дозволяє узагальнити коефіцієнт на інтервал часу тривання з довільною шириною, так само більшою, як і меншою від одиниці календарного часу, що буде періодом спостереження. Найвигідніше розглядати інтервали часу тривання та період спостереження на однаковій довжині. Про цю групу можна вже сказати, що одиниці, які її становлять, досягли на початку періоду середній час тривання $x_r+c_r/2$, у кінці $x_r+c_r/2+1$. Одиниці, що зазнали події мають натомість закінчений час тривання рівний "середньому" $x_r+c_r/2+1/2$ одиниць часу, причому одиниця часу приймає довжину періоду спостереження. У випадку, коли одиницею часу є рік (річними є інтервали часу тривання та період спостереження) спостереження охоплює суто один календарний період, хоча стосується двох інтервалів часу тривання: (x_r, x_{r+1}) та (x_{r+1}, x_{r+2}) (r=0, 1, ..., w). Слід відзначити, що перехідний коефіцієнт не охоплює всіх випадків настання події у цих інтервалах часу тривання, як це було у випадку часткового коефіцієнта.

Приймемо, що на початку календарного періоду кожна одиниця в даному класі (групі) часу тривання має таку саму ймовірність $q_r^{(z)}$ настання події в цьому періоді, а також одиниці зазнають події незалежно. Кількість одиниць, що зазнають події в досліджуваному періоді запишемо за допомогою біноміального рівняння $bin(P_r^z; q_r^{(z)})$, а середню похибку очікування w_r будемо обчислювати за допомогою виразу:

$$S^2(w_r) = \frac{w_r(1-w_r)}{P_x^z} = \frac{{}_\delta D_r^z + {}_\alpha D_{r+1}^z}{(P_x^z)^2} - \frac{({}_\delta D_r^z + {}_\alpha D_{r+1}^z)^2}{(P_x^z)^3}, \tag{36}$$

а у випадку одиничних інтервалів скористаємося рівнянням:

$$S^2(w_r^*) = \frac{(P_r^z - P_{r+1}^{z+1})P_{r+1}^{z+1}}{(P_x^z)^3}. \tag{37}$$

На основі оцінок можна оцінити функцію існування, а далі інші функції моделі існування в системі сукупності. Оцінка функції існування вимагає застосування статистичного оператора, який ґрунтується на добутку ймовірностей протилежних подій для $t_r+c_r/2$. Якщо необхідні очікування для цілковитих величин часу тривання t_r, t_{r+1} і т.д., то потрібен спеціальний підхід, що ґрунтується на поєднанні ймовірностей існування підінтервалів, наприклад:

$$\hat{P}_{r,r+1} = (1-w_{r-1})^{1/2}(1-w_r)^{c_r-1/2}. \tag{38}$$

Концепція розв'язання проблеми є простою, але запровадження її на практиці досить складне. Цей метод докладніше розглянуто у відповідній літературі [3; 5].

Сукупний коефіцієнт подій, відомий у демографії як коефіцієнт смертності стосовно покоління чи груп поколінь [4], може використовуватися в аналізі існування як оператор інтенсивності виходів. Визначимо його як частку кількості випадків настання події у даній множині протягом досліджуваного періоду до кількості членів сукупності на середньому етапі цього періоду. Узагальнимо раніше подане рівняння (5):

$$z_r = \frac{{}_{\delta}D_x^z + {}_{\delta}D_{x+1}^z}{\frac{1}{2}(P_r^z + P_{r+}^{z+1})}. \tag{39}$$

У цьому рівнянні P_{r+}^{z+1} означає чисельність сукупності в кінці періоду дослідження (початок наступного періоду), а також в інтервалі часу тривання (x_r+1, x_{r+1}+1), межами якого є зміщення 0+1 стосовно до вихідного інтервалу (t_r, t_{r+1}). Коли інтервали багаторічні, хоча інтервал із такими межами не мусить виникати з групування часу тривання (наприклад, коли у групуванні виділено інтервал 5–10 років, тоді протікання річного періоду спостереження мусимо взяти до уваги інтервал 6–11 років). Величину P_{r+}^{z+1} не потрібно трактувати окремо від P_r^z. Якщо час тривання згрупований у річні інтервали, то $P_{r+}^{z+1} = P_{r+1}^{z+1}$. У такій ситуації рівняння (5) можна записати у вигляді:

$$z_r = \frac{P_x^z - P_{x+1}^{z+1}}{\frac{1}{2}(P_x^z + P_{x+1}^{z+1})}. \tag{40}$$

Для обчислення цього коефіцієнта підійдуть такі самі дані, що й для перехідного коефіцієнта, оскільки $P_{r+}^{z+1} = P_r^z - {}_{\delta}D_r^z - {}_{\alpha}D_{r+1}^z$. Для чисел розміщених на *рис. 2* величина коефіцієнта становить (8+5)/ 1/2(59+46)=0.248. Коефіцієнт z_r має таку саму будову як і оператор сталої інтенсивності подій. Його чисельник становить вихід у "середньому" інтервалі часу тривання (t_r+c_r/2, t_r+c_r/2+1), знаменник натомість виражає середню кількість одиниць у цьому інтервалі, яку слід трактувати як сумарний час, прожитий P_r^z одиницями протягом періоду спостереження. Отже, сукупний коефіцієнт подій є оператором найбільшої ймовірності сталої інтенсивності подій в інтервалі (t_r+c_r/2, t_r+c_r/2+1). При даних P_r^z та P_{r+}^{z+1} функція ймовірності має такий вигляд:

$$L = [\lambda_r \exp(-\lambda_r / 2)]^{P_r^t - P_{r+}^{z+1}} [\exp(-\lambda_r)]^{P_{r+}^{z+1}} . \qquad (41)$$

$$s^2(z_r) = \frac{4(P_r^z - P_{r+}^{z+1})}{(P_r^z + P_{r+}^{z+1})^2} . \qquad (42)$$

Коефіцієнт z_r тісно пов’язаний із коефіцієнтом w_r. Якщо w_r розуміємо як умовну ймовірність "зазнання" події в середньому інтервалі часу тривання від $t_r+c_r/2$ до $t_r+c_r/2+1$, то z_r є оператором сталої інтенсивності виходів у цьому інтервалі. Для цього достатньо скористатися рівнянням (21), яке для a_r =1/2 та c_r=1 (для спрощення згрупуємо час тривання в річні інтервали) приймає вигляд:

$$\hat{\lambda}_r^{(2)} = \frac{2(1-\hat{p}_r^*)}{(1+\hat{p}_r^*)} , \qquad (43)$$

де $\hat{p}_r^*$ – оператор умовної ймовірності настання події в одиничному інтервалі, скажімо від t_r+1/2 до t_r+3/2. Беремо $\hat{p}_r^* = 1 - w_r^* = P_{r+1}^{z+1} / P_r^z$ (див. рівняння 35) й підставимо у наведене вище співвідношення, отримуємо сукупний коефіцієнт виходів z_r^*, який описаний виразом (40). Можна констатувати, що z_r та w_r пов’язані співвідношеннями:

$$z_r = \frac{2w_r}{1 - w_r} , \qquad (44)$$

$$w_r = \frac{2z_r}{2 + z_r} , \qquad (45)$$

які є зміненою і спрощеною формою відомого рівняння q_i=1–exp(–λ_i).

Труднощі, що виникають при використанні коефіцієнта w_r ґрунтуються на тому, що оцінка функцій ймовірності моделі тривання стосується інтервалів чи етапів "з половиною", є також частка коефіцієнта z_r. Якщо відносно z_r, як оператора інтенсивності подій, хочемо її уникнути, то вистачить у рівнянні (43) прийняти до уваги $\hat{p}_r$ згідно виразу (38).

Таким чином можна вважати, що представлені всі можливі способи застосування даних перепису для оцінки моделі існування. Один із підходів використання окремої модифікації даних, що ґрунтується на положенні Р. Прессата [4], запропонував А. Балицький [3].

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. *Keyfitz N.* Introduction to the Methematics of Population (with Revisions) / N. Keyfitz ; Addison-Wesley Publishing Company, Reading. — Massachusetts, 1977.
2. *Abergauz* i in. Rachunek probabilistyczny : poradnik. — Warszawa : Wydawnictwo MON, 1973.
3. *Balicki A.* Analiza kohortowa w badaniu plynnosci kadr / A. Balicki // Studia Demograficzne. — 1986. — № 2.
4. *Pressat R.* Analiza demograficzna / R. Pressat. — Warszawa : PWN, 1966.
5. *Bartholomew D. J.* Statistical techniques for manpower planning / D. J. Bartholomew, A. F. Forbes. — London : Wiley, 1981.