# МОДИФІКАЦІЯ ОЦІНКИ КАПЛАНА-МЕЙЄРА ДЛЯ МОДЕЛІ СУМІШЕЙ ЗІ ЗМІННИМИ КОНЦЕНТРАЦІЯМИ

Р. Є. МАЙБОРОДА AND В. Г. ХІЗАНОВ

Анотація. Запропонована модифікація оцінки Каплана-Мейера для оцінювання розподілу компонентів суміші зі змінними концентраціями за цензурованими даними. Доведена консистентність цих оцінок у рівномірній нормі та отримана верхня межа для швидкості їх збіжності.

A modification of Kaplan-Meier estimator is proposed for mixture components CDFs estimation by censored data in the case when mixing probabilities vary from observation to observation. Consistency of the estimators in the sup-norm is demonstrated and an upper bound for the convergence rate is derived.

## 1. Вступ

Задача оцінювання розподілу тривалості життя за спостереженнями, випадково цензурованими зправа, природно виникає у актуарній та медічній статистиці, а також при аналізі надійності приладів [1, 2]. У випадку, коли цензуровані спостереження є незалежними та однаково розподіленими, найбільш поширеною оцінкою для функції розподілу є оцінка Каплана-Мейєра [3]. У даній роботі ми розглядаємо модифікацію цієї оцінки для випадку, коли спостережувані об'єкти вибирають з суміші, причому концентрації компонентів у цій суміші змінюються від спостереження до спостереження. Схожа задача розглядалась у роботі А. Рижова [6]. Але запропоновану у [6] техніку оцінювання можна використовувати лише у випадку, коли концентрації компонентів у суміші приймають значення з деякого скінченного набору чисел. У даній роботі пропонується оцінка, яку можна використовувати для довільних концентрацій. При побудові цієї оцінки використані результати роботи Р. Д. Гілла та С. Йохансена [9], де оцінка Каплана-Мейєра трактується як продакт-інтеграл від оцінки інтегральної функції ризику. Крім того, ми використовуємо для оцінювання навантажені емпіричні розподіли, застосування яких для сумішей зі змінними концентраціями розглядалось у роботах Р. Майбороди та О. Сугакової [4, 5]. Про задачі, що приводять до моделі суміші зі змінними концентраціями див. також [7, 8].

Далі у п. 2 формально описана ймовірнісна модель даних, та запропонована модифікація оцінки Каплана-Мейєра для них. Тут також сформульована теорема про умови рівномірної консистентності цієї оцінки та швидкість її збіжності. У п. 3 містяться допоміжні відомості про продакт-інтеграли. П. 4 присвячено доведенню теореми про консистентність. У п. 5 описано результати дослідження поведінки оцінок на модельованих вибірках.

---

2000 *Mathematics Subject Classification.* 62N05 ; 62G05.

*Ключові слова і фрази.* Оцінка Каплана-Мейєра, моделі сумішей зі змінними концентраціями, консистентність, цензурування

Kaplan-Meier estimator, mixtures with varying concentrations. consistency, censoring.

## 2. Постановка задачі

Розглянемо вибірку з $n$ об'єктів $O_1, \dots, O_n$ з суміші зі змінними концентраціями, тобто будемо вважати, що кожен $O_j$ належить одній з $M$ популяцій (компонентів суміші). Позначимо $\mathrm{ind}(O_j)$ — номер популяції, якій належить $O_j$. Вважаємо, що значення $\mathrm{ind}(O_j)$ невідомі, але відомі концентрації компонентів у суміші при відборі $O_j$, тобто $w_j^m = \mathbb{P}(\mathrm{ind}(O_j) = m), m = 1, \dots, M$. Для кожного $O_j$ існує невід'ємна характеристика $\xi_j = \xi(O_j)$, яку будемо називати тривалістю життя $O_j$. Розподіл $\xi(O)$ залежить від того, якому компоненту належить $O$:

$$F_m(A) = \mathsf{P}(\xi(O) \in A \mid \mathrm{ind}(O) = m)$$

Ми вважаємо, що при спостереженні $O_j$ відбувається випадкове цензурування справа, тобто спостерігається пара $\xi_j^* = \min(\xi_j, c_j)$ та $\delta_j = \mathbb{1}\{\xi_j^* \le c_j\}$, де $c_j$ — невід'ємна випадкова величина, що зветься цензором. При цьому розподіл $c_j$ також може бути різним для різних компонентів:

$$G_m(A) = \mathsf{P}(c_j \in A \mid \mathrm{ind}(O_j) = m)$$

Задача полягає в тому, щоб оцінити функцію розподілу $\xi(O)$ для $m$-ого компонента, тобто $F_m(x) = F_m((-\infty, x])$, за спостереженнями $(\xi_j^*, \delta_j)_{j=1}^n$. При цьому випадкові вектори $(\xi_j, c_j)$ вважаються незалежними при різних $j$, $\xi_j$ та $c_j$ незалежні при фіксованому $\mathrm{ind}(O_j)$. Концентрації $w_j^m$ $j = 1, \dots, n$, $m = 1, \dots, M$ вважаються відомими, розподіли $F_m$ та $G_m$ — невідомими.

Ми будемо досліджувати асимптотичну поведінку оцінок при зростанні обсягу вибірки: $n \to \infty$. При цьому спостережувані об'єкти $O_j$ та відповідні концентрації $w_j^k$ при різних $n$ ніяк між собою не пов'язані, тобто $\xi_j = \xi_{j;n}$, $w_j^m = w_{j;n}^m$. Для спрощення позначень індекс $n$ надалі писати не будемо.

Надалі функцію розподілу будемо позначати тією ж літерою, що і відповідний розподіл. Наприклад, $G_m(t) = G_m((-\infty, t])$, $G_m(t-) = \lim_{s \uparrow t} G_m(s) = G_m((-\infty, t))$ — неперервна зліва версія функції розподілу. Відповідна функція виживання позначається рискою вгорі: $\overline{G}_m(t) = 1 - G_m(t), \overline{F}_m(t) = 1 - F_m(t)$.

Позначимо також

$$N_m(t) = \mathsf{P}(\xi_j^* \le t, \delta_j = 1 \mid \mathrm{ind}(O_j) = m) = \int_{(0,t]} \overline{G}_m(s-) F_m(ds)$$

$$Y_m(t) = \mathsf{P}(\xi^* \ge t \mid \mathrm{ind}(O_j) = m) = \overline{G}_m(t-) \overline{F}_m(t-)$$

Тоді

$$\overline{F}_k(t) = \prod_{(0,t]} \left(1 - \frac{N_k(ds)}{Y_k(s)}\right), \tag{1}$$

де $\prod(1 + d\alpha)$ позначає продакт-інтеграл за диференціалом адитивної функції $\alpha$ (докладніше див. у п. 3).

Зафіксуємо $k$ — номер компонента, для якого проводиться оцінювання. Для того, щоб оцінити $\overline{F}_k(t)$ (і, відповідно, $F_k(t) = 1 - \overline{F}_k(t)$) ми оцінимо $N_k$ та $Y_k$ і підставимо відповідні оцінки у (1) замість справжніх значень. Як показано у [4], на роль оцінок $N_k$ та $Y_k$ можна використовувати відповідні навантажені емпіричні функції розподілу:

$$\hat{N}_{k,n}(t) = \frac{1}{n} \sum_{j=1}^n a_j^k \mathbb{1}\{\xi_j^* \le t, \delta_j = 1\}, \tag{2}$$

$$\hat{Y}_{k,n}(t) = \frac{1}{n} \sum_{j=1}^n a_j^k \mathbb{1}\{\xi_j^* \ge t\} \tag{3}$$

Тут $\mathbf{a}^k = (a_1^k, \dots, a_n^k)$ — невипадкові вагові коефіцієнти, що не залежать від спостережень, але залежать від концентрацій $w_j^m$. Для того, щоб оцінки (2) та (3) були незміщеними, досить, щоб

$$\frac{1}{n}\sum_{j=1}^{n} w_j^m a_j^k = \mathbb{1}\{k = m\}, \quad m = 1, \dots, M \tag{4}$$

Як показано у [4], у класі всіх незміщених оцінок мінімаксними є оцінки з ваговими коефіцієнтами

$$\mathbf{a}^k = \mathbf{e}^k \mathbf{\Gamma}_n^{-1} \mathbf{W}_n^\top, \tag{5}$$

де $\mathbf{e}^k = (\mathbb{1}\{k = \ell\})_{\ell=1}^M$ — вектор-рядок, $\mathbf{W}_n = (w_j^m)_{j=1\dots,n;m=1\dots,M}$ — $n \times M$ матриця концентрацій, $\mathbf{\Gamma}_n = (\frac{1}{n}\sum_{j=1}^n w_j^m w_j^l)_{m,l=1}^M$ (за умови, що $\det \mathbf{\Gamma}_n \neq 0$).

Надалі ми будемо розглядати переважно оцінки (2), (3) з ваговими коефіцієнтами вигляду (5), але теорема 2.1 є вірною для будь-яких вагових коефіцієнтів, що задовольняють її умови.

Таким чином, оцінка для $F_k(t)$ набуває вигляду:

$$\hat{F}_{k,n}(t) = 1 - \prod_{(0,t]}\left(1 - \frac{d\hat{N}_{k,n}(s)}{\hat{Y}_{k,n}(s)}\right) = 1 - \prod_{j:\xi_j^* \le t}\left(1 - \frac{a_j^k \delta_j}{n - \sum\limits_{i:\xi_i^* < t} a_i^k}\right) \tag{6}$$

Зауважимо, що у випадку, коли всі $\xi_j^*$ однаково розподілені, а $a_j^k = 1$, оцінка (6) перетворюється на звичайну оцінку Каплана-Мейєра для функції розподілу за однорідними цензурованими спостереженнями.

Позначимо $\ell_n = \sqrt{\frac{\ln n}{n}}$, $\overline{a}_n^k = \sup_{j=1,\dots,n}|a_j^k|$, $\tau_k = \sup\{t : \overline{F}_k(t)\overline{G}_k(t) > 0\}, k = 1, \dots, M$.

**Теорема 2.1.** *Нехай $T_k \in \mathbb{R} : 0 < T_k < \tau_k$ і для $a_j^k$ виконуються умови (4) та $\ell_n(\overline{a}_n^k)^2 \to 0$ при $n \to \infty$. Тоді існує така випадкова величина $0 < \eta < \infty$ м. н., що*

$$\sup_{t\in[0,T_k]} |\hat{F}_{k,n}(t) - F_k(t)| \le \eta \ell_n (\overline{a}_n^k)^3$$

**Наслідок 2.1.** *Якщо вагові коефіцієнти $a_j^k$ визначаються (5) і $\ell_n(\det \mathbf{\Gamma}_n)^{-3} \to 0$, то $\hat{F}_{k,n}(t)$ є рівномірно сильно консистентною оцінкою для $F_k(t)$ на відрізку $[0, T_k]$.*

Дійсно, оскільки ймовірності $0 \le w_j^m \le 1$, то для $a_j^k$ заданих (5), $\overline{a}_n^k \le C \det \Gamma_n$.

Відмітимо, що при $\det \mathbf{\Gamma}_n = 0$ вектори концентрацій компонентів $\mathbf{w}^m = (w_1^m, \dots, w_n^m)$, $m = 1, \dots, M$ є лінійно залежними. У цьому (виродженому) випадку задача оцінювання функцій розподілу компонентів стає неідентифіковною навіть при відсутності цензурування [4] і, отже, консистентне оцінювання неможливе. Тим не менше, наслідок показує, що оцінки $\hat{F}_{k,n}$ будуть консистентними, якщо $\det \mathbf{\Gamma}_n \to 0$ але не занадто швидко.

## 3. Допоміжні відомості

Слідуючи [9], нагадаємо поняття мультиплікативного інтегралу, яке буде використовуватися у доведенні теореми про сильну консистентність модифікованої оцінки Каплана-Мейєра. Розглянемо розбиття інтервалу $(0, t]$ та позначимо його через

$$P = \{0 = t_0 < t_1 < \ldots < t_{|P|} = t\} \tag{7}$$

Через $\mathcal{P}$ будемо позначати клас усіх розбиттів таких, що

$$\max_{t_i \in P} |t_{i+1} - t_i| \to 0, |P| \to \infty \tag{8}$$

**Означення 3.1.** Функція $\alpha(u,v)$, визначена для $0 \le u < v \le t$, називається адитивною функцією, якщо для всіх $0 \le u < v < s < \infty$

$$\alpha(u,s) = \alpha(u,v) + \alpha(v,s).$$

**Означення 3.2.** Нехай $\alpha$ — адитивна неперервна справа функція. Тоді мультиплікативним інтегралом називається границя:

$$\prod_{(0,t]}(1+\alpha(ds)) = \lim_{\substack{P\in\mathcal{P}:\\ |P|\to\infty}} \prod_{i=1}^{|P|}(1+\alpha(t_{i-1},t_i)) \tag{9}$$

Для довільних функцій $f, g : [0,t] \to \mathbb{R}$ позначимо

$$\prod_{(0,t]}(1+f(s)dg(s)) = \prod_{(0,t]}(1+d\alpha),$$

де $\alpha(u,v) = \int_u^v f(s)dg(s)$.

Надалі фіксуємо $T_k$, що відповідає умові теореми 2.1 і будемо позначати $\|x\|_\infty = \sup_{t\in[0,T_k]}|x(t)|$, $\|x\|_v$ — повна варіація $x(t)$ на інтервалі $[0,T_k]$.

**Лема 3.1.** *Нехай $\alpha, \alpha_n$ — адитивні функції. Покладемо $\mu_n(t) = \prod_{(0,t]}(1+d\alpha_n)$, $\mu(t) = \prod_{(0,t]}(1+d\alpha)$. Тоді*

$$\|\mu_n - \mu\|_\infty \le 4\|\mu\|_v \exp(\|\alpha_n\|_\infty)\|\alpha_n\|_v\|\alpha_n - \alpha\|_\infty$$

*Доведення* випливає з наступних нерівностей, отриманих у [9] при доведенні теореми 7:

$$\|\mu_n - \mu\|_\infty \le 4\|\mu_n\|_v\|\mu\|_v\|\alpha_n - \alpha\|_\infty$$

та

$$\|\mu_n\|_v \le \exp(\|\alpha_n\|_\infty)\|\alpha_n\|_v$$

□

## 4. Доведення теореми

Позначимо $\Lambda_k(s,t) = \int_{(s,t]} \frac{dN_k(u)}{Y_k(u)}$, $\Lambda_k(t) = \Lambda_k(0,t)$ і $\hat\Lambda_{k,n}(s,t) = \int_{(s,t]} \frac{d\hat N_{k,n}(u)}{\hat Y_{k,n}(u)}$, $\hat\Lambda_{k,n}(t) = \hat\Lambda_{k,n}(0,t)$. Функцію $\Lambda_k(t)$ називають (інтегральною або кумулятивною) функцією ризику для тривалості життя з розподілом $F_k$.

**Лема 4.1.** *Нехай для $a_j^k$ виконуються умови незміщеності (4). Тоді існує така випадкова величина $\eta_0 : 0 < \eta_0 < \infty$ м. н., що*

$$\|\hat N_{k,n} - N\|_\infty \le \eta_0 \ell_n \overline{a}_n^k$$
$$\|\hat Y_{k,n} - Y\|_\infty \le \eta_0 \ell_n \overline{a}_n^k$$

*Доведення.* розглянемо випадкові вектори $\zeta_j = (\xi_j^*, \delta_j)$. Легко бачити, що набір $(\zeta_1, \dots, \zeta_n)$ являє собою вибірку з суміші зі змінними концентраціями:

$$\mathbb{P}(\zeta_j \le \mathbf{z}) = \sum_{k=1}^m w_j^k H_k(\mathbf{z}),$$

де $\mathbf{z} = (z_1, z_2) \in \mathbb{R}^2$,

$$H_k(\mathbf{z}) = \mathbb{P}(\min(\xi(O_j), c_j) \le z_1, \mathbb{1}\{\xi(O_j) \le c_j\} \le z_2 \mid \mathrm{ind}(O_j) = k)$$

— функція розподілу $\zeta_j$ за умови, що $O_j$ належить $k$-ому компоненту суміші. Застосувавши до $\zeta_j$ наслідок 2.2.4 з [5], отримуємо, що існує така випадкова величина $\eta_1 < \infty$ м. н., для якої

$$\sup_{\mathbf{z}\in\mathbb{R}^2} |\hat H_{k,n}(\mathbf{z}) - H_k(\mathbf{z})| \le \eta_1 \ell_n \overline{a}_n^k, \tag{10}$$

де $\hat{H}_{k,n}(\mathbf{z}) = \frac{1}{n}\sum_{j=1}^{n} a_j^k \mathbb{1}\{\zeta_j \le \mathbf{z}\}$. Оскільки $\hat{Y}_{k,n}(t) = 1 - \hat{H}_{k,n}(t,1)$, $Y(t) = 1 - H_k(t,1)$, $\hat{N}_{k,n}(t) = \hat{H}_{k,n}(t,1) - \hat{H}_{k,n}(t,0)$, $N_k(t) = H_k(t,1) - H_k(t,0)$, з (10) отримуємо твердження леми. □

**Лема 4.2.** *В умовах теореми 2.1 існують такі випадкові величини $\eta_2 : \eta_2 < \infty$ м. н. і $n_0 < \infty$ м.н., що для всіх $n > n_0$*

$$\|\hat{\Lambda}_{k,n} - \Lambda_k\|_\infty \le \eta_2 \ell_n \overline{a}_n^k$$

*Доведення.* За означенням $\Lambda$ і $\hat{\Lambda}$ отримуємо

$$\|\hat{\Lambda}_{k,n} - \Lambda_k\|_\infty = \sup_{t\in[0,T]} \left| \int_{(0,t]} \frac{d\hat{N}_{k,n}(s)}{\hat{Y}_{k,n}(s)} - \int_{(0,t]} \frac{dN_k(s)}{Y_k(s)} \right| \le I_1 + I_2, \tag{11}$$

де

$$I_1 = \sup_{t\in[0,T]} \left| \int_{(0,t]} \left( \frac{1}{\hat{Y}_{k,n}(s)} - \frac{1}{Y_k(s)} \right) d\hat{N}_{k,n}(s) \right|,$$

$$I_2 = \sup_{t\in[0,T]} \left| \int_{(0,t]} \frac{d(\hat{N}_{k,n}(s) - N_k(s))}{Y_k(s)} \right|$$

Оцінимо $I_1 \le \|\frac{\hat{Y}_{k,n} - Y_k}{\hat{Y}_{k,n} Y_k}\|_\infty \|\hat{N}_{k,n}\|_v$. Оскільки $T_k < \tau_k$, $\sigma_k = \inf\{t \in [0, T_k] \mid Y_k > 0\}$ і за лемою 4.1 отримуємо, що $\inf_{t\in[0,T]} |\hat{Y}_{k,n}(t)| > \sigma_k + \eta_0 \ell_n \overline{a}_n^k$. Отже, при достатньо великих $n$,

$$\|\frac{\hat{Y}_{k,n} - Y_k}{\hat{Y}_{k,n} Y_k}\|_\infty \le \frac{\|\hat{Y}_{k,n} - Y_k\|_\infty}{\sigma_k(\sigma_k - \eta_0 \ell_n \overline{a}_n^k)} \le \frac{\eta_0 \ell_n \overline{a}_n^k}{\sigma_k(\sigma_k - \eta_0 \ell_n \overline{a}_n^k)}$$

Функція $\hat{N}_{k,n}$ є сталою на інтервалах між стрибками зі стрибками висоти $\frac{1}{n} a_j^k$. Тому

$$\|\hat{N}_{k,n}\|_v \le \frac{1}{n} \sum_{j=1}^{n} |a_j^k| \le \overline{a}_n^k$$

Таким чином,

$$I_1 \le \frac{\eta_0 \ell_n (\overline{a}^k)_n^2}{\sigma_k(\sigma_k - \eta_0 \ell_n \overline{a}_n^k)} \tag{12}$$

Оцінимо $I_2$ інтегруванням частинами:

$$I_2 = \sup_{t\in[0,T_k]} \left| \frac{\hat{N}_{k,n}(s) - N_k(s)}{Y_k(s)} \right|_{s=0+}^{t} - \int_{(0,t]} (\hat{N}_{k,n}(s) - N_k(s)) d\frac{1}{Y_k(s)} \Bigg|$$
$$\le 2\frac{\|\hat{N}_{k,n} - N_k\|_\infty}{\sigma_k} + \|\hat{N}_{k,n} - N_k\|_\infty \|\frac{1}{Y_k}\|_v$$

Оскільки $Y_k$ є монотонно неспадною функцією і $Y_k > \sigma_k$ при $t \in [0, T_k]$, то $\|\frac{1}{Y_k}\|_v < \infty$.

Отже, за лемою 4.1

$$I_2 \le \frac{2}{\sigma_k} \eta_0 \ell_n \overline{a}_n^k + \eta_0 \ell_n \overline{a}_n^k \|\frac{1}{Y_k}\|_v \tag{13}$$

З умови (4) випливає, що $\overline{a}_n^k \ge 1$. Тому, об'єднуючи (11), (12), (13), отримуємо твердження леми. □

Оскільки

$$F_k(t) = 1 - \prod_{(0,t]} (1 - d\Lambda_k(s)),$$

$$\hat{F}_{k,n}(t) = 1 - \prod_{(0,t]} (1 - d\hat{\Lambda}_{k,n}(s)),$$

то за лемою 3.1 отримуємо

$$\|\hat{F}_{k,n} - F_k\|_\infty \le 4\|F_k\|_v \exp(\|\hat{\Lambda}_{k,n}\|_\infty)\|\hat{\Lambda}_{k,n}\|_v\|\hat{\Lambda}_{k,n} - \Lambda_k\|_\infty$$

Зрозуміло, що $\|F_k\|_v \le 1$, $\|\hat{\Lambda}_{k,n}\|_\infty \le \|\Lambda_k\|_\infty + \eta_2\ell_n(\overline{a}_n^k)^2$ (за лемою 4.2) і

$$\|\hat{\Lambda}_{k,n} - \Lambda_k\|_\infty \le \eta_2\ell_n(\overline{a}^k)^2.$$

Оцінимо $\|\hat{\Lambda}_{k,n}\|_v$ — повну варіацію $\hat{\Lambda}_{k,n}(t) = \int_{(0,t]} \frac{d\hat{N}_{k,n}(t)}{\hat{Y}_{k,n}(t)}$. Це стала функція на інтервалах між стрибками зі стрибками висоти

$$\frac{1}{n}\frac{a_j^k}{\hat{Y}_{k,n}(\xi_j^*)} \le \frac{1}{n}\frac{\overline{a}^k}{\sigma_k},$$

отже $\|\hat{\Lambda}_{k,n}\|_v \le \frac{\overline{a}_n^k}{\sigma_k}$. Тому

$$\|\hat{F}_{k,n} - F_k\|\infty \le 4\exp(\|\Lambda_k\|_\infty + \eta_2\ell_n(\overline{a}_n^k)^2)\frac{\overline{\mathbf{a}}^k}{\sigma_k}\eta_2\ell_n(\overline{a}_n^k)^2.$$

Враховуючи, що, за умовою теореми, $\ell_n(\overline{a}_n^k)^2 \to \infty$, отримуємо твердження теореми. □

## 5. Результати імітаційного моделювання

Для перевірки поведінки запропонованих оцінок на вибірках помірного обсягу було проведене невелике дослідження методом імітаційного моделювання. Розглядалась модель суміші двох компонентів. Перший компонент мав хі-квадрат розподіл з трьома ступенями вільності, другий — півнормальний розподіл. Для обох компонентів розподіл цензора був однаковим — експоненційним з параметром $\lambda = 0.1$. Концентрації компонентів у суміші дорівнювали $w_j^1 = \frac{j}{n}$ та $w_j^2 = 1 - \frac{j}{n}$ $(j = 1, \dots, n)$ відповідно. Для кожного розглянутого обсягу вибірки $n$ було згенеровано 1000 вибірок з суміші зі змінними концентраціями, за якими проводилось оцінювання розподілів першого та другого компонента з використанням оцінки $\hat{F}_{k,n}(t)$, визначеної 6.

Для характеризації якості оцінювання використовуються два підходи. При першому підході оцінка обчислюється у фіксованій точці $t = 1.85$ і підраховується медіана (Median) її відхилення від справжнього значення функції розподілу та інтерквартильний розмах (IQR). (Ці робастні міри середнього положення та розкиду використані для того, щоб усунути вплив викидів, що виникають при малих обсягах вибірок). При другому підході для кожної модельованої вибірки обчислюється супремум різниці між оцінкою та справжньою функцією розподілу і знаходиться медіана цих супремумів по всіх вибірках (Med-Sup).

Результати моделювання наведені у таблиці.

Результати моделювання

| | Перший компонент | | | Другий компонент | | |
|---|---|---|---|---|---|---|
| n | Median | IQR | Med-Sup | Median | IQR | Med-Sup |
| 100 | 0,00891 | 0,13706 | 0,18477 | 0,01178 | 0,13682 | 0,17725 |
| 250 | 0,00573 | 0,08855 | 0,11843 | 0,00601 | 0,08707 | 0,11373 |
| 500 | 0,00444 | 0,06563 | 0,08587 | 0,00018 | 0,05965 | 0,07835 |
| 1000 | 0,00339 | 0,04426 | 0,06088 | 0,00072 | 0,04428 | 0,05666 |
| 2000 | 0,00063 | 0,03114 | 0,04429 | -0,0003 | 0,03106 | 0,04010 |

Ці результати свідчать про досить швидку збіжність оцінок до справжніх значень оцінюваних функцій.

## 6. Висновки

Ми побудували модифікацію оцінки Каплана-Мейера для оцінювання функцій розподілу компонентів суміші зі змінними концентраціями за цензурованими даними і довели її консистентність за досить широких умов. Результати моделювання свідчать про достатньо хорошу поведінку оцінки при помірних обсягах вибірки. Можна сподіватись, що техніка, запропонована у [9] для доведення асимптотичної нормальності звичайних оцінок Каплана-Мейера за однорідними вибірками, дозволить отримати аналогічний результат і для нашої модифікації. Це має бути предметом подальших досліджень.

## Література

[1] Korosteleva O. *Clinical Statistics: Introducing Clinical Trials, Survival Analysis, and Longitudinal Data Analysis*, Jones and Bartlett Publishers: Sudbury, MA, 2008, 120 p.

[2] Huber C., Limnios N., Mesbah M., Nikulin M. *Mathematical methods in survival analysis, reliability and quality of life* ISTE & Wiley, London & Hoboken, NJ, 2008, 370p.

[3] Shao J. *Mathematical statistics.* - Springer-Verlag: New York, 1998. - 530 p.

[4] Maiboroda R., Sugakova O. Statistics of mixtures with varying concentrations with application to DNA microarray data analysis, *Journal of Nonparametric Statistics*, 2012,24:1, 201-215

[5] Майборода Р.Є., Сугакова О.В. *Оцінювання та класифікація за спостереженнями із суміші.* - К.: ВПЦ "Київський ун-т 2008. - 213 с.

[6] Рижов А.Ю. Оцінки розподілів компонент суміші по цензурованим даним.- *Теорія ймовірностей та математична статистика.*— 2003, Вип. 69.— с.154-161.

[7] Щербіна А.М. Оцінювання середнього у моделі сумішей зі змінними концентраціями// Теорія ймовірностей та математична статистика.– 2011.– Т. 84.– С. 142-154.

[8] Autin F., Pouet C. Minimax rates over Besov spaces in ill-conditioned mixture-models with varying mixing-weights// Journal of Statistical Planning and Inference V 146, 2014, p. 20–30.

[9] Gill R. D., Johansen S., A Survey of Product-Integration with a View Toward application in Survival Analysis, *Ann. Statist.* 1990, 18 N 4, 1501-1555.

Кафедра теорії ймовірностей, математичної статистики і актуарної математики механіко-математичний факультет, Національний університет ім. Тараса Шевченка, проспект Глушкова, 6, Київ 03127, Україна

*E-mail address*: mre@univ.kiev.ua

Кафедра теорії ймовірностей, математичної статистики і актуарної математики механіко-математичний факультет, Національний університет ім. Тараса Шевченка, проспект Глушкова, 6, Київ 03127, Україна

*E-mail address*: vl.khizanov@gmail.com