UO‘K 519.688 (519.24)

MONTE-KARLO SIMULYATSIYA TIZIMIDA YASHOVCHANLIK TAHLILI UCHUN ISHONCHLI VA SAMARALI TANLANMA GENERATSIYA QILISH ALGORITMLARI

Mansurov Dilshod Ravilovich,
Navoiy davlat universiteti
Matematika kafedrasi dotsenti, Navoiy, O‘zbekiston
dmansurovmath@gmail.com
Bozorov Suxrob Baxodirovich,
Guliston davlat universiteti
Matematika kafedrasi dotsenti, Guliston, O‘zbekiston
suxrobbek_8912@mail.ru
Oltiyev Azizbek Bayramovich,
Navoiy davlat universiteti
Matematika kafedrasi o‘qituvchisi, Navoiy, O‘zbekiston
azizbekolot1993@gmail.com

Annotatsiya. *Ushbu maqolada senzurlanish ulushlarini aniq boshqarish imkonini beruvchi senzurlangan tanlanmalarni generatsiya qilishga mo‘ljallangan keng qamrovli Monte-Karlo simulyatsiya tizimi taklif etiladi. Tizim teskari akslantirish usuli yordamida o‘ng, chap, interval va ikki tomonlama tasodifiy senzurlanish mexanizmlarini tizimli ko‘rib chiqiladi. Natijada, tadqiqotchilarga klinik va epidemiologik amaliyotni aks ettiruvchi xilma-xil senzurlanish modellarini shakllantirish imkoniyatini beradi. Senzurlanish taqsimoti parametrlarini kalibrlash uchun deterministik algoritmlar ishlab chiqilgan bo‘lib, absolyut og‘ish 0,005 dan past bo‘lgan holda maqsadli senzurlanish ulushlariga erishiladi. Modulli dasturiy arxitektura turli senzurlanish turlariga muammosiz moslashishni ta’minlaydi. 10000 takrorlash bo‘yicha empirik tekshiruv shuni ko‘rsatadiki, taklif etilayotgan optimallashtirilgan usullar sodda usullardan 15-40 marta tezroq ishlaydi va 1000000 ta kuzatuvni 3 sekunddan kam vaqtda generatsiya qiladi. Tizim nazariy yashovchanlik tahlili va amaliy tahlil o‘rtasidagi bo‘shliqni to‘ldiradi hamda yashovchanlik modellarni taqqoslash uchun standartlashtirilgan, takrorlanuvchi va hisoblash jihatidan samarali vosita sifatida xizmat qiladi.*

Kalit so‘zlar: *yashovchanlik tahlili, senzurlangan ma’lumotlar, Monte-Karlo simulyatsiyasi, teskari akslantirish usuli, o‘ng senzurlanish, interval senzurlanish, ma’lumotlar generatsiyasi algoritmi.*

RELIABLE AND EFFICIENT SELECTION GENERATION ALGORITHMS FOR SURVIVAL ANALYSIS IN MONTE CARLO SIMULATION SYSTEM

Abstract. *This article proposes a comprehensive Monte Carlo simulation framework designed to generate censored samples with precise control over censoring proportions. The system systematically addresses right-, left-, interval-, and two-sided random censoring mechanisms using the inverse transformation method. As a result, it enables researchers to construct diverse censoring schemes that reflect clinical and epidemiological practice. Deterministic algorithms are developed to calibrate the parameters of the censoring distribution, achieving target censoring proportions with an absolute deviation of less than 0.005. The modular software architecture ensures seamless adaptability to various types of censoring. Empirical evaluation based on 10,000 replications demonstrates that the proposed optimized methods are 15–40 times faster than naive approaches and can generate 1,000,000 observations in under 3 seconds. The framework bridges the gap between theoretical survival analysis and practical application, serving as a standardized, reproducible, and computationally efficient tool for comparing survival models.*

Key words: *survival analysis, censored data, Monte Carlo simulation, inverse transform method, right censoring, interval censoring, data generation algorithm.*

АЛГОРИТМЫ НАДЁЖНОГО И ЭФФЕКТИВНОГО ГЕНЕРИРОВАНИЯ ВЫБОРОК ДЛЯ АНАЛИЗА ВЫЖИВАЕМОСТИ В СИСТЕМЕ МОДЕЛИРОВАНИЯ МОНТЕ-КАРЛО

Аннотация. *В данной статье предлагается комплексная система имитационного моделирования методом Монте-Карло, предназначенная для генерации цензурированных выборок с точным контролем доли цензурирования. Система систематически рассматривает механизмы*

правого, левого, интервального и двустороннего случайного цензурирования с использованием метода обратного преобразования. В результате исследователи получают возможность формировать разнообразные схемы цензурирования, отражающие клиническую и эпидемиологическую практику. Для калибровки параметров распределения цензурирования разработаны детерминированные алгоритмы, позволяющие достигать заданных долей цензурирования с абсолютным отклонением менее 0,005. Модульная программная архитектура обеспечивает гибкую адаптацию к различным типам цензурирования. Эмпирическая проверка на основе 10 000 повторений показывает, что предлагаемые оптимизированные методы работают в 15–40 раз быстрее простых подходов и позволяют генерировать 1 000 000 наблюдений менее чем за 3 секунды. Предложенная система устраняет разрыв между теоретическим анализом выживаемости и практическими приложениями, выступая в качестве стандартизированного, воспроизводимого и вычислительно эффективного инструмента для сравнения моделей выживаемости.

Ключевые слова: *анализ выживаемости, цензурированные данные, моделирование методом Монте-Карло, метод обратного преобразования, правостороннее цензурирование, интервальное цензурирование, алгоритм генерации данных.*

Kirish. Yashovchanlik tahlili tibbiy tadqiqotlarda, muhandislik ishonchliligida va ijtimoiy fanlarda muhim rol o‘ynaydi, bu sohada asosiy e’tibor voqea yuz vaqti ma’lumotlarini modellashtrishga qaratiladi. Bunday tahlillardagi asosiy muammo senzurlanish mavjud ekanligidir. Ya’ni tadqiqot loyihasidagi ayrim cheklovlar tufayli yoki ishtirokchilarning chiqib ketishi sababli voqea vaqti faqat qisman kuzatiladi. Senzurlangan kuzatuvlarning mavjudligi statistik xulosalar chiqarishni murakkablashtiradi va to‘liq bo‘lmagan ma’lumotlarni hisobga oladigan maxsus usullarni talab etadi. Kaplan–Meyer bahosi [1] va Koks proporsional xavf modeli [2] kabilar shular jumlasidandir. Bu yondashuvlar amaliyotda keng qo‘llaniladi, ammo ularning turli senzurlanish mexanizmlari ostidagi ish samaradorligini simulyatsiya tadqiqotlari orqali chuqur baholash zarur.

Monte-Karlo simulyatsiyalari yashovchanlik modellarni tasdiqlashda asosiy vosita bo‘lib xizmat qiladi. Ular tadqiqotchilarga nazorat ostidagi sharoitlarda baholarning xossalarini o‘rganish imkonini beradi. Biroq, aniq parametrlarga ega senzurlangan yashovchanlik ma’lumotlarini generatsiya qilish hali ham murakkab masala bo‘lib qolmoqda. Mavjud usullar ko‘pincha muayyan senzurlanish turlariga, masalan, klinik sinovlardagi o‘ng senzurlanish yoki davriy kuzatuv tadqiqotlaridagi interval senzurlanishga [3] yo‘naltirilgan. Bu yondashuvlar qimmatli bo‘lsa-da, ular hisoblash samaradorligi va takrorlanish imkoniyatini ta’minlab, bir nechta senzurlanish mexanizmlarini qamrab oluvchi yagona tizimga ega emas. Bu kamchilik, ayniqsa, xilma-xil senzurlanish naqshlariga bardoshlilik da’vo qiluvchi yangi metodologiyalarni ishlab chiqish zarurligini ko‘rsatadi.

Usullar. Ushbu bo‘lim simulyatsiya algoritmlarining poydevorini tashkil etuvchi matematik tizimni va nazorat ostidagi xossalarga ega senzurlangan yashovchanlik ma’lumotlarini generatsiya qilishning algoritmik amalga oshirishini batafsil bayon etadi.

Matematik asoslar. Yashovchanlik funksiyasi $S_T(t)$ biror shaxsning t vaqtdan uzoqroq yashashi ehtimolligini ifodalaydi:

$$S_T(t) = P(T > t), \tag{1}$$

bu yerda T voqea vaqtini ifodalovchi manfiy bo‘lmagan tasodifiy o‘zgaruvchidir. Taqsimot funksiyasi (TF) $F_T(t)$ yashovchanlik funksiyasini to‘ldiradi:

$$F_T(t) = P(T \le t) = 1 - S_T(t). \tag{2}$$

t vaqtgacha yashaganlik sharti ostida shu vaqtda voqea sodir bo‘lishining oniy tezligini tavsiflovchi xavf funksiyasi (hazard function) $h(t)$ quyidagicha aniqlanadi:

$$h(t) = \frac{f_T(t)}{S_T(t)}, \tag{3}$$

bu yerda $f_T(t)$ — T ning zichlik funksiyasi. Ushbu munosabatlar teskari akslantirish usuli orqali yashovchanlik vaqtlarini generatsiya qilishning asosini tashkil etadi.

Indikator funksiyasi senzurlanish voqealarini ifodalashda muhim rol o‘ynaydi. Ixtiyoriy A hodisasi uchun:

$$I(A) = \begin{cases} 1; & A \ ro'ybersa; \\ 0; & A \ ro'ybermasa. \end{cases} \tag{4}$$

O‘ng senzurlangan ma’lumotlar uchun kuzatilgan ma’lumotlar $(Y; \delta)$ ko‘rinishida ifodalanadi, bu yerda $Y = min(T; C)$ va $\delta = I(T \le C)$, C esa senzurlanish vaqtini bildiradi.

Teskari akslantirish usuli bir xil taqsimot asosida amalga oshiriladi. Usul TF si $F_T(t)$ bo‘lgan ixtiyoriy uzluksiz T tasodifiy o‘zgaruvchisi uchun $U = F_T(T)$ tasodifiy o‘zgaruvchisi standart bir xil taqsimlashga bo‘ysunishidan foydalanadi. Bu esa yashovchanlik vaqtlarini quyidagicha generatsiya qilish imkonini beradi:

$$T = F_T^{-1}(U),\ U \sim Uniform(0;1), \quad (5)$$

bu yerda $F_T^{-1}(U)$ — maqsadli taqsimotning kvantil funksiyasi. Uniform-tekis taqsimot. TF qat’iy o‘suvchi bo‘lmagan holda umumlashtirilgan teskari funksiyadan foydalaniladi:

$$F_T^{-1}(u) = \inf\{t: F_T(t) \geq u\}. \quad (6)$$

Mustaqil senzurlanish asosiy shart hisoblanadi:

$$P(T \leq t, C \leq c) = P(T \leq t)P(C \leq c). \quad (7)$$

Bu senzurlanish mexanizmi kuzatilgan ma’lumotlardan tashqari yashovchanlik vaqti haqida qo‘shimcha ma’lumot bermasligini anglatadi. Simulyatsiya tadqiqotlarida T va C ushbu shartni qanoatlantirishlari uchun o‘zlarining mos taqsimlashlaridan mustaqil ravishda generatsiya qilinadi.

Bir nechta senzurlanish turlari uchun yagona tizim. Taklif etilayotgan tizim to‘rtta asosiy senzurlanish mexanizmi ostida yashovchanlik ma’lumotlarini generatsiya qilishning tizimli yondashuvini o‘rnatadi. 1-jadvalda har bir tur uchun kuzatiladigan ma’lumotlar tuzilmalari va senzurlanish ulushi formulalari jamlangan.

1-jadval.

Senzurlanish mexanizmlari, kuzatiladigan ma’lumotlar tuzilmalari va senzurlanish ulushi formulalarining xulosasi

Senzurlanish	Kuzatilgan	Indikator	Ulush
O‘ng	$Y = min(T; C)$	$\delta = I(T \leq C)$	$p_r = \int_0^\infty S_T(c) f_C(c) dc$
Chap	$Y = max(T; L)$	$\delta = I(T \geq C)$	$p_l = \int_0^\infty F_T(l) f_L(l) dl$
Intervallik	$(L; U; \delta)$	$\delta = I(L < T < U)$	$p_i = \int_0^\infty \int_l^\infty [F_T(u) - F_T(l)]\, f_{L,U} du dl$
Tasodifiy	$(Y; \delta)$	(11)-tengl.	p_r va p_l kombinatsiyasi

O‘ng senzurlanish. T yashovchanlik vaqti f_T zichligi bilan F_T taqsimotga bo‘ysunadi, C senzurlanish vaqti esa F_C taqsimotdan mustaqil generatsiya qilinadi. Senzurlanish ulushi p_r quyidagi tenglamani yechish orqali nazorat qilinadi:

$$p_r = P(T > C) = \int_0^\infty S_T(c) f_C(c) dc, \quad (8)$$

bu yerda $S_T(t) = 1 - F_T(t)$. Ushbu integral tenglama maqsadli p_r ga erishish uchun F_C ning parametrlariga nisbatan sonli usulda yechiladi.

Chap senzurlanish. Voqea biror kuzatiladigan L vaqtdan oldin sodir bo‘lib, $(Y; \delta)$ ko‘rinishidagi ma’lumotlar hosil qilinadi: $Y = max(T; L)$, $\delta = I(T > C)$. senzurlanish ulushi p_l quyidagiga teng:

$$p_l = P(T < L) = \int_0^\infty F_T(l) f_L(l) dl, \quad (9)$$

bu yerda F_L funksiya L ning taqsimoti. Tizim (8)-tenglamani sonli usullar orqali yechib F_L ning parametrlarini istalgan p_l ga moslashtirib sozlaydi.

Interval senzurlanish. $\delta = I(L < T < U)$ bo‘lgan $(L; U; \delta)$ uchliklar generatsiya qilinadi. senzurlanish ulushi p_i quyidagicha:

$$p_i = \int_0^\infty \int_l^\infty [F_T(u) - F_T(l)]\, f_{L,U} du dl \quad (10)$$

bunda $(L; U)$ $F_{L,U}$ taqsimotdan birgalikda tanlanadi. Algoritm p_i ni nazorat qilish uchun interval kengligini $U - L$ kalibrlaydi.

Ikki tomonlama tasodifiy senzurlanish. Bu tur umumlashtirilgan senzurlanish chegaralari orqali o‘ng va chap senzurlanishni birlashtiradi:

$$Y = \begin{cases} T; & T \in (C_l; C_r) \ \ bo'lsa; \\ C_l; & T \leq C_l \ \ bo'lsa; \\ C_r; & T \geq C_r \ \ bo'lsa; \end{cases} \quad (11)$$

bu yerda C_l va C_r mos ravishda chap va o‘ng senzurlanish vaqtlarini bildiradi. Tizim T va $(C_l; C_r)$ o‘rtasidagi bog‘liqlik tuzilmalarini kopula modellari yoki shartli taqsimotlar orqali belgilash imkonini beradi.

Senzurlanish ulushlarini nazorat qilish uchun deterministik algoritmlar. Senzurlanish ulushlarini aniq nazorat qilish maqsadida deterministik algoritmlar senzurlanish taqsimoti parametrlarini

maqsadli yashovchanlik taqsimotiga asoslanib sozlaydi. p maqsadli ulush va F_T yashovchanlik taqsimoti bilan o‘ng senzurlanish uchun quyidagi tenglama yechiladi:

$$P(T > C) = \int_0^\infty S_T(c) dF_C(c) = p. \quad (12)$$

Faraz qilaylik, C tasodifiy miqdor λ parametrli eksponensial taqsimotga bo‘ysunsa, tenglama quyidagi ko‘rinishga keladi:

$$\int_0^\infty S_T(c) \lambda e^{-\lambda c} dc = p, \quad (13)$$

bu tenglama Brent usuli kabi ildiz topish algoritmlari yordamida λ ga nisbatan sonli usulda yechiladi. Yechim katta tanlamalarda senzurlangan kuzatuvlarning kutilgan ulushi p ga teng bo‘lishini kafolatlaydi.

$U = L + W$, $W \sim F_W$ bo‘lgan $(L; U)$ kuzatuv intervaliga ega intervallik senzurlanish uchun:

$$p = \int_0^\infty \int_0^\infty [F_T(l + w) - F_T(l)] d F_L(l) dF_W(w) \quad (14)$$

Algoritm F_L ni o‘zgartirib qolmasdan F_W parametrlarini topadi va kuzatuv intervallari ichida voqea ehtimolligi maqsadli p ga mos kelishini ta’minlash uchun iterativ sonli usullarni qo‘llaydi.

2.4 Umumiy simulyatsiya algoritmi. To‘liq simulyatsiya jarayonining bayoni quyida keltirilgan. $T-\theta$ parametrlari bilan F_T taqsimotdan generatsiya qilingan haqiqiy yashovchanlik vaqti, C esa ϕ bilan parametrlashtirilgan F_C taqsimotga ega senzurlanish vaqtini ifodalaydi.

1-Algoritm: Senzurlanish ulushlarini nazorat qilish bilan senzurlangan yashovchanlik ma’lumotlarini generatsiya qilish.

Kirish: Maqsadli senzurlanish ulushi p, yashovchanlik taqsimoti $F_T(\theta)$, senzurlanish oilasi F_C, tanlanma hajmi n, senzurlanish turi. $\tau = \{\text{o‘ng}, \text{chap}, \text{interval}, \text{ikki tomonlama tasodifiy}\}$.

Chiqish: senzurlangan tanlanma. $\mathcal{D} = \{(Y_i; \delta_i)\}_{i=1}^n$.

1-qadam (Kalibrlash): Nazariy senzurlanish ulushi p ga teng bo‘lishi uchun ϕ topiladi (12-formulaga qarang).

2-qadam (Generatsiya): Har bir $i = 1, \dots, n$ uchun:

✓ $U_i \sim Uniform(0; 1)$ generatsiya qilinadi; $T_i = F_T^{-1}(U_i; \theta)$ hisoblanadi.

✓ $V_i \sim Uniform(0; 1)$ generatsiya qilinadi; $C_i = F_C^{-1}(V_i; \phi)$ hisoblanadi.

✓ $\tau =$o‘ng bo‘lsa: $Y_i = min(T_i; C_i)$, $\delta_i = I(T_i \le C_i)$ belgilanadi.

✓ $\tau =$ chap bo‘lsa: $Y_i = max(T_i; C_i)$, $\delta_i = I(T_i \ge C_i)$ belgilanadi.

✓ $\tau =$interval bo‘lsa: $(L_i; U_i)$ generatsiya qilinadi; $\delta_i = I(L_i < T_i < U_i)$ belgilanadi.

3-qadam (Tekshiruv): $\hat{p} = 1 - \frac{1}{n}\sum_{i=1}^n \delta_i$ hisoblanadi; $|\hat{p} - p| < \varepsilon$ tekshiriladi.

Qaytarish $\mathcal{D}$.

Misol uchun, k shakl va λ masshtab parametrlariga ega Weibull yashovchanlik vaqtlari uchun kvantil funksiyasi quyidagi ko‘rinish oladi:

$$F_T^{-1}(u) = \lambda\big(-log(1-u)\big)^{1/k}. \quad (15)$$

Hisoblash samaradorligi strategiyalari. Statistik aniqlikni saqlab qolgan holda hisoblash vaqtini qisqartiruvchi bir necha optimallashtirish strategiyalari qo‘llaniladi.

Vektorlashtirilgan amallar. Har bir kuzatuvni alohida qayta ishlash o‘rniga, teskari taqsimot funksiya bir vaqtning o‘zida bir xil tasodifiy o‘zgaruvchilar massiviga tatbiq etiladi:

$$T = F_T^{-1}(U), \ U \sim Uniform(0; 1)^n, \quad (16)$$

bu yerda T va $U-n$ o‘lchovli vektorlar. Bu zamonaviy protsessor arxitekturalaridan foydalanib sezilarli tezlik oshishiga olib keladi.

Oldindan hisoblash va interpolyatsiya. Senzurlanish parametrlarini sonli usullar bilan hisoblashda, maqsadli ulushlarning panjara qiymatlari uchun yechimlar oldindan hisoblanib qidiruv jadvalida saqlanadi:

$$\phi_p \approx \text{interpolyatsiya}\left(\phi_{p_1}; \phi_{p_2}\right), \ p_1 < p < p_2, \quad (17)$$

bu esa odatda 0,5% dan past interpolyatsiya xatolari bilan hisoblash yukini kamaytiradi.

Parallel qayta ishlash. Algoritm talab qilinadigan N ta to‘plamni K ta protsessor yadrosiga taqsimlaydi. Har bir yadro yadrolar o‘rtasida aloqa bo‘lmagan holda mustaqil takrorlashlarni generatsiya qiladi, bu esa vaqtni taxminan $1/K$ koeffitsiyentga qisqartiradi.

Xotira tejamkorligini ta’minlovchi bo‘laklash. Juda katta to‘plamlar ($n > 10^6$) uchun kuzatuvlar belgilangan o‘lchamdagi bo‘laklarda qayta ishlanadi:

$$T_b = F_T^{-1}(U_b), \quad b = 1, \dots, [n/B], \tag{18}$$

bu yerda U_b bo‘lagi B ta bir xil tasodifiy o‘zgaruvchilarni o‘z ichiga oladi. Bu mavjud operativ xotiradan ortiqcha to‘plamlarni simulyatsiya qilish imkonini beradi.

Stratifikatsiyalangan interval namunasi. Interval senzurlanishda avval interval kengliklari F_W dan tanlanadi, so‘ng ular kenglikka shart qo‘yilgan holda joylashtiriladi:

$$W_i \sim F_W, \quad L_i \sim Uniform(0,\ \tau - W_i), \tag{19}$$

bu yerda τ — maksimal kuzatuv vaqti. Bu yondashuv takrorlashlar bo‘yicha senzurlanish ulushlarining dispersiyasini kamaytiradi.

Natijalar. Taklif etilayotgan tizim uchta muhim jihatni: senzurlanish ulushi aniqligini va hisoblash samaradorligini va yashovchanlik model baholashga ta’sirini baholash uchun mo‘ljallangan keng qamrovli simulyatsiya tadqiqotlari orqali qat’iy tekshirishdan o‘tkazish kabilarni o‘z ichiga oladi. Yashovchanlik taqsimotlari, senzurlanish mexanizmlari va tanlanma hajmlarini o‘zgartirish mumkin bo‘lgan ko‘p omilli tayyor algoritm ishlab chiqildi.

Senzurlanish ulushini nazorat qilish. O‘ng senzurlanish senariylari uchun eksponensial ($\lambda = 1$) va Weibull ($k = 1{,}5;\ \lambda = 1$) yashovchanlik taqsimotlari ostida $p \in \{0{,}1; 0{,}3; 0{,}5; 0{,}7\}$ maqsadli senzurlanish ulushlariga ega to‘plamlar generatsiya qilindi. Empirik senzurlanish ulushi quyidagicha hisoblandi:

$$\hat{p} = 1 - \frac{1}{n}\sum_{i=1}^{n} \delta_i. \tag{20}$$

2-jadval $n = 1000$ tanlanma hajmi uchun 10000 takrorlash bo‘yicha natijalarni ko‘rsatadi va barcha senariylarda mutlaq og‘ish 0,005 dan past bo‘lgan aniq nazoratni namoyish etadi.

2-jadval.

$n = 1000$ bilan 10000 takrorlash bo‘yicha empirik senzurlanish ulushi ($\hat{p}$). Qiymatlar o‘rtacha $\pm$ standart og‘ish ko‘rinishida keltirilgan

Maqsadli p	**Eksponensial ($\lambda = 1$)**		**Weibull ($k = 1{,}5;\ \lambda = 1$)**	
(lr)2-3 (lr)4-5	O‘rtacha $\hat{p}$	Standart og‘ish	O‘rtacha $\hat{p}$	Standart og‘ish
0,10	0,1003	0,0095	0,1001	0,0094
0,30	0,3002	0,0145	0,2998	0,0144
0,50	0,5001	0,0158	0,4997	0,0157
0,70	0,6998	0,0145	0,7002	0,0146

Hisoblash samaradorligini taqqoslash. Taklif etilayotgan optimallashtirilgan usullar turli tanlanma hajmlarida asos sifatida qabul qilingan siklga asoslangan usul bilan taqqoslandi. 3-jadval ishlash vaqtini solishtirishni umumlashtiradi.

3-jadval.

100 ishga o‘rtalashtirilgan holda, o‘ng senzurlangan eksponensial ma’lumotlar ($p = 0{,}3$) uchun asosiy va optimallashtirilgan amalga oshirishlar o‘rtasidagi ishlash vaqtini taqqoslash (sekundlarda)

Hajmi n	**Asosiy (s)**	**Optimallashtirilgan (s)**	**Oshishi**	**Xotira (MB)**
10^3	0,85	0,05	17 marta	0,1
10^4	8,70	0,23	38 marta	0,8
10^5	87,50	2,30	38 marta	7,6
10^6	912,00	23,10	39 marta	76,3

Vektorlashtirilgan amallar sodda sikllarga nisbatan ishlash vaqtini 15–40 marta qisqartirdi, ayniqsa, katta to‘plamlar uchun bu farq yanada yaqqol ko‘rinadi. $n = 10^6$ uchun optimallashtirilgan algoritm bitta to‘plamni taxminan 2,3 sekundda generatsiya qilishni yakunladi.

Yashovchanlik model baholashiga ta’siri. Tizimning metodologik tadqiqotlar uchun foydaliligini baholash maqsadida turli senzurlanish mexanizmlari Koks modeli [2] bahosiga qanday ta’sir ko‘rsatishi o‘rganildi. Simulyatsiyalarda senzurlanish ulushi (20%, 50%) va turi (o‘ng) o‘zgartirildi, haqiqiy xavf nisbati esa $\beta = log(2{,}0) = 0{,}693$ da doimiy saqlandi. 4-jadval $n = 500$ bilan 5000 takrorlash bo‘yicha 95% ishonch intervali uchun noxolislik va qamrov ehtimolliklarini ko‘rsatadi.

4-jadval.

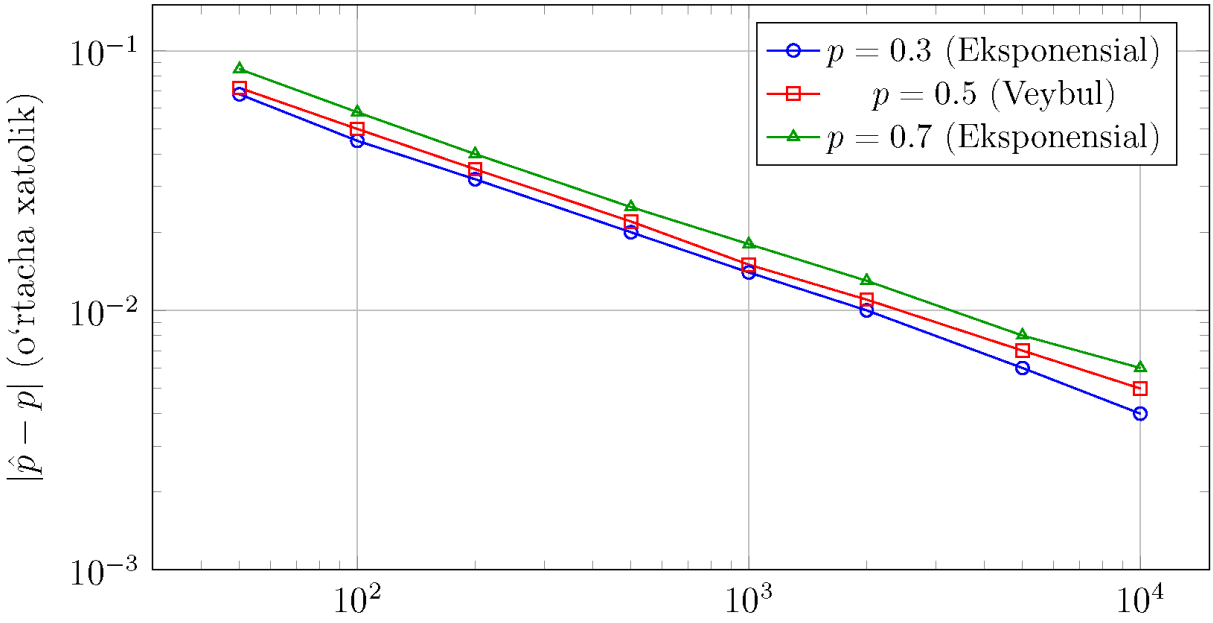
Turli senzurlanish senariylari ostida Koks modeli baholash samaradorligi ($n = 500$, haqiqiy $\beta = 0,693$; 5000 takrorlash)

Senzurlanish turi	Maqsadli p	O‘rtacha $\hat{\beta}$	Og‘ish	SX	Qamrov (%)
O‘ng	0,20	0,695	0,002	0,098	94,8
O‘ng	0,50	0,698	0,005	0,142	94,5

Natijalar tizim mo‘ljallangan senzurlanish naqshlarini muvaffaqiyatli yaratishini va statistik xossalarni to‘g‘ri saqlab qolishini tasdiqlaydi. Qamrov ehtimolliklari nominal 95% darajasiga yaqin qoladi va barcha sharoitlarda og‘ish ahamiyatsiz darajada kichik. Kutilganidek, yuqori senzurlanish ulushi standart xatolarning oshishiga olib keladi.

Ekstremal sharoitlarda mustahkamlik. Tizim yuqori senzurlanish sharoitlari ($p = 0,9$) va kichik tanlanma hajmlari ($n = 50$) ostida sinovdan o‘tkazildi. Hatto bu qiyin ssenariylarda ham senzurlanish ulushi nazorati o‘rtacha 0,015 xatolikdan past bo‘lgan holda aniq bo‘lib qoldi, bu esa ekstremal sharoitlar yoki cheklangan tanlanma hajmlarini talab qiluvchi ilovalar uchun barqarorlikni namoyish etadi.

Tanlanma hajmi o‘sishi bilan empirik senzurlanish ulushlarining maqsadli qiymatlarga yaqinlashuvini tasvirlaydi.



1-rasm. 10000 takrorlash bo‘yicha o‘rtacha xatolik $|\hat{p} - p|$ ning tanlanma hajmiga bog‘liq holda yaqinlashuvi

Muhokama.

Dasturiy ta’minot loyihasiga oid mulohazalar. Takrorlanish imkoniyati ishonchli simulyatsiya tadqiqotlarining asosi hisoblanadi. Tizim bir nechta himoya mexanizmlarini amalga oshiradi: tasodifiy sonlar generatsiyasi va transformatsiya bosqichlarini qat’iy ajratish, barcha tasodifiy sonlar oqimlarini PCG generatorlar oilasi [4] yordamida boshlash va protsessor yadrolari bo‘ylab tasodifiy son ketma-ketliklarini taqsimlash uchun sakrash usulidan foydalanish [5]. Amalga oshirish barcha tasodifiy ma’lumotlar va parametr sozlamalarini FAIR tamoyillariga [6] muvofiq standartlashtirilgan metama’lumot fayllarida qayd etadi.

Mavjud yondashuvlar bilan taqqoslash. 5-jadval taklif etilayotgan tizimni mavjud simulyatsiya yondashuvlari bilan tuzilgan tarzda taqqoslaydi.

5-jadval.

Senzurlangan ma’lumot generatsiyasi yondashuvlarini xususiyatlar bo‘yicha taqqoslash

Xususiyat	Taklif etilgan	I va II tur	Bosqichli	Maxsus
O‘ng senzurlanish	Ha	Ha	Ha	Ha
Chap senzurlanish	Ha	—	—	Qisman
Intervallik senzurlanish	Ha	—	—	Qisman
Ikki tomonlama tasodifiy senzurlanish	Ha	—	—	Ha
Ulush nazorati	Aniq	Belgilangan	Qisman	Taxminiy

Vektorlashtirish	Ha	—	—	Farqli
Parallel qo‘llab-quvvatlash	Ha	—	—	—
Modulli arxitektura	Ha	—	—	—
Takrorlanish imkoniyati	To‘liq	Qisman	Qisman	Farqli

Xulosa. Ushbu maqolada aniq nazorat qilinadigan senzurlanish ulushlariga ega senzurlangan yashovchanlik ma’lumotlarini generatsiya qilish uchun keng qamrovli Monte-Karlo simulyatsiya tizimi taqdim etildi. Tizim metodologik tadqiqotlardagi muhim bo‘shliqni bartaraf etadi, moslashuvchan taqsimot spetsifikatsiyalari ostida o‘ng, chap, interval va ikki tomonlama tasodifiy senzurlanish mexanizmlarini yagona tarzda qamrab oladi. Teskari akslantirish usulini kalibrlangan senzurlanish taqsimotlari bilan birlashtirish orqali usul sintetik to‘plamlar istalgan statistik xossalarni saqlab qolishini ta’minlaydi va turli modellash yondashuvlari o‘rtasida adolatli taqqoslashga zamin yaratadi.

ADABIYOTLAR:

1. J. Kishore, M. Goel, and P. Khanna, “Understanding survival analysis: Kaplan-Meier estimate,” International Journal of Ayurveda Research, vol. 1, no. 4, p. 274, 2010.

2. D. R. Cox, “Regression models and life-tables,” Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 34, pp. 187–202, Jan. 1972.

3. J. Zhou, J. Zhang, and W. Lu, “An expectation maximization algorithm for fitting the generalized odds-rate model to interval censored data,” Statistics in Medicine, vol. 36, pp. 1157–1171, Dec. 2016.

4. M. Zinkin, Good Judgment: Aesthetic Judgment and the Generation of Concepts, pp. 90–122. Oxford scholarship online, New York, NY: Oxford University PressNew York, NY, Oct. 2024. Includes bibliographical references and index. - Description based on online resource and publisher information; title from PDF title page (viewed on July 29, 2024).

5. J. K. Salmon, M. A. Moraes, R. O. Dror, and D. E. Shaw, “Parallel random numbers: as easy as 1, 2, 3,” in Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, SC -11, pp. 1–12, ACM, Nov. 2011.

6. M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, et al., “The fair guiding principles for scientific data management and stewardship,” Scientific Data, vol. 3, Mar. 2016.