TOSHKENT IRRIGATSIYA VA QISHLOQ XO‘JALIGINI MEXANIZATSIYALASH MUHANDISLARI INSTITUTI MILLIY TADQIQOT UNIVERSITETI HUZURIDAGI ILMIY DARAJALAR BERUVCHI DSc.03/31.03.2022.T/FM.10.04 RAQAMLI ILMIY KENGASH

NAVOIY DAVLAT PEDAGOGIKA INSTITUTI

MANSUROV DILSHOD RAVILOVICH

TO‘LIQ BO‘LMAGAN KUZATILMALI SXEMALARDA STATISTIK MODELLASHTIRISH METODLARI YORDAMIDA BAHOLARNI TAHLIL QILISH

05.01.07 – Matematik modellashtirish. Sonli usullar va dasturlar majmui

FIZIKA-MATEMATIKA FANLARI BO‘YICHA FALSAFA DOKTORI (PhD) DISSERTATSIYASI AVTOREFERATI

NAVOIY – 2023

UDK: 519.688 (519.24)

Fizika-matematika fanlari bo'yicha falsafa doktori (PhD) dissertatsiyasi avtoreferati mundarijasi

Оглавление автореферата диссертации доктора философии (PhD) по физико-математическим наукам

Content of dissertation abstract of doctor of philosophy (PhD) on physical-mathematical sciences

TOSHKENT IRRIGATSIYA VA QISHLOQ XO‘JALIGINI MEXANIZATSIYALASH MUHANDISLARI INSTITUTI MILLIY TADQIQOT UNIVERSITETI HUZURIDAGI ILMIY DARAJALAR BERUVCHI DSc.03/31.03.2022.T/FM.10.04 RAQAMLI ILMIY KENGASH

NAVOIY DAVLAT PEDAGOGIKA INSTITUTI

MANSUROV DILSHOD RAVILOVICH

TO‘LIQ BO‘LMAGAN KUZATILMALI SXEMALARDA STATISTIK MODELLASHTIRISH METODLARI YORDAMIDA BAHOLARNI TAHLIL QILISH

05.01.07 – Matematik modellashtirish. Sonli usullar va dasturlar majmui

FIZIKA-MATEMATIKA FANLARI BO‘YICHA FALSAFA DOKTORI (PhD) DISSERTATSIYASI AVTOREFERATI

NAVOIY – 2023

Fizika-matematika fanlari bo‘yicha falsafa doktori (PhD) dissertatsiyasi mavzusi O‘zbekiston Respublikasi Vazirlar Mahkamasi huzuridagi Oliy attestatsiya komissiyasida B2022.4.PhD/FM825 raqam bilan ro‘yxatga olingan.

Dissertatsiya Navoiy davlat pedagogika institutida bajarilgan.

Dissertatsiya avtoreferati uch tilda (o‘zbek, rus, ingliz (xulosa)) Ilmiy kengashning veb-sahifasida (https://www.ifar.uz) va “ZiyoNet” Axborot ta’lim portalida (http://www.ziyonet.uz) joylashtirilgan.

Ilmiy rahbar:	**Abdushukurov Abduraxim Axmedovich** fizika-matematika fanlari doktori, professor
Rasmiy opponentlar:	**Eshmatov Farxod Xasanovich** fizika-matematika fanlari doktori, professor
	Nurmuxamedova Nargiza Saydillayevna fizika-matematika fanlari bo‘yicha PhD, dotsent
Yetakchi tashkilot:	**Toshkent davlat transport universiteti**

Dissertatsiya himoyasi “TIQXMMI” Milliy tadqiqot universiteti huzuridagi fundamental va amaliy tadqiqotlar instituti huzuridagi DSc.03/31.03.2022.T/FM.10.04 raqamli Ilmiy kengashning 2023-yil “____” ____ soat ____ dagi majlisida bo‘lib o‘tadi (Manzil: 100000 Toshkent shahri, Qori Niyoziy ko‘chasi, 39-uy, Fundamental va amaliy tadqiqotlar instituti. __-katta majlislar zali; tel. (+99871) 237-09-61; faks (+998 71) 234-48-67; e-mail: info@ifar.uz).

Dissertatsiya bilan “TIQXMMI” Milliy tadqiqot universiteti huzuridagi fundamental va amaliy tadqiqotlar instituti Axborot-resurs markazida tanishish mumkin (___ raqami bilan ro‘yxatga olingan). Manzil: 100000 Toshkent shahri, Qori Niyoziy ko‘chasi, 39-uy, Fundamental va amaliy tadqiqotlar instituti, __-katta majlislar zali; tel. (+99871) 237-09-61.

Dissertatsiya avtoreferati 2023-yil “ ___” ___ kuni tarqatildi.
(2023-yil “___” ____ dagi ___ -raqamli reyestr bayonnomasi)

B.J.Ahmedov
Ilmiy darajalar beruvchi ilmiy
kengash raisi f.-m.f.d., professor

E.X.Karimbayev
Ilmiy darajalar beruvchi ilmiy
kengash ilmiy kotibi f.-m.f. bo‘yicha PhD

A.R.Hayotov
Ilmiy darajalar beruvchi ilmiy
kengash huzuridagi ilmiy seminar raisi
f.-m.f.d., professor

KIRISH (falsafa doktori (PhD) dissertatsiyasi annotatsiyasi)

Dissertatsiya mavzusining dolzarbligi va zarurati. Jahon miqyosida olib borilayotgan ko‘plab ilmiy va amaliy tadqiqotlar aksariyat hollarda statistik masalalarni tadqiq qilishga keltiriladi. Bizga ma’lumki, statistik tadqiqotlarning asosiy vazifasi kuzatuv natijalarini qayta ishlash, qaror qabul qilish va uni qulay shaklda taqdim etishdir. Biroq hayotda har doim ham to‘liq ma’lumotlar uchrayvermaydi. Gohida to‘liq bo‘lmagan kuzatuvlarni tadqiq qilish va ular yordamida bir yoki bir nechta noma’lum taqsimot qonuni haqida xulosa chiqarish kerak bo‘ladi. To‘liq bo‘lmagan kuzatishlarni biz umr davomiyligi, sug‘urta ishi, sifat nazorati, demografiya, tuproqshunoslik, astronomiya va boshqa sohalarga oid tadqiqotlarda ko‘plab uchratishimiz mumkin.

Matematik statistika muammolari, xususan, senzuralangan kuzatishlar bo‘yicha noparametrik baholash to‘liq kuzatuvli tanlanmalar bilan taqqoslaganda o‘ziga xos xususiyatlarga ega. Empirik taqsimot funksiyaga o‘xshash xususiyatlarga ega bo‘lgan to‘liq bo‘lmagan tanlanmalarga asoslangan bahoni qurish ko‘plab statistiklarning tadqiqot mavzusiga aylangan. Bu kabi mualliflarning to‘liq bo‘lmagan kuzatilmalarga asoslangan baholariga e’tibor qaratadigan bo‘lsak, bahoning empirik taqsimot funksiyaga nisbatan anchayin murakkab tarkibga ega ekanligini ko‘rishimiz mumkin. Shuning uchun ham ushbu baholarni o‘rganishda qaralayotgan modelning o‘ziga xos xususiyatlarini hisobga olgan holda noan’anaviy usullar va maxsus yondashuvlar zarur bo‘ladi. Ammo shunday murakkab bo‘lishiga qaramasdan amaliy masalalarning juda ko‘pchiligi to‘liq bo‘lmagan tanlanmali modellarga olib kelinadi. Bu esa baholarning xossalarini analitik usulda yechilmasada, sonli usullar yordamida o‘rganishni taqozo qiladi. Shu maqsadda ushbu dissertatsiya ishi ham, ayni paytda, mavjud bo‘lgan va amaliyotda qo‘llanilib kelayotgan baholarni solishtirish, ularning kuchli va zaif tomonlarini ajratib ko‘rsatish hamda ular uchun muvofiqlik mezonlarini ishlab chiqishda sonli usullardan foydalanib mavjud muammolarga qisman yechim berishni maqsad qiladi.

O‘zbekiston Respublikasi Prezidentining 2017-yil 7-fevraldagi PQ-4947-sonli “O‘zbekiston Respublikasini yanada rivojlantirish bo‘yicha harakatlar strategiyasi to‘g‘risida”; 2017-yil 17-fevraldagi PQ-2789-sonli “Fanlar akademiyasi faoliyati, ilmiy-tadqiqot ishlarini tashkil etish, boshqarish va moliyalashtirishni yanada takomillashtirish chora-tadbirlari to‘g‘risida”; 2017-yil 20-apreldagi PQ-2909-sonli “Oliy ta’lim tizimini yanada rivojlantirish chora-tadbirlari to‘g‘risida”; 2018-yil 27-apreldagi PQ-3682-sonli “Innovatsion g‘oyalar, texnologiyalar va loyihalarni amaliyotga joriy qilish tizimini yanada takomillashtirish chora-tadbirlari to‘g‘risida”; 2020-yil 7-maydagi PQ-4708-sonli “Matematika sohasidagi ta’lim sifatini oshirish va ilmiy tadqiqotlarni rivojlantirish chora-tadbirlari to‘g‘risida”gi va 2020-yil 6-oktabrdagi PQ-4851-sonli “Axborot texnologiyalari sohasida ta’lim tizimini yanada takomillashtirish, ilmiy tadqiqotlarni rivojlantirish va ularni IT industriya bilan integratsiya qilish chora-tadbirlari to‘g‘risida”gi qarorlari hamda mazkur faoliyatga tegishli boshqa normativ-huquqiy hujjatlarda belgilangan vazifalarni amalga oshirishda ushbu dissertatsiya tadqiqoti muayyan darajada xizmat qiladi.

Tadqiqotning respublika fan va texnologiyalari rivojlanishi ustuvor yo‘nalishlariga bog‘liqligi. Mazkur tadqiqot respublika fan va texnologiyalar rivojlanishining IV “Matematika, mexanika va informatika” ustuvor yo‘nalishi doirasida bajarilgan.

Muammoning o‘rganilganlik darajasi. Statistikada dastlab “senzurlanganlik” va “kesilganlik” atamalarining tizimli qo‘llanilishi A.Hald va B.Friedmanga tegishli ekanligini ta’kidlab o‘tamiz. Keyinchalik senzurlanishli tanlanmalarga oid ishlarda, asosan, o‘ng tomondan senzurlanishli modellar N.Breslow, M.Burke, G.Campbell, S.Csörgő, D.Dabrowska, B.Efron, A.Földes, L.Rejto, R.Gill, M.Hollander, E.Kaplan, P.Meier, R.Liu, F.J.Rubio K.S.Sagidullayev, R.S.Murodov, F.A.Abdukalikov kabi olimlar tomonidan o‘rganilgan. Buning sababi o‘ng tomondan tasodifiy senzurlanishli tanlanmalarning tabiatda ko‘p uchrashi hamda ularning metodologik jihatdan tushuntirishning sodda ekanligida. Shuningdek, o‘ng tomondan senzurlanishga doir ishlarni chap tomon uchun ham oson umumlashtirish mumkinligidir. Ma’lumki, to‘liq kuzatishlar sharoitida taqsimot funksiyaning noparametrik bahosi empirik taqsimot funksiya hisoblanadi va uning xossalari ko‘plab mualliflar Y.Blagoveshchensky, R.Gill, R.Dudley, A.Dvoretzky, W.Stute, J.Van Rayzin, P.Major, S.Csörgő tomonidan o‘rganilgan. Amaliyotda ko‘p qo‘llanilishi talabidan senzurlangan ma’lumotlar asosida quriladigan noparametrik baholarni topish asosiy statistik masalaga aylandi. Izlanishlarning natijasi sifatida 1958-yilda Kaplan va Meier birinchilardan bo‘lib o‘ng tomonlama tasodifiy senzurlanish modeli uchun o‘zlarining PL-(product-limit, ya’ni ko‘paytma ko‘rinishidagi) bahosini taklif qilishgan hamda keyinchalik u ko‘plab mualliflar tomonidan chuqur o‘rganilgan. PL-bahoning turli xil o‘zgartirilgan versiyalari mavjud. Ko‘pgina olimlarning ishlari ushbu bahoni o‘rganish va undan foydalanishga bag‘ishlangan. Buni dissertatsiya oxiridagi adabiyotlar ro‘yxati ham tasdiqlaydi. Ko‘p sonli olimlar orasida Y.K.Belyayev, Y.N.Blagoveshchensky, R.D.Gill, O.Aalen, M.Csörgő, S.Csörgő, N.Breslow, B.Efron, P.K.Andersen, M.D.Burke, L.Horváth, A.Földes, L.Rejto, D.M.Dabrowska, J.K.Ghorai, V.Susarla, J.Van Rayzin, P.Major, E.G.Phadia, W.Stute, J.Wellner, N.Veraverbeke mualliflarni, O‘zbekistonda esa V.I.Romanovskiy, S.H.Sirojiddinov, A.A.G‘ofurov, K.S.Sagidullayev, R.S.Murodov, F.A.Abdukalikov kabi olimlarni alohida ta’kidlab o‘tishimiz kerak. Abdushukurov tomonidan o‘ng tomondan tasodifiy senzurlanishli modellar uchun baholar taklif etilgan. Xususan, muallif proporsional intensivlik modeli uchun darajali ko‘rinishdagi baho taklif etgan. Keyinchalik (S.Csörgőning ishlarida) ushbu baho ACL (ACL-Abdushukurov – Cheng – Lin) bahosi deb nomlanadi. Bu va bundan keyingi o‘rinlarda ushbu bahoni darajali baho deb yuritamiz.

Yuqorida nomi tilga olingan Kaplan-Meier bahosi amaliy jihatdan juda keng qo‘llaniladigan baho bo‘lib, uning kovariatali ko‘rinishi Beran tomonidan umumlashtirilgan. Yuqorida keltirilgan adabiyotlardan ko‘rinib turganidek, ushbu baholar chuqur o‘rganilgan va amaliyotga tatbiq qilingan. Biroq darajali baho Kaplan-Meir bahosiga qaraganda bir qator afzalliklarga ega. Uning umumlashmasi bo‘lgan kovariatali ko‘rinishi deyarli o‘rganilmagan. Ushbu dissertatsiya ishi yuqorida aytib o‘tilgan baholarni statistik modellashtirish hamda sonli usullar yordamida o‘rganishni va ma’lum darajada solishtirishni maqsad qiladi.

Yuqorida ta’kidlab o‘tilgan ko‘pgina olimlarning senzurlangan tanlanmalar bo‘yicha tadqiqotlari bilan bog‘liq ishlari tufayli hozirda matematik statistikaning to‘liq bo‘lmagan kuzatishlar statistikasi deb ataladigan alohida sohasi yuzaga keldi.

Dissertatsiya ishining maqsadi. Darajali bahoning ikki tomonlama senzurlanishdagi o‘xshashini tadqiq qilish va mavjud baholar bilan taqqoslash hamda u uchun muvofiqlik mezonlarini ishlab chiqish va o‘rganish. Darajali bahoning kovariatali ko‘rinishdagi umumlashmasi uchun optimal “oyna kengligi” parametri mavjud ekanligi, uning qanday parametrlarga bog‘liq ekanligini o‘rganish va uni topish uchun adaptiv algoritmni ishlab chiqishdan iborat.

Dissertatsiya ishining vazifasi. Maqsadga muvofiq quyidagi vazifalarni hal etish rejalashtirilgan:

o‘ng tomondan tasodifiy senzurlanishli informativ modelda mavjud baholarni xususiyatlarini o‘rganish. Ushbu baholarning afzalliklari va kamchiliklarini solishtirish (darajali bahoni Kaplan-Meier bahosi bilan solishtirish);

informativ model baholarini senzurlanishning noma’lum parametrlariga bog‘liqligini o‘rganish hamda eksponensial taqsimotni parametrik baholash;

darajali baho uchun muvofiqlik mezonlarini qurish va ularning xossalarini tadqiq qilish;

taqsimot funksiyasini kovariatali bahosi xususiyatlarini o‘rganish hamda “oyna kengligi” parametrini tanlashning taqsimot funksiyaning kovariatali bahosi xususiyatlariga ta’sirini tadqiq etish;

taqsimot funksiyaning kovariatali bahosi uchun mos keluvchi optimal “oyna kengligi” parametrini topish algoritmini ishlab chiqish hamda taklif qilingan algoritmdan foydalangan holda dasturiy mahsulot yaratish.

Tadqiqotning obyekti. O‘ng tomondan tasodifiy senzurlanishli informativ modelning darajali bahosi va uning tabiiy umumlashmalari. Kaplan-Meier bahosi va uning umumlashmalari.

Tadqiqotning predmeti. O‘ng tomonlama tasodifiy senzurlanishli modelda darajali baho va Kaplan-Meier baholarini solishtirish ular uchun muvofiqlik mezonlari xossalarini o‘rganish. Baholarning kovariatali ko‘rinishlari va ularning xossalarini aniqlashdan iborat.

Metodologiya va tadqiqot usullari. Belgilangan vazifalarni hal qilish uchun matematik statistika, ehtimollar nazariyasi, matematik va statistik modellashtirish (dasturlash) usullari qo‘llanildi.

Tadqiqotning ilmiy yangiligi quyidagilardan iborat:

ilk marotaba ikki tomonlama tasodifiy senzurlanishli informativ modelda bosh to‘plam eksponensial taqsimoti parametrik baholangan;

ilk marotaba Kolmogorov statistikasida empirik taqsimot funksiya o‘rniga darajali baho qo‘yilganda statistikaning limit taqsimoti taqsimot funksiyaga va uning parametrlariga bog‘liq emasligi, bog‘liq bo‘lsa ham amaliy jihatidan ahamiyatli emasligi sonli usullar yordamida ko‘rsatib berilgan;

darajali bahoga asoslangan muvofiqlik mezoni taklif qilingan;

taqsimot funksiyasining kovariatali bahosining statistik xususiyatlarining tanlangan yadro funksiyalari turiga, tanlanma hajmiga, eksperiment rejasiga, nazorat nuqtalari soniga, regressiyaga bog‘liqligi ko‘rsatib berilgan va o‘rganilgan;

“oyna kengligi” parametrining optimal qiymatini tanlash uchun adaptiv algoritm taklif etilgan va joriy qilingan. Bu esa eksperimental kuzatishlar natijalarini sifatli yarimparametrik bahosini tuzish imkonini berdi.

Tadqiqotning amaliy natijalari quyidagilardan iborat:

darajali baho uchun muvofiqlik mezonlarini tekshiruvchi to‘liq funksionalga ega bo‘lgan dasturiy mahsulot ishlab chiqildi;

optimal “oyna kengligi” parametrini tanlashning adaptiv algoritmi ishlab chiqildi;

taklif qilingan algoritmlarni amalga oshiradigan dasturiy ta’minot yaratildi.

Tadqiqot natijalarining ishonchliligi quyidagilar bilan ta’minlanadi:

muvofiqlik mezonlarini ishlab chiqishda va ularning statistik xossalari hamda taqsimotlarini o‘rganish uchun matematik apparat va statistik modellashtirish usullarini to‘g‘ri qo‘llanilishi;

statistik modellashtirish natijalarining oldindan ma’lum nazariy natijalar bilan mos kelishi;

matematik fikrlashning qat’iyligi;

olingan natijalarning dalillari matematik (nazariy) jihatdan to‘g‘riligi;

olingan natijalarning boshqa mualliflar tomonidan olingan qismiy natijalari bilan mazmunan mos kelishi.

Tadqiqot natijalarining ilmiy va amaliy ahamiyati. Tadqiqot natijalarining ilmiy ahamiyati shundan iboratki, ishda olingan ilmiy-amaliy natijalar keyingi to‘liq bo‘lmagan tanlanmalarga bog‘liq bo‘lgan tadqiqotlarda foydalaniladi. Shuningdek, ishlab chiqilgan dasturiy mahsulot yordamida amaliy masalalarni yechishda qo‘llanilishi mumkin.

Tadqiqot natijalarini amaliyotga tatbiq etish. Dissertatsiya ishi natijasi sifatida olingan informativ va informativ bo‘lmagan modellarda baholarni taqqoslash natijalari hamda proporsional intensivlik bahosi uchun ishlab chiqilgan muvofiqlik mezonidan F4-40 raqamli “O‘lchovlik funksiyalar sinfida indekslangan integral empirik protsesslarning asimptotik xossalarini tadqiq etish” mavzusidagi fundamental loyihada foydalanilgan (Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universitetining 2023-yil 6-fevraldagi 04/11-566-sonli ma’lumotnomasi). Ilmiy natijalarning qo‘llanilishi Koziol-Grinning proporsional intensivliklar xususiy modeli bahosini Kaplan-Meier bahosi bilan taqqoslash imkonini bergan;

informativ va informativ bo‘lmagan modellar uchun sun’iy tanlanmalar hosil qilib ular yordamida muvofiqlik mezonlarini tekshiruvchi dasturiy mahsulotdan hamda to‘liq bo‘lmagan kuzatilmali modellar baholari qo‘llanilgandagi empirik jarayonning kovariatsion ko‘rinishidan Yo-F4-07 raqamli “Kopula funksiyalari yordamida statistik baholash va gipotezalarni tekshirish” mavzusidagi fundamental yoshlar loyihasida foydalanilgan (Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universitetining 2023-yil 13-maydagi 04/11-2875-sonli ma’lumotnomasi). Ilmiy natijalarning qo‘llanilishi taqsimot funksiyalar oilasi haqidagi mezonlarni tekshirishga yordam bergan.

Tadqiqot natijalarini aprobatsiya qilish. Mazkur tadqiqot natijalari 8 ta, jumladan, 3 ta xalqaro va 5 ta respublika ilmiy-amaliy anjumanlarida muhokamadan o‘tkazilgan.

Tadqiqot natijalarini nashr etish. Dissertatsiya mavzusi bo‘yicha jami 16 ta ilmiy ish chop etilgan, shulardan, O‘zbekiston Respublikasi Oliy attestatsiya komissiyasining doktorlik dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan ilmiy nashrlarda 5 ta maqola, jumladan, 2 ta xorijiy va 3 ta respublika jurnallarida nashr etilgan hamda EHM uchun yaratilgan dasturning rasmiy ro‘yxatdan o‘tkazilganligi to‘g‘risidagi 3 ta guvohnoma olingan.

Dissertatsiyaning hajmi va tuzilishi. Ushbu dissertatsiya ishi kirish, 3 ta bob, xulosa va ilovalardan iborat. Dissertatsiyaning umumiy hajmi 120 bet bo‘lib, u 47 ta rasm va 21 ta jadvalni o‘z ichiga oladi. Adabiyotlar ro‘yxati 110 ta nomdan iborat.

DISSERTATSIYANING ASOSIY MAZMUNI

Kirish qismida dissertatsiya mavzusining dolzarbligi va zarurati asoslangan bo‘lib, tadqiqotning respublika fan va texnologiyalari rivojlanishining ustuvor yo‘nalishlariga mosligi ko‘rsatilgan. Muammoning o‘rganilganlik darajasi keltirilgan. Tadqiqot maqsadi, vazifalari, obyekti va predmeti tavsiflangan. Tadqiqotning ilmiy yangiligi va amaliy natijalari bayon qilingan. Olingan natijalarning nazariy va amaliy ahamiyati ochib berilgan. Tadqiqot natijalarining joriy qilinishi, nashr etilgan ishlar va dissertatsiya tuzilishi bo‘yicha ma’lumotlar keltirilgan.

Dissertatsiyaning “**Senzurlanishli tanlanmalar**” deb nomlangan birinchi bobida dissertatsiya mavzusini to‘la yoritish uchun zarur bo‘lgan asosiy ta’riflar va muhim tushunchalar keltirilgan. Shuningdek, dissertatsiya ishida ko‘p marotaba murojaat qilinadigan taqsimot funksiyalari va ularning xossalari, umr davomiyligi ko‘rinishidagi tanlanmalar va ularni sun’iy hosil qilish algoritmlari bayon qilingan.

Bosh to‘plam taqsimoti $F(t)$ ning o‘rniga uning turli baholarini ishlatishimiz mumkin. U holda bahoning bosh to‘plam taqsimotiga qay darajada yaqinligini yoki bir nechta baholardan qaysi biri yaxshiroq ekanligini tekshirishga to‘g‘ri keladi. Shu maqsadda bahoning “sifati” o‘lchovini:

$$d\left(F_1, F_2\right) = \sup_{-\infty<t<\infty} \left|F_1(t) - F_2(t)\right| \quad (1)$$

kabi olamiz.

Statistik tajribalarda kerakli aniqlikka erishish uchun tajribalarni bir necha bor takrorlash zarur. Shunga ko‘ra, takrorlanishlar soni

$$N \le \left[\frac{t_\gamma^2}{\varepsilon^2}\right] + 1 \quad (2)$$

ko‘rinishda bo‘lishi kerak ekanligi ko‘rsatib o‘tilgan. Ushbu ishda $\varepsilon = 0,01$ va $t_\gamma = 0,99$ deb olingan. Bularni (2)ga qo‘yib, takrorlanishlarning minimal soni $N = 16600$ ekanligini hisoblaymiz. Dissertatsiyadagi barcha natijalar eng kamida 20000 ta tajribaning o‘rtachasi sifatida olingan.

Tadqiqotchi kuzatilayotgan obyektni tajriba boshidan uning yakuniga qadar kuzatishi mumkin. Bunday ko‘rinishdagi kuzatishlar to‘liq kuzatish (klassik statistika) deyiladi va tajriba yakunini anglatuvchi **“ishdan chiqish” (Отказ)** deb ataluvchi qandaydir hodisa bilan yakunlanadi. Ishdan chiqish subyektiv tushuncha bo‘lib, tajriba davomida foydalanuvchi tomonidan kiritiladi va ta’riflanadi. Ishdan chiqishga misol qilib texnik qurilmaning buzilishi, iqtisodiy masalalarda bankrotlik holati, meditsinada bemorning o‘lim holati va shu kabi bir qancha holatlarni keltirishimiz mumkin.

Ammo amaliyotda har doim ham obyektni tajriba boshidan oxirigacha kuzatishimizning imkoni bo‘lmaydi. Ya’ni ayrim obyektlar tajriba yakuniga yetmasdan kuzatishdan chiqib ketadi (“ishdan chiqish”ga uchramasdan). Bu obyektlar haqida faqatgina kuzatishdan chiqib ketgan vaqt momentigacha bo‘lgan ma’lumotlar qoladi. Bunday holatga misol qilib hali davolanishni yakunlamagan bemorning qandaydir sabablarga ko‘ra (doimiy yashash manzili o‘zgarganligi, boshqa klinikaga o‘tganligi va hakozo) klinikani tark etishini keltirishimiz mumkin. Bu holatda bemorning klinikaga kelgan vaqtidan boshlab, klinikani tark etgunigacha bo‘lgan ma’lumotlar bor. Ammo ketganidan keyingi ma’lumotlar mavjud emas. Klinikada faqatgina bemorning chiqib ketgan vaqt momentigina mavjud xolos. Bunday kuzatilmalar **senzurlangan kuzatilmalar** deyiladi. Sodda qilib aytadigan bo‘lsak, obyektning kuzatuvchining ko‘rish maydonidan g‘oyib bo‘lishi (klinikani tark etishi) yoki sinov boshlanganidan keyin bir muncha vaqt o‘tgach kuzatuv ostiga tushishi (bemorda kasallikning dastlabki simptomlari yuzaga chiqqan vaqt momenti noma’lum) senzurlangan kuzatilmalarga misol bo‘ladi. Birinchi hol ma’lumotlarning **o‘ngdan senzurlanishi**, ikkinchi hol esa **chapdan senzurlanishidir**. O‘ng tomondan senzurlanish umr davomiyligi bilan bog‘liq masalalarda ko‘p qo‘llanilganligi hamda uning chap tomon uchun ham osongina umumlashtirish mumkin bo‘lganligi uchun ushbu ishda, asosan, o‘ng tomondan senzurlanishli modellar qaralgan.

O‘ng tomondan senzurlanishli tanlanmalar (umr davomiyligi ko‘rinishidagi kuzatilmalar) deb, quyidagicha aniqlanuvchi C^n to‘plamga aytiladi:

$$C^n = \left\{ \left(Z_i, \delta_i \right), i = \overline{1,n} \right\} \tag{3}$$

bu yerda $Z_i = \min\left(T_i, Y_i\right)$, $T_i -$“ishdan chiqish” yuz bergan vaqt momenti, $Y_i -$ senzurlanish yuz bergan vaqt momenti (*i*-obyekt uchun kuzatishning tugatilish vaqti) va $\delta_i = I\left(T_i \le Y_i\right) -$ hodisa indikatori.

Agar tajriba davomida barcha obyektlarning “ishdan chiqish” vaqt momentlari kuzatilgan bo‘lsa ($Z_i = T_i,\ \delta_i = 1$), u holda bunday kuzatilma **to‘liq kuzatilma** deyiladi (klassik statistika). Agarda qandaydir $Y_i \le T_i$ vaqt momentida kuzatishning tugatilganligi tufayli T_i noma’lum bo‘lsa ($Z_i = Y_i,\ \delta_i = 0$) u holda C^n kuzatilma **o‘ngdan senzurlangan** deyiladi.

O‘ng tomondan senzurlanishni quyidagi uchta asosiy turga bo‘lish mumkin:

I tur senzurlanish. Bu holatda barcha obyektlar oldindan aniqlab olingan qandaydir Y vaqt momentigacha kuzatiladi. Bu vaqt momentigacha “ishdan chiqish”ga uchramagan barcha obyektlarni o‘ng tomondan senzurlangan deb qabul qilinadi ($Z_i = Y,\quad \delta_i = 0$).

II tur senzurlanish. Bu holatda kuzatish ishlari dastlabki $k\ (k<n)$ ta "ishdan chiqish" yuz berguncha amalga oshiriladi. Boshqa barcha obyektlar senzurlangan deb topilib, ularning senzurlanish vaqt momenti $Z_i=T_{(k)},\quad \delta_i=0$ deb qabul qilinadi. Bu yerda $T_{(k)}-$ oxirgi kuzatilgan $k-$obyektining senzurlanishga uchragan vaqt momenti.

III tur senzurlanish. Har bir kuzatilma $Z_i=\min(T_i,\ Y_i)$ kabi tuzib olinadi. Bu yerda T_i- "ishdan chiqish" vaqt momenti hamda Y_i- senzurlanish vaqt momenti bo'lib, ularning taqsimot funksiyalari mos ravishda $F(t)$ va $G(t)$. T_i va Y_i bog'liqsiz tasodifiy miqdorlardir. Bu turdagi senzurlanish o'z navbatida **informativ** va **informativ bo'lmagan** turlarga bo'linadi. Agarda $F(t)$ va $G(t)$ taqsimot funksiyalar bir-biri bilan funksional bog'langan bo'lsa informativ, aks holda informativ bo'lmagan senzurlanish modeli deyiladi.

Biz ushbu dissertatsiya ishida aynan III turdagi senzurlanishdan foydalanamiz. Buning sababi ayni ushbu ko'rinishdagi senzurlanish meditsina, biostatistika, injenerlik tadqiqotlarida va boshqa amaliy masalalarda ko'p uchraydi.

Ushbu to'liq $T=\{T_1,T_2,\ \dots\ ,T_n\}$ tanlanmaga ko'ra $F(t)$ funksiyaning ko'rinishi haqida hech narsa deb bo'lmaydi. Ammo noma'lum $F(t)$ taqsimot funksiyani $F_n^{э}(t)$ empirik taqsimot funksiya yordamida quyidagicha baholash mumkin:

$$F_n^{э}(t)=\frac{1}{n}\sum_{i=1}^{n}I(T_i\le t) \tag{4}$$

bu yerda $I(A)-A$ hodisaning indikatori: $I(A)=\begin{cases}1, & A\ \text{ro'y bersa},\\ 0, & \overline{A}\ \text{ro'y bersa}.\end{cases}$

Dissertatsiya ishining asosiy obyekti senzurlanishli tanlanmalar yordamida taqsimot funksiyani baholovchi empirik taqsimot funksiya (4)ga o'xshash baholarini tadqiq qilishdir. Bular haqida dissertatsiyaning **"Senzurlangan tanlanmalar yordamida taqsimot funksiyani baholash"** nomli ikkinchi bobida so'z boradi:

Kaplan-Meier bahosi. 1958-yili Kaplan va Meier tomonidan, F taqsimot funksiya uchun taklif etilgan "product-limit" (PL) bahosi

$$F_n^{PL}(t)=\begin{cases}1-\prod\limits_{\{j:Z_{(j)}\le t\}}\left(1-\dfrac{\delta_{(j)}}{n-j+1}\right), & t\le Z_{(n)},\\ 1, & t>Z_{(n)},\quad \delta_{(n)}=1,\\ \text{aniqlanmagan}, & t>Z_{(n)},\quad \delta_{(n)}=0,\end{cases} \tag{5}$$

bu yerda $Z_{(1)}\le\dots\le Z_{(n)}-$tartibli statistika, $\delta_i=I(X_i\le Y_i)$.

Darajali baho. Faraz qilaylik, $\{(X_k,L_k,Y_k),k\ge1\}$ ketma-ketlik (X,L,Y) uchlikning bog'liqsiz amaldagi qiymati va

$$S^{(n)}=\{(Z_i,\Delta_i),i=1,...,n\}$$

– kuzatilgan tanlanma bo‘lsin. Bu yerda $Z_i=\max\{L_i,\min\{X_i,Y_i\}\}$, $\Delta_i=\left(\delta_i^{(0)},\delta_i^{(1)},\delta_i^{(2)}\right)$, $\delta_i^{(0)}=I\left(\min(X_i,Y_i)<L_i\right)$, $\delta_i^{(1)}=I(L_i\le X_i<Y_i)$, $\delta_i^{(2)}=I(L_i\le Y_i<X_i)$ va $I(A)-$ A hodisaning indikatori. Bunday ko‘rinishdagi modelni tavsiflash maqsadida Z_i va $V_i=\min(X_i,Y_i)$ tasodifiy miqdorlarni kiritib ularning taqsimot funksiyalarini H hamda N orqali belgilab olamiz. H va N mos ravishda maksimum hamda minimumning taqsimot funksiyalari bo‘lganligidan oson tushunish mumkinki,

$$\begin{gathered}H(x)=K(x)N(t),\\ N(t)=1-(1-F(t))(1-G(t)),t\in\mathbb{R}^1.\end{gathered}\tag{6}$$

Endi (6) tenglikka qo‘shimcha tarzda shunday musbat noma’lum θ va β parametrlar topilib, barcha $t\in\mathbb{R}^1$ uchun quyidagi tenglik bajarilishini talab qilamiz:

$$\begin{cases}1-G(t)=(1-F(t))^{\theta},\\ K(t)=(N(t))^{\beta},\end{cases}\tag{7}$$

bu yerda β parametr chap tomondan, θ esa o‘ng tomondan senzurlanish darajasini aniqlaydi. Ushbu parametrlarning qiymatlarini nolga intilishi mos tomondan senzurlanishning kuchsizlanganini anglatadi. Bunday ko‘rinishdagi ikki tomonlama tasodifiy senzurlanishli maxsus model A.A.Abdushukurov tomonidan kiritilgan. (6) va (7) formulalardan oson tushunish mumkinki, o‘rniga qo‘yish usuli yordamida F uchun quyidagi ifodani topamiz:

$$1-F(t)=\left[1-(H(t))^{\lambda}\right]^{\gamma},t\in\mathbb{R}^1,\tag{8}$$

bu yerda $\lambda=\dfrac{1}{1+\beta}$ va $\gamma=\dfrac{1}{1+\theta}$ bo‘lib, λ hamda γ ning 1 ga intilishi mos tomondan senzurlanishning kuchsizlanishini anglatadi. F taqsimot funksiya uchun topilgan (8) formuladagi $(H(x),\lambda,\gamma)$ ifodalarni $S^{(n)}$ tanlanma yordamida baholarini topib, F taqsimot funksiya uchun yarimparametrik bahoni hosil qilamiz. Buning uchun modelning eng asosiy xarakteristikasidan foydalanamiz:

Teorema 1. (7) *tenglik bajarilishi uchun* Z_i *va* Δ_i *tasodifiy miqdorlarning bog‘liqsiz bo‘lishi zarur hamda yetarli.*

Faraz qilaylik, (7) tenglik va teorema 1 ning natijalaridan kelib chiqqan holda λ hamda θ parametrlarni quyidagicha baholab olishimiz mumkin:

$$\lambda_n=1-p_n^{(0)},\quad \gamma_n=p_n^{(1)}\left(1-p_n^{(0)}\right)^{-1}\tag{9}$$

bu yerda $p_n^{(m)}=\dfrac{1}{n}\sum_{i=1}^{n}\delta_i^{(m)},\ m=0,1,2$.

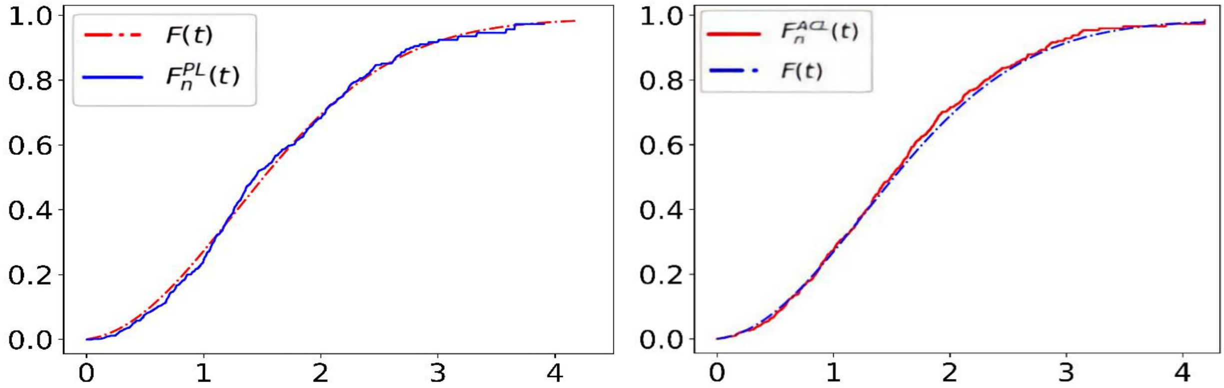
(8) ifodaga (9) ni qo‘yib $F(t)$ uchun quyidagi bahoga ega bo‘lamiz:

$$F_n^{ACL}(t) = 1 - \left[1 - \left(H_n(t)\right)^{\lambda_n}\right]^{\gamma_n}, t \in \mathbb{R}^1. \tag{10}$$

Agarda (7) tenglikda $\beta = 0$, ya'ni $\lambda = 1$ bo'lsa, u holda chap tomondan senzurlanish yo'q ekanligini anglatadi. Bu holatda baho quyidagi ko'rinishni oladi:

$$F_n(t) = 1 - \left[1 - H_n(t)\right]^{\gamma_n}, t \in \mathbb{R}^1. \tag{11}$$

Python programmalash tili yordamida o'ng tomondan senzurlanishli model bo'yicha tanlanma beruvchi modul yaratib, har ikkala bahoning chizmasini chizib olamiz (1-rasm).



a) Kaplan-Meier bahosi b) Darajali baho

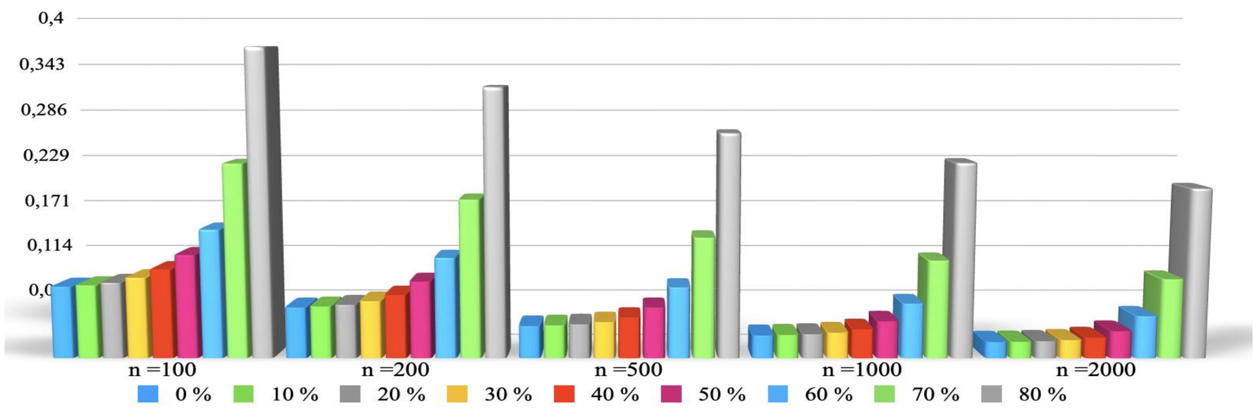
1-rasm. $F(t)$ – Veybull taqsimoti bahosi, $n = 500$, senzurlanish 30 %.

Rasmdan (1-rasm, a), b)) ko'rinib turganidek, har ikkala baho taqsimot funksiyaga juda yaxshi yaqinlashadi. Ammo Kaplan-Meier bahosining asosiy kamchiliklaridan biri bu eng katta variatada senzurlanish bo'lsa, keyingi nuqtalarda taqsimot funksiyani o'zida aks ettira olmaydi (1-rasm a)). Darajali bahoda bunday kamchilik kuzatilmaydi. U butun son o'qida aniqlangan. Endi baholarni qaysi biri taqsimot funksiyaga yaqinroq ekanligini bilish maqsadida taqsimot funksiya va baho orasidagi (1) masofaning (2)da aytilganiga ko'ra, 20000 ta tajribadagi o'rtacha qiymatlaridan ushbu jadvalni (1-jadval) tuzib olamiz.

1-jadval

Kaplan-Meier bahosining nazariy taqsimot funksiyadan chetlashishi ($F(t)$ – Veybull taqsimoti)

Veybull taqsimoti ($c = 2$)					
O'tacha senzurlanish darajasi	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
0 %	0,08531	0,06062	0,03845	0,02729	0,01941
10 %	0,08704	0,06189	0,03940	0,02793	0,01973
20 %	0,09004	0,06393	0,04061	0,02876	0,02043
30 %	0,09591	0,06815	0,04323	0,03056	0,02174
40 %	0,10581	0,07545	0,04863	0,03451	0,02453
50 %	0,12317	0,09151	0,06073	0,04430	0,03249
60 %	0,15315	0,11967	0,08471	0,06535	0,05047
70 %	0,23079	0,18854	0,14359	0,11655	0,09443
80 %	0,36622	0,31974	0,26643	0,23165	0,20170

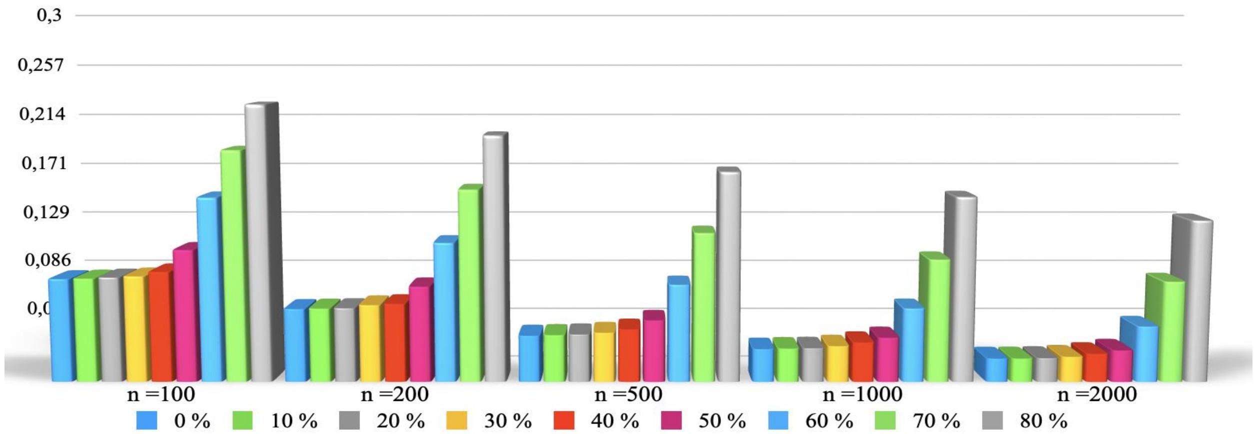


2-rasm. Kaplan-Meier bahosining nazariy taqsimot funksiyadan chetlashishi ($F(t)$ – Veybull taqsimoti).

2-jadval

Darajali bahoning nazariy taqsimot funksiyadan chetlashishi (Veybull taqsimoti)

Veybull taqsimoti ($c=2$)					
O‘rtacha senzurlanish darajasi	$n=100$	$n=200$	$n=500$	$n=1000$	$n=2000$
0 %	0,08526	0,06053	0,03840	0,02726	0,01940
10 %	0,08569	0,06089	0,03896	0,02778	0,0195
20 %	0,08664	0,06127	0,03952	0,02803	0,01983
30 %	0,08752	0,06374	0,04087	0,02976	0,02095
40 %	0,09111	0,06467	0,04368	0,03271	0,02337
50 %	0,10931	0,07905	0,05118	0,03680	0,02629
60 %	0,15230	0,11517	0,08066	0,06104	0,04614
70 %	0,19078	0,15876	0,12329	0,10141	0,08323
80 %	0,22785	0,20286	0,17335	0,15289	0,13364



3-rasm. Darajali bahoning nazariy taqsimot funksiyadan chetlashishi (Veybull taqsimoti).

Yuqoridagi 2- va 3-rasmlar 1- va 2-jadvallar asosida chizilgan bo‘lib, har bir tanlanma hajmida 9 xil senzurlanish darajasida taqsimot funksiya va baho orasidagi farq tasvirlangan. Unga ko‘ra, bahoning taqsimot funksiya bilan orasidagi farqi senzurlanish darajasi oshishi bilan ko‘rsatkichli ko‘rinishda oshishini ko‘rishimiz mumkin. Jadvallardagi har bir ustun va har bir satrlarni o‘zaro taqqoslab, informativ modelda darajali baho Kaplan-Meier bahosiga nisbatan taqsimot funksiyaga yaqinroq degan xulosani qat’iy aytishimiz mumkin.

Yuqoridagi taqqoslash ishlari informativ modelda olib borildi. Informativ modelda darajali baho Kaplan-Meier bahosiga nisbatan yaxshiroq ekanligi ko‘rsatib berildi. Informativ bo‘lmagan (tasodifiy senzurlanish) modelda ham baholarni taqqoslash maqsadida “umr davomiyligi”ni standart parametrli eksponensial taqsimot va senzurlanish vaqt momenti taqsimot funksiyasini esa kerakli parametrli ($\int_{-\infty}^{+\infty}\left(1-G\left(t,\ \theta\right)\right)f\left(t\right)dt=1-r$ tenglikdan topiladi) Gamma taqsimotdan foydalanib, tasodifiy senzurlanishli tanlanma hosil qilib, ular yordamida quyidagi taqsimot funksiya va baho orasidagi “masofa” jadvalini tuzib olamiz:

3-jadval

Kaplan-Meier bahosining nazariy taqsimot funksiyadan chetlashishi (Eksponensial taqsimot, senzurlanish vaqt momenti Gamma taqsimotga ega)

Eksponensial taqsimot $(\lambda=1)$					
O‘rtacha senzurlanish darajasi	$n=100$	$n=200$	$n=500$	$n=1000$	$n=2000$
0 %	0,08534	0,06105	0,03827	0,02737	0,01933
10 %	0,08718	0,05921	0,03897	0,02788	0,01970
20 %	0,08726	0,06213	0,03923	0,02807	0,01995
30 %	0,09448	0,07071	0,04418	0,03124	0,02241
40 %	0,10656	0,07779	0,05115	0,03683	0,02643
50 %	0,12372	0,09155	0,06136	0,04434	0,03217
60 %	0,14420	0,10884	0,07419	0,05488	0,04029

4-jadval

Darajali bahoning nazariy taqsimot funksiyadan chetlashishi (Eksponensial taqsimot, senzurlanish vaqt momenti Gamma taqsimotga ega)

Eksponensial taqsimot $(\lambda=1)$					
O‘rtacha senzurlanish darajasi	$n=100$	$n=200$	$n=500$	$n=1000$	$n=2000$
0 %	0,08579	0,06115	0,03845	0,02726	0,01946
10 %	0,08621	0,06252	0,03812	0,02977	0,01953
20 %	0,08723	0,06759	0,04052	0,03036	0,01976
30 %	0,09285	0,07075	0,04287	0,03127	0,02188
40 %	0,09649	0,07148	0,04833	0,03587	0,02782
50 %	0,11017	0,07976	0,05133	0,03658	0,02938
60 %	0,13820	0,10474	0,07406	0,05446	0,04113

Yuqoridagi 3- va 4-jadvallarni sinchkovlik bilan ko‘zdan kechirib, noinformativ (tasodifiy senzurlanish) modelda darajali baho tanlanma hajmi kichik bo‘lganida va senzurlanish darajasi yuqori bo‘lganida Kaplan-Meier bahosiga nisbatan taqsimot funksiyaga yaqinroq ekanligi haqida xulosa chiqarishimiz mumkin.

Endi (7) modelda $X_i \sim F(x,\alpha)=1-e^{-\frac{x}{\alpha}},\ x\geq 0, \alpha>0$, $Y_i \sim G(x,\alpha)=1-\left(1-F(x,\alpha)\right)^{\theta}, \theta>0$ va $L_i \sim K(x,\alpha)=\left(N(x,\alpha)\right)^{\beta}, \beta>0$ bo‘lgan holni qaraymiz. U holda (8)ga ko‘ra, biz $1-e^{-\frac{x}{\alpha}}=1-\left[1-\left(H(x)\right)^{\lambda}\right]^{\gamma}$ ga ega bo‘lamiz. Bu tenglikdan $H(x)$ni topib, uning uchun haqiqatga maksimal o‘xshashlik metodini qo‘llagan holda α parametr uchun quyidagi yarimparametrik bahoga ega bo‘lamiz:

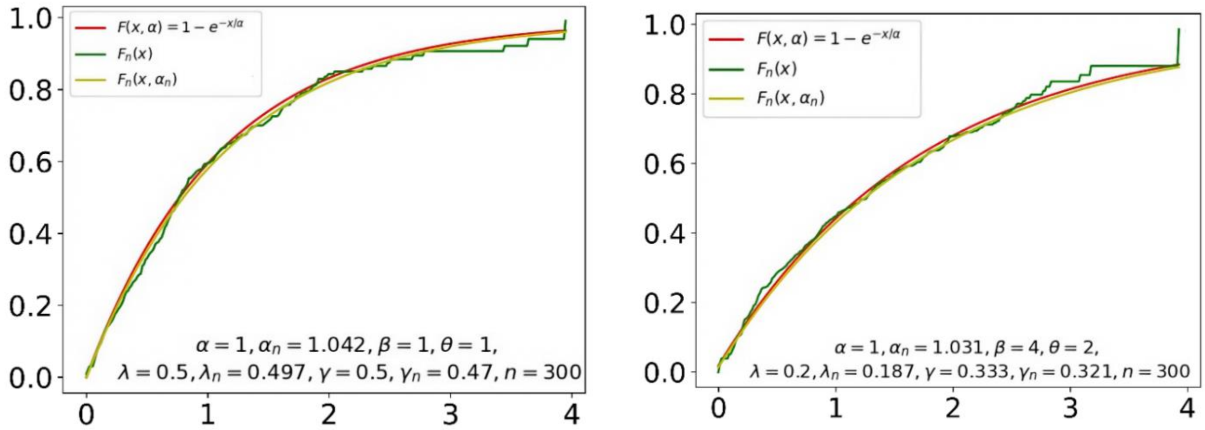
$$\alpha_n=\frac{1}{\lambda_n\gamma_n n}\cdot\sum_{i=1}^{n} Z_i\left(1-\frac{1-\lambda_n}{\left(H_n(Z_i)\right)^{\lambda_n}}\right). \tag{12}$$

(12) formuladan $F(x,\alpha)$ taqsimot funksiya uchun $F(x,\alpha_n)$ bahoga ega bo‘lamiz. Endi biz sonli usullar yordamida model bo‘yicha tanlanma beruvchi dastur tuzib, (10) baho va (12)ni qo‘llagan holda olingan bahoni o‘zaro taqqoslab ko‘ramiz. Bahoning qay darajada taqsimot funksiyaga yaqinligini baholash maqsadida (1) ifodaning 20000 ta tajribadagi o‘rtacha qiymati bo‘yicha jadval tuzib olamiz. Bu esa bahoning bosh to‘plam taqsimotidan uzoqlashishining “o‘lchovi” sifatida qaraladi.

5-jadval

Informativ modelda eksponensial taqsimotning yarimparametrik bahosining nazariy taqsimot funksiyadan uzoqlashishi

Eksponensial taqsimot ($\lambda = 1$)					
O‘rtacha senzurlanish darajasi	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
0 %	0,03041	0,02046	0,01400	0,00977	0,00681
10 %	0,03263	0,02160	0,01405	0,01002	0,00701
20 %	0,03353	0,02317	0,01443	0,01070	0,00717
30 %	0,03694	0,02502	0,01600	0,01083	0,00787
40 %	0,03938	0,02769	0,01740	0,01238	0,00874
50 %	0,04154	0,02815	0,01822	0,01366	0,00960
60 %	0,04657	0,03283	0,02056	0,01470	0,01029
70 %	0,05507	0,03903	0,02358	0,01696	0,01227
80 %	0,06493	0,04738	0,02953	0,02143	0,01494



a) Eksponensial taqsimot ($\beta = 1 \theta = 1$)   b) Eksponensial taqsimot $\beta = 4, \theta = 2$

4-rasm. Ikki tomonlama tasodifiy senzurlanishli informativ modelda eksponensial taqsimotni parametrik bahosini o‘rganish.

Yuqoridagi 2- va 5-jadval, 4- va 5-rasmlardan xulosa qilishimiz mumkinki, $F(x, \alpha_n)$ baho juda yaxshi ekan. Agarda F taqsimot funksiya eksponensial taqsimotga ega bo‘lsa, u holda (12) yordamida olingan yarimparametrik bahoni qo‘llash tavsiya qilinadi. Chunki yuqoridagilardan ko‘rinib turganidek, uning senzurlanishga nisbatan sezgirligi past hisoblanadi.

Ma’lumki, Donsker teoremasiga ko‘ra, $\left\{\sqrt{n}\cdot\left(F_n^{\ni}(x) - F(x)\right), x \in R^1\right\}$ tasodifiy jarayon o‘rtacha qiymati nol va kovariatsiyasi $F\left(\min(x_1, x_2)\right) - F(x_1)F(x_2)$ bo‘lgan $\mathbb{B}\left(F(x)\right)$ Broun ko‘prigiga kuchsiz yaqinlashadi. Bunga qo‘shimcha ravishda quyidagi teoremani keltirish mumkin:

Teorema 2 (A. N. Kolmogorov). *Agar $F(t)$ – uzluksiz taqsimot funksiya bo‘lsa, u holda barcha $t \in \mathbb{R}$ lar uchun*

$$\lim_{n\to\infty} P\left\{\sqrt{n}\sup_{u\in\mathbb{R}}\left|F(u) - F_n^{\ni}(u)\right| < t\right\} = K(t) \tag{13}$$

limit o‘rinli. Bu yerda $K(t) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2t^2}$ – Kolmogorov taqsimoti bo‘lib, limitning natijasi $F(t)$ taqsimot funksiyaga bog‘liq emas. Bu natijalar $F(t)$ taqsimot funksiya uchun ishonch intervallarini qurishda va $H_0 : F = F_0$, ko‘rinishidagi mezonlarni tekshirishda muhim hisoblanadi. Endi biz $F_n^{\ni}(x)$ empirik taqsimot funksiya o‘rniga (10) bahoni qo‘ysak, (13) ifodaning limit taqsimotida qanday o‘zgarish bo‘lishini qiziqib ko‘ramiz. Ayni shu maqsadda dissertatsiyaning **“O‘ng tomondan tasodifiy senzurlanishli informativ model bahosi va uning kovariatali umumlashmasi”** nomli uchinchi bobida $Q_n(x) = \sqrt{n}\cdot\left(F_n(x) - F(x)\right)$ jarayon qaraladi. Bu jarayonning kovariatsion tarkibi dissertatsiya ishining asosiy natijasi sifatida quyidagi teorema tarzida keltirib o‘tilgan:

Teorema 3. Faraz qilaylik, $q > 0$ bo‘lsin. U holda $\{Q_n(x), x \in D\}$ tasodifiy jarayonlar ketma-ketligi $\{A(x), x \in D\}$ markaziy Gauss jarayoniga kuchsiz yaqinlashadi va uning kovariatsiyasi $\forall x_1, x_2 \in D$ uchun quyidagicha bo‘ladi:

$$\operatorname{cov}\{A(x_1),A(x_2)\}=(1-F(x_1))(1-F(x_2))\times$$
$$\times\{a(x_1)a(x_2)\left[H(\min(x_1,x_2))-H(x_1)H(x_2)\right]+b(x_1)b(x_2)\times$$
$$\times p^{(0)}\left(1-p^{(0)}\right)+c(x_1)c(x_2)p^{(1)}\left(1-p^{(1)}\right)-$$
$$-2b(\min(x_1,x_2))c(\min(x_1,x_2))p^{(0)}p^{(1)}\},$$

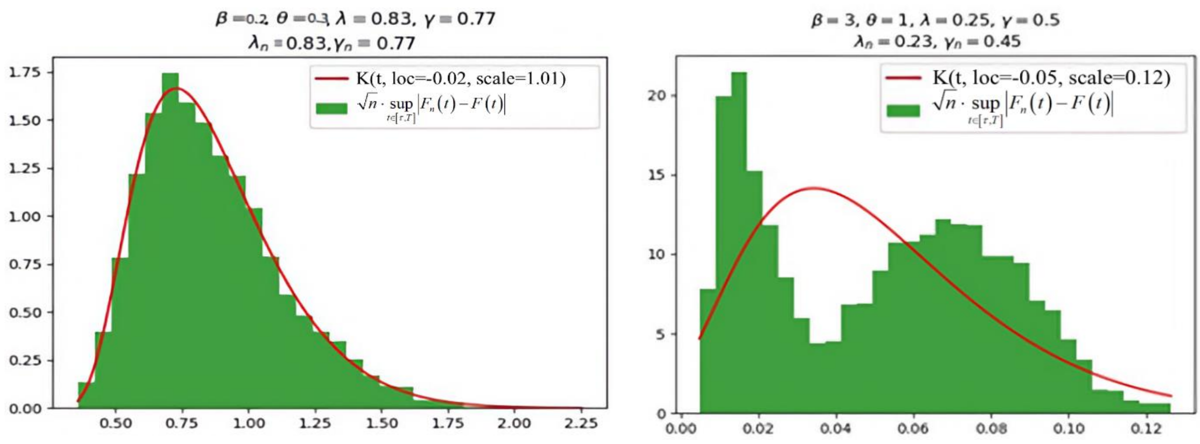
bu yerda

$$a(x)=p^{(1)}\left[(H(x))^{p^{(0)}}-H(x)\right]^{-1},$$

$$b(x)=-\left[\frac{p^{(1)}}{\left(1-p^{(0)}\right)}c(x)+\frac{a(x)}{\left(1-p^{(0)}\right)}H(x)\ln H(x)\right],$$

$$c(x)=-\frac{1}{\left(1-p^{(0)}\right)}\ln\left[1-(H(x))^{1-p^{(0)}}\right].$$

Yuqoridagi teoremadan ko‘rinib turibdiki, kovariatsion ko‘rinishi murakkab ko‘rinishga ega. Uning xossalarini analitik o‘rganish juda murakkab masala bo‘lgani uchun uni sonli usullar yordamida o‘rganamiz. Shuni eslatib o‘tamizki, yuqoridagi kovariatsiyada har ikkala tomondan senzurlanishni yo‘q deb hisoblasak ($\beta=\theta\equiv 0$, $\lambda=\gamma\equiv 1$, $p^{(0)}=p^{(2)}\equiv 0$, $p^{(1)}\equiv 1$, $H(x)\equiv F(x)$), u holda uning kovariatsion tarkibi Broun ko‘prigi bilan mos tushadi. Sonli o‘rganish maqsadida $\sqrt{n}\cdot\sup\limits_{t\in[\tau,T]}\left|F_n(t)-F(t)\right|$ tasodifiy miqdorning taqsimotini turli senzurlanish darajalarida Kolmogorov taqsimoti bilan taqqoslab ko‘ramiz.



(a) $\beta=0.2,\ \ \theta=0.3$ (b) $\beta=3,\ \ \theta=1$

5-rasm. $\sqrt{n}\cdot\sup\limits_{t\in[\tau,T]}\left|F_n(t)-F(t)\right|$ tasodifiy miqdor, $n=5000$.

Yuqorida keltirilgan gistogrammalarni (5-rasm) analiz qilib shuni ko‘rishimiz mumkinki, parametrlarning $\beta,\theta\in[0,1]$ qiymatlarida limit taqsimot Kolmogorov taqsimotidan deyarli farq qilmayapti. Ammo boshqa holatlarda, ya’ni senzurlanish darajasi oshishi bilan ular orasidagi farq sezilarli bo‘lmoqda. Demak, senzurlanish darajasi yuqori bo‘lganda Kolmogorov taqsimotidan foydalanib bo‘lmaydi.

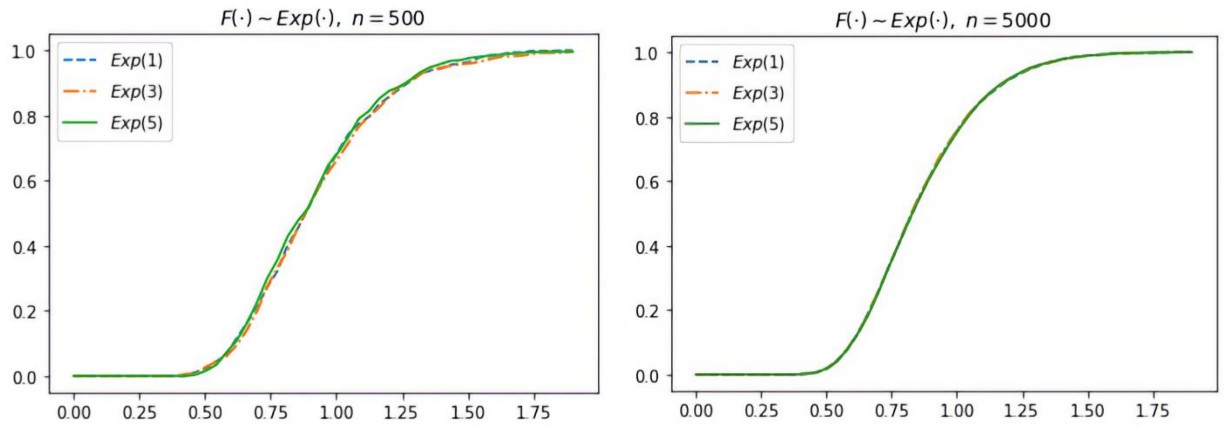
Endi muvofiqlik mezonlarini qurish uchun quyidagicha belgilash kiritib olamiz:

$$D_n^{ACL}=\sup_{t\le T}\left|F(t)-F_n^{ACL}(t)\right|=O\left(\sqrt{\frac{\ln\ln n}{n}}\right). \tag{14}$$

Biz yuqoridagi (14) intilish tezligini hisobga olib,

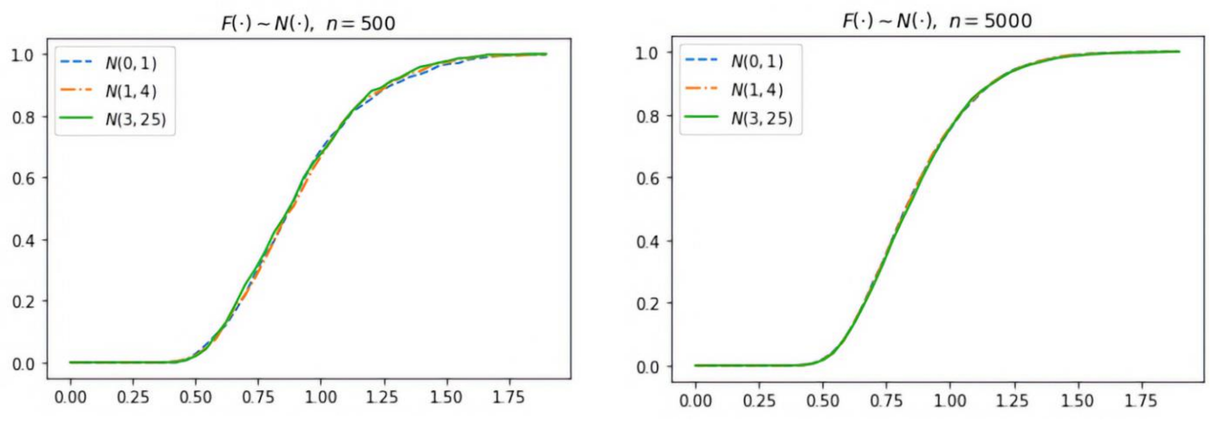
$$\sqrt{\frac{n}{\ln\ln n}}\cdot\sup_{t\le T}\left|F(t)-F_n^{ACL}(t)\right| \tag{15}$$

tasodifiy miqdorni hosil qilamiz. (15) tasodifiy miqdorning taqsimot funksiyasi va limit taqsimoti haqida xulosa chiqarish maqsadida (15)ning ko‘rinishida tanlanma beruvchi dastur tuzib, ushbu dastur yordamida turli $F(t)$ va n lar uchun tanlanmalar olamiz. Olingan tanlanmalar asosida xulosa chiqarishga urinamiz. (15)ning limit taqsimoti olingan bosh to‘plam taqsimoti $F(t)$ning parametrlariga bog‘liqmi yoki yo‘q ekanligini tekshirish uchun $F(t)$ sifatida eksponensial va normal taqsimotlarni tanlab, ular yordamida $n=500$ (6a-rasm) hamda $n=5000$ (6b-rasm) hajmli (15) ko‘rinishdagi tasodifiy miqdorni hosil qilib, ularning empirik taqsimot funksiyalarini chizib ko‘ramiz.



a) tanlanma hajmi n=500 b) tanlanma hajmi n=5000

6-rasm. $F(t)-$eksponensial taqsimlangan holda (15) ning limit taqsimoti.

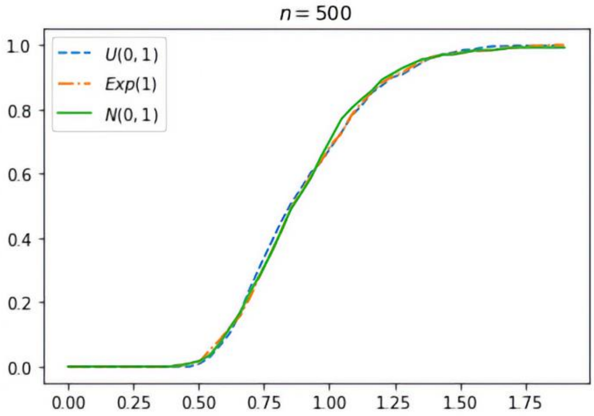


a) tanlanma hajmi n=500 b) tanlanma hajmi n=5000

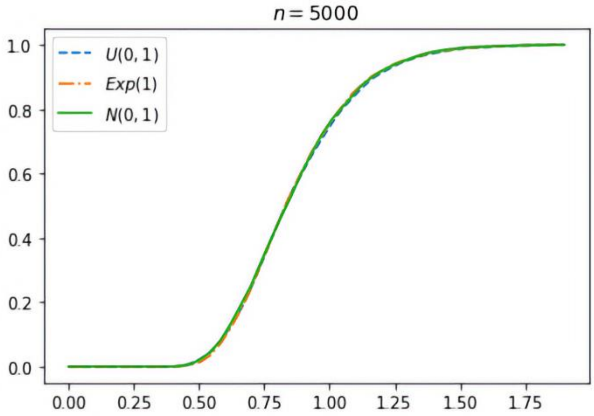
7-rasm. $F(t)-$normal taqsimlangan holda (15) ning limit taqsimoti.

Yuqoridagi 6- va 7-rasmlarni umumlashtirib (15)ning limit taqsimoti $F(t)$ ning parametrlariga bog‘liq emas deb xulosa chiqarishimiz mumkin. Yana shuni alohida ta’kidlab o‘tish kerakki, yuqoridagi chizmalar yoki umuman olganda alohida boshqa holat aytilmagan bo‘lsa, barcha chizmalar o‘rtacha 50 % senzurlanishli tanlanmalar bilan chizilgan.

Endi taqsimot funksiyalarni turli tanlab ko‘ramiz. Yuqoridagi parametrlarga bog‘liq emasligidan umumiylikka ziyon yetkazmagan holda barcha taqsimot funksiyalarni standart parametrda chizamiz.



a) tanlanma hajmi n=500

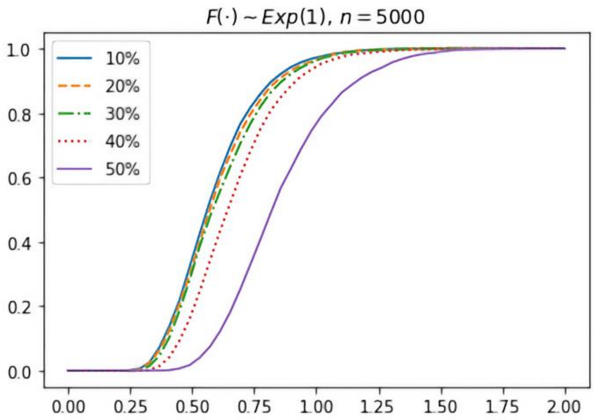


b) tanlanma hajmi n=5000

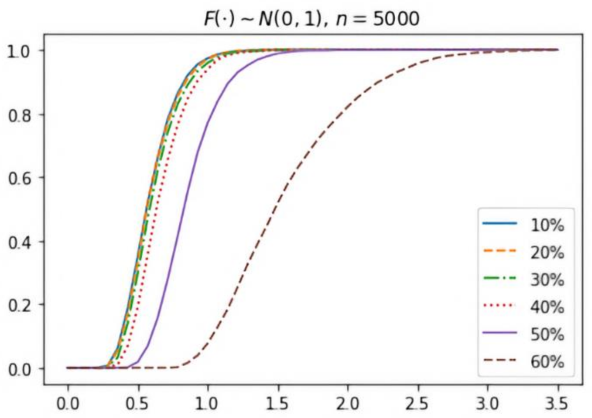
8-rasm. $F(t)-$turli taqsimlangan holda (15) ning limit taqsimoti.

Endi (15) da $F(t)$ ni tekis, eksponensial va normal olganimizda (15)ning taqsimot funksiyasi usma-ust tushmoqda. **Bu chizmalar asosida (15) $F(t)$ ga bog‘liq emas yoki bog‘liq bo‘ladigan bo‘lsa ham amaliy jihatdan sezilarli emas deb xulosa qilishimiz mumkin.** Buni, ayniqsa, tanlanma hajmi n=5000 bo‘lganda (8b-rasm) yaqqol ko‘rish mumkin.

Yuqorida alohida ta’kidlab o‘tganimizdek, barcha chizmalar senzurlanish darajasini o‘zgartirilmagan holda chizilgan edi. Endi limit taqsimotning senzurlanish darajasiga qay darajada bog‘liq ekanligini o‘rganish maqsadida bir turdagi $F(t)$ uchun (eksponensial va normal) turli senzurlanish darajasida limit taqsimotni chizamiz.



a) $F(t)-$eksponensial taqsimot



b) $F(t)-$normal taqsimot

9-rasm. Turli senzurlanish darajali tanlanmalarda (15) ning limit taqsimoti.

Grafiklarda masshtab xatoliklari bo‘lishi mumkinligini e’tiborga olib, uzil-kesil xulosaga kelish maqsadida dasturiy ta’minot yordamida quyidagi jadvalni tuzib olamiz. Unda turli $F(t)-$taqsimot funksiyalarda (15)ning limit taqsimoti qurilib, uning $t = 0,3;\ 0,5;\ 0,7;\ 1;\ 1,5$ bo‘lgandagi qiymatlari hisoblandi:

6-jadval

Limit taqsimotning turli nuqtadagi qiymatlari

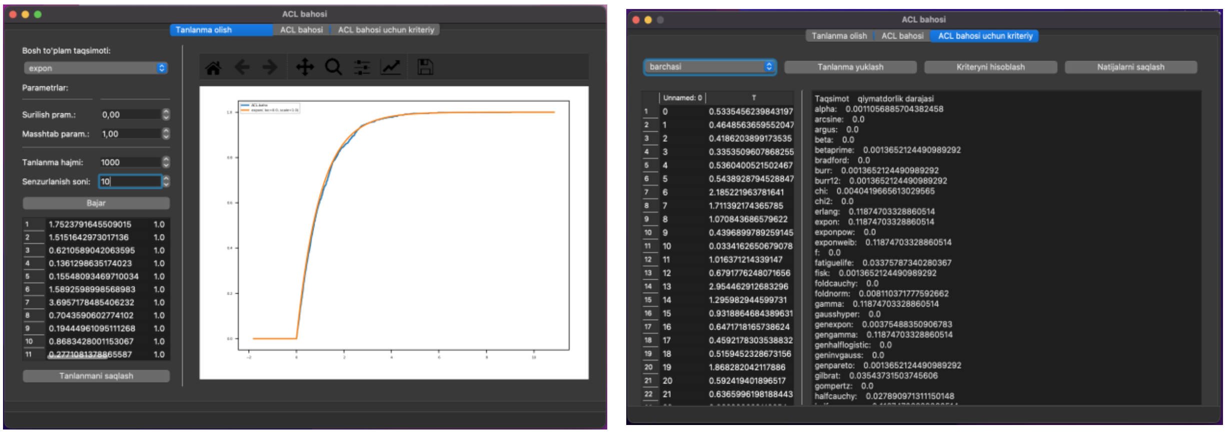
	0,3	0,5	0,7	1	1,5
N(1, 4)	0,0100	0,3455	0,7630	0,9749	0,9999
Exp(1)	0,0099	0,3449	0,7624	0,9747	0,9999
N(0, 1)	0,0100	0,3452	0,7621	0,9750	0,9999
Exp(1/4)	0,0100	0,3450	0,7623	0,9751	0,9999
N(3, 5)	0,0100	0,3454	0,7626	0,9751	0,9999
Exp(5)	0,0100	0,3450	0,7621	0,9747	0,9999
Exp(3)	0,0100	0,3456	0,7628	0,9749	0,9999

Yakuniy xulosa qilib aytadigan bo‘lsak, (15)ning limit taqsimoti bosh to‘plam taqsimot funksiyasi $F(t)$ ning parametrlariga va uning o‘ziga bog‘liq bo‘lmasdan senzurlanish darajasiga kuchli bog‘liq bo‘lar ekan. Senzurlanish darajasi 50 % va undan oshgani hamonoq limit taqsimotda keskin o‘zgarish ko‘rinib qolmoqda. Chiqarilgan xulosalar ayni kerak bo‘ladigan xossadir. Biroq limit taqsimotning senzurlanish darasiga juda kuchli bog‘langanligi limit taqsimotni topishda qiyinchilik tug‘diradi. Ya’ni senzurlanish darajasini limit taqsimotning parametri sifatida kiritish kerak bo‘ladi. Bahoning empirik taqsimotga qaraganda murakkab tarkibga ega ekanligi limit taqsimotning oshkor ko‘rinishini (analitik) topishga xalaqit beradi.

Bunday vaziyatda masalani amaliy jihatdan quyidagicha yechish mumkin:

1. Dasturiy mahsulot yordamida juda yuqori aniqlikda kritik nuqtalarning jadvalini tuzish (ma’lum senzurlanish darajalari uchun).
2. Dasturiy mahsulot unga berilgan tanlanma hajmi va senzurlanish darajasi asosida bir nechta sun’iy tanlanmalar qurib, ular yordamida limit taqsimotni sonli qurib olish va kerakli aniqlikda mezonni tekshirish.
3. $\sqrt{\dfrac{n}{\ln\ln n}}\cdot\sup\limits_{t\le T}\left|F(t)-F_n^{ACL}(t)\right|$ tasodifiy miqdordan ko‘rinib turganidek, senzurlanish darajasi oshishi bilan uning qiymati ham oshadi. Qiymatining o‘shish tezligini baholab, $\varphi(\theta)\sqrt{\dfrac{n}{\ln\ln n}}\cdot\sup\limits_{t\le T}\left|F(t)-F_n^{ACL}(t)\right|$ ko‘rinishdan $\varphi(\theta)$ funksiyani topish.

Ushbu ishda barcha yechimlar keltirib o‘tilgan. Xususan, ikkinchi holatni yechimi sifatida dasturiy mahsulot olindi (№ DGU 11510, № DGU 14754). Quyida dasturning tarkibiy qismlari haqida qisqacha tavsif keltirib o‘tamiz.



a) “Tanlanma olish” bo‘limi   b) “ACL bahosi uchun mezon” bo‘limi

10-rasm. O‘ng tomondan tasodifiy senzurlanishli model baholarini tahlil qiluvchi elektron dastur tasnifi.

10-rasmdan ko‘rinib turganidek, programma uchta oynadan iborat bo‘lib, Birinchisi oyna proporsional intensivlik modeli uchun sun’iy tanlanma hosil qiluvchi oyna bo‘lib, unda o‘ng tomonlama tasodifiy senzurlanishli informativ model uchun tanlanma olinadi. Uning bosh to‘plam taqsimoti bo‘limida musbat aniqlanadigan 63 turdagi taqsimotlar o‘rin olgan. Keyingi oynada (“ACL bahosi”) mavjud tanlanmani yuklab olib u asosida darajali baho va Kaplan-Meier baholarini taqqoslash ishlarini amalga oshirish hamda natijalarni saqlash imkonini beradi. Keyingi oynada excel formatida kiritilgan tanlanma asosida mezonda olg‘a surilayotgan bosh to‘plam taqsimoti (63 turdagi taqsimotlar mavjud) tanlanib, mezonni natijalarini ko‘rish va natijalarni saqlash mumkin.

Amaliy statistik masalalar bilan ishlayotganda nafaqat senzurlargan ma’lumotlar bilan, balki obyektni biror ta’sir ostida kuzatishga to‘g‘ri keladi. Bunday ta’sir faktorlariga misol qilib, ishlab chiqarish sohasida namlik, harorat, bosim, mexanik kuchlanish; meditsinada dorining miqdori, davolash usuli, bemorning jinsi, bemorning qon bosimi va yoshini keltirishimiz mumkin. Bunday ta’sir faktorlari **kovariata** deb ataladi.

Biz kovariatani $x=\left(x_1,x_2,\ldots,x_m\right)$ orqali belgilaymiz va umumiylikka ziyon yetkazmagan holda ularni o‘shish tartibida $0\le x_1\le x_2\le\ldots\le x_m\le 1$ deb faraz qilamiz. Ravshanki, n tanlanma hajmi bo‘lsa, u holda $m\le n$ bo‘ladi. Ushbu dissertatsiya ishida, asosan, kovariatalarning vaqtga bog‘liq bo‘lmagan va $m=n$ holi qaraladi.

Kompyuter modellashtirishida kovariata sifatida $[0,\ 1]$ kesmada tekis taqsimlangan tasodifiy miqdorlar olinadi.

Kovariatali to‘liq tanlanmalar deb quyidagi ko‘rinishda olingan tanlanmalarga aytiladi:

$$C_n=\left\{\left(T_1,x_1\right),\left(T_2,x_2\right),\ldots,\left(T_n,x_n\right)\right\},$$

bu yerda T_i –“ishdan chiqish” vaqt momenti, x_i –kovariata.

Kovariatali o‘ng tomondan senzurlangan ma’lumotlar deb quyidagi tanlanmaga aytiladi:

$$C_n = \{(Z_1,\delta_1,x_1),(Z_2,\delta_2,x_2),\ldots,(Z_n,\delta_n,x_n)\},$$

bu yerda $Z_i = min(T_i,Y_i)$, kuzatilmaning qiymati. $T_i -$"ishdan chiqish" vaqt momenti, $Y_i -$senzurlanish vaqt momenti, $\delta_i = I(T_i \le Y_i)$ hodisa indikatori va $x_i -$ kovariata.

Dissertatsiya ishining keyingi vazifasi quyidagi kovariatali bahoning xossalarini tadqiq qilishdir:

$$F_{xh}(t) = 1 - \left\{1 - \left[H_{xh}(t)\right]^{\lambda_{xh}}\right\}^{\gamma_{xh}}, t \ge 0. \tag{16}$$

bu yerda

$$\omega_{ni}(x;h_n) = \int_{x_{i-1}}^{x_i} \frac{1}{h_n}\pi\left(\frac{x-y}{h_n}\right)dy \Big/ \int_{x_0}^{x_n} \frac{1}{h_n}\pi\left(\frac{x-y}{h_n}\right)dy, i = 1,\ldots,n, \ \sum_{i=1}^{n}\omega_{ni}(x;h_n) = 1$$

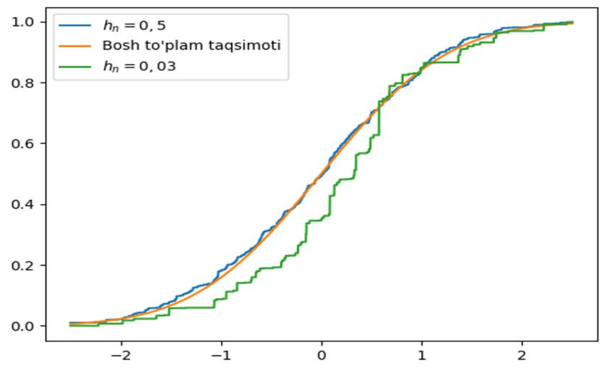
$$H_{xh}(t) = \sum_{i=1}^{n} I(\xi_i \le t)\omega_{ni}(x;h_n), \quad P_{xh}^{(m)} = \sum_{i=1}^{n}\delta_i^{(m)}\omega_{ni}(x;h_n), m = 0,1,2$$

$$\lambda_{xh} = 1 - p_{xh}^{(0)}, \gamma_{xh} = p_{xh}^{(1)}\left(1 - p_{xh}^{(0)}\right)^{-1}.$$

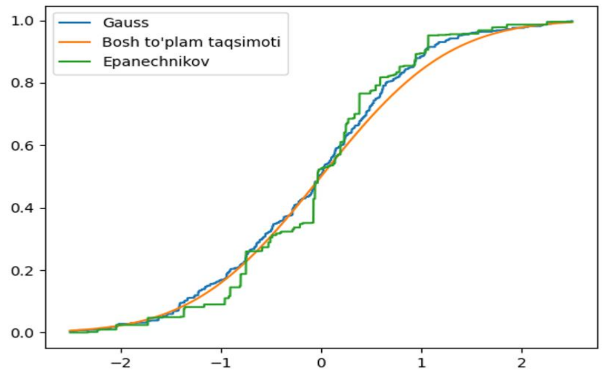
Xususan, $\omega_{ni}(x;h_n) = \dfrac{1}{n}$ (yoki kovariata bo‘lmasa) (16) oddiy empirik bahoga aylanadi.

Ushbu dissertatsiya ishida Gesser-Müller vazn funksiyasini hisoblash uchun to‘g‘ri to‘rtburchakli, Uchburchakli, Epanechnikov, Kvadratik, Uch vaznli, Gauss, Kosinusli yadro funksiyalardan foydalaniladi.

Endi bahoni xossalarini o‘rganishni boshlaymiz. Birinchi turli yadro funksiyalari va turli oyna kengligi bilan bahoni chizib ko‘ramiz.



a) Turli "oyna kengligi"da



b) Turli yadro funksiyada

11-rasm. Kovariatali bahoning turli "oyna kengligi" va yadrodagi chizmasi.

Yuqoridagi 11-rasm va 7-jadvallarni sinchiklab tekshirib, bahoning sifati tanlangan yadro funksiyasiga, bahoni qurishda tanlangan kovariataga hamda tanlangan "oyna kengligi" parametrlariga kuchli bog‘liq ekanligini kuzatishimiz mumkin. Bu yerda bahoni qurayotganimizda olingan tanlanma va kovariatani tanlash bizga bog‘liq bo‘lmasligi mumkin. Ammo bahoni qurayotganimizda yadro funksiyasi va "oyna kengligi" parametrlarini tanlash bizning ixtiyorimizdagi ishdir. Demak, bahoni imkoni qadar yaxshilash maqsadida optimal yadro funksiyasini va "oyna kengligi" parametrini topishga harakat qilamiz. Ustunlarning har birini diqqat bilan kuzatib optimal “oyna kengligi” mavjud ekanligini xulosa qilishimiz mumkin. Ular jadvalda qalin qilib belgilab ko‘rsatilgan. Boshqa ustunlarni ham sinchiklab kuzatib “oyna kengligi” parametrining kovariataning qiymatiga ham kuchli bog‘liq ekanligini ko‘ramiz.

7-jadval

Bahoning va nazariy taqsimotdan chetlashishi

(Koks modeli $S(t)=\left(S_0(t)\right)^{r(x,\beta)}$, $S_0(t)-$eksponensial, $r(x,\beta)=e^{\beta_0+\beta_1 \log x}$, $\beta_0=1, \beta_1=2$, $n=1000$, senzurlanish 10 %**)**

To‘g‘ri to‘rtburchakli yadro funksiya							
h_n \ x	0,1	0,2	0,3	0,5	0,7	0,9	1
0,1	0,08368	0,08458	0,08422	0,08394	0,08334	0,08498	0,11643
0,3	0,06046	0,05330	0,05046	0,04932	0,04827	0,06094	0,06848
0,5	0,04897	0,04548	0,04189	0,03778	0,04285	0,04935	0,05378
0,7	0,04263	0,03994	0,03813	**0,03669**	0,03813	0,04194	0,04507
0,9	0,03803	**0,03643**	0,03774	0,03800	**0,03687**	0,03806	0,03982
1	**0,03647**	0,03820	**0,03669**	0,03829	0,03793	**0,03750**	0,03839
3	0,03783	0,03837	0,03804	0,03838	0,03855	0,03808	**0,03725**
5	0,03824	0,03845	0,03818	0,03793	0,03877	0,03882	0,03814

Xulosa qilib aytadigan bo‘lsak, kovariatali bahoni turli kovariata, turli yadro va “oyna kengligi” parametrlarida tekshirib ko‘rib, har qanday holatda ham baho yadro funksiyasi “Gauss yadro funksiyasi” bo‘lganida eng yaxshi bo‘lishini ko‘rdik. Endigi asosiy maqsadimiz optimal oyna kengligi parametrini tanlashdan iboratdir. Optimal oyna kengligi parametrini topishning bir qancha metodlari mavjud bo‘lib, ulardan eng muhimlarini quyida muhokama qilamiz:

– terib ko‘rish metodi. Bunda barcha holatlarni o‘rganib shular ichidan eng yaxshisini tanlash;
– referent tanlash;
– o‘rtacha integral xatolikni minimallashtirish metodlari.

Endi yuqoridagi metodlarning har birining yutuq va kamchiliklarini muhokama qilamiz. Birinchi usul yetarli darajada aniqlikdagi oyna kengligi parametrini topib berishi mumkin. Ammo bu usul juda katta hisoblash ishlarini talab qiladi. Bu hattoki amaliy jihatdan imkonsiz bo‘lishi mumkin. Bundan tashqari bu usulda agarda umr davomiyligi va kovariata orasida kuchli kovaratsion bog‘lanish mavjud bo‘lsa, uni hisobga olgan holda terib ko‘radigan usulni qayta qurib chiqish zarurdir (bog‘lanish turiga ko‘ra). Aytilganlarni hisobga olsak bu usul optimal oyna kengligi parametrini topish uchun yaramaydi.

Ikkinchi referent tanlash usulida Beran bahosidagi kabi:

$$h_{n,opt}=\frac{C}{\sqrt[5]{n}}$$

formuladan foydalanish mumkin. Ammo mualliflar C parametrni tanlashning biror bir algoritmi haqida hech qanday ma'lumot keltirmaganlar. Faqatgina parametr empirik topiladi deyilgan xolos. Bu usul oyna kengligini kovariataga bog'liqligini o'z ichiga olmaydi.

Uchinchi usul o'rtacha minimal integral xatolik:

$$h_n = \hat{\sigma} \sqrt[5]{\frac{8\sqrt{\pi}R(K)}{2\mu_2(K)^2 n}} \tag{17}$$

bu yerda $\mu_2(K) = \int y^2 K(y)dy;$ $R(K) = \int K^2(y)dy$ va $\hat{\sigma}-$o'rtacha kvadratik chetlanish. Bu usulda ko'rinib turganidek $\mu_2(K)$ va $R(K)$ funksiyalarni hisoblash juda qiyin masaladir. Xulosa qilib aytadigan bo'lsak, yuqorida ta'kidlab o'tilgan usullar bizning talablarimizni qanoatlantirmaydi. Quyida biz optimal "oyna kengligi" parametrini topish uchun algoritm taklif etamiz.

Optimal oyna kengligi parametrini topish uchun biz taqsimotning teskari funksiyasi $F^{-1}(p)$ va "ishdan chiqish" vaqt momenti Z_i orasidagi "masofani" minimallashtirib ko'ramiz. Bu turdagi yondashuvdan tushunarliki, u kovariataning istalgan qiymatida ham o'rinli bo'ladi. Demak, maqsadimiz

$$\sum_{i=1}^{n} \delta_i \left| \hat{f}\left(F_{xh}(Z_i)\right) - Z_i \right| \tag{18}$$

yig'indining qiymatini minimallashtiruvchi $h-$"oyna kengligi" parametrini topishdan iboratdir. Bu yerda $\hat{f}(\bullet) = \frac{1}{n}\sum_{i=1}^{n}\omega(\bullet)Z_i$ umr davomiyligi funksiyasiga teskari funksiyasi, $F_{xh}(Z_i)-$taqsimot funksiya bahosi, $\omega(\bullet)-$vazn funksiya va δ_i- senzurlanishni ifodalovchi indikator.

Bu yerda $\omega(\bullet)$ o'rniga turli vazn funksiyalardan foydalanishimiz mumkin. Xusasan, ushbu dissertatasiya ishida Nadarya-Vatson (Nadaraya-Watson) vazn funksiyasidan foydalanamiz:

$$\omega_{ni}(x;a_n) = \frac{\sum_{i=1}^{n} Y_i \pi\left(\frac{x-x_i}{a_n}\right)}{\sum_{i=1}^{n} \pi\left(\frac{x-x_i}{a_n}\right)}.$$

Yuqoridagi vazn funksiyani hisoblashda a_n-"oyna kengligi" parametri sifatida (17) formulani ishlatamiz. Endi yuqoridagi aytilganlarni haqiqatda ham optimal "oyna kengligi" topib berishini tekshirish maqsadida bahoning modeli qurilib, ushbu model asosida quyidagi tekshirish ishlari olib borildi.

8-jadval

Optimal “oyna kengligi” parametri va ushbu parametr qo‘llanilganida bahoning nazariy taqsimotdan chetlashishi

(Koks modeli $S(t)=\left(S_0(t)\right)^{r(x,\beta)}$, $S_0(t)-$eksponensial, $r(x,\beta)=e^{\beta_0+\beta_1 \log x}$, $\beta_0=1, \beta_1=2$, $n=1000$, senzurlanish 10 %**)**

$K(\bullet)$ \ x		0,1	0,2	0,3	0,5	0,7	0,9	1
To‘g‘ri to‘rtburchakli yadro	$h_{n,opt.}$	0,9917	0,7404	2,9660	0,8954	0,7655	1,9410	2,9338
	D_n	0,0356	0,0360	0,0362	0,0363	0,0358	0,0365	0,0354

Yuqoridagi jadvalga ishlab chiqilgan algoritm yordamida topilgan optimal “oyna kengligi” parametri va bahoning taqsimot funksiyadan “chetlashishi” har bir yadro funksiyasi uchun keltirilgan bo‘lib, u yuqoridagi 7-jadvalda keltirilgan natijalar bilan aynan usma-ust tushmoqda. Demak, taklif etilayotgan usul yordamida haqiqatdan ham optimal oyna kengli parametrini topishimiz mumkin ekan.

ХУЛОСА

“To‘liq bo‘lmagan kuzatilmali sxemalarda statistik modellashtirish metodlari yordamida baholarni tahlil qilish” mavzusidagi falasafa doktori darajasini (PhD) olish uchun yozilgan dissertatsiya ishi bo‘yicha o‘tkazilgan tadqiqotlarning asosiy natijalari quyidagilardan iborat:

1. O‘ng tomondan tasodifiy senzuralanishli informativ modelda mavjud baholar (Kaplan-Meier, darajali baho, RR, PR) taqqoslandi. Ularning yutuq va kamchiliklari turli hollarda o‘rganildi. Informativ modelda darajali bahoning Kaplan-Meier bahosiga nisbatan afzalliklari ko‘rsatib berildi. Informativ bo‘lmagan holda darajali baho tanlanma hajmi kichik va senzurlanish darajasi yuqori bo‘lganida Kaplan-Meier bahosiga nisbatan afzalliklarga ega ekanligi aniqlandi.

2. Ikki tomonlama tasodifiy senzurlanishli informativ model bahosining senzurlanishning noma’lum parametrlariga qay darajada bog‘liq ekanligi tadqiq etildi. Tadqiqot ishlarini amalga oshirish maqsadida ikki tomonlama senzurlanishli informativ model uchun sun’iy tanlanma beruvchi dasturiy mahsulot ishlab chiqildi.

3. Ikki tomonlama tasodifiy senzurlanishli informativ modelda bosh to‘plam eksponensial taqsimlangan holda yarimparametrik baho ishlab chiqildi hamda uning xossalari tadqiq etildi. Ushbu bahoning senzurlanishga nisbatan “sezgirligi” past ekanligi ko‘rsatib berildi. Shuningdek, bosh to‘plam eksponensial taqsimlangan holda ushbu bahodan foydalanish kerakligi tavsiya qilindi (boshqa baholar bilan taqqoslagan holda).

4. Darajali baho uchun Kolmogorov muvofiqlik mezoni xossalari tadqiq qilindi. Erishilgan natijalari asosida muvofiqlik mezoni ishlab chiqildi va mezonni tekshiruvchi dasturiy mahsulot yaratildi.

5. Taqsimot funksiyasini kovariatali bahosining xususiyatlari o‘rganildi. Ushbu bahoning sifati “oyna kengligi” parametriga, tanlangan yadro funksiyasiga hamda kovariataga kuchli bog‘liq ekanligi ko‘rsatib berildi. Ushbu bahoning sifatini oshiruvchi optimal “oyna kengligi” parametrini topish algoritmi ishlab chiqildi.

НАУЧНЫЙ СОВЕТ DSc.03/31.03.2022.T/FM.10.04 ПО ПРИСУЖДЕНИЮ УЧЕНЫХ СТЕПЕНЕЙ ПРИ ИНСТИТУТЕ ФУНДАМЕНТАЛЬНЫХ И ПРИКЛАДНЫХ ИССЛЕДОВАНИЙ НАЦИОНАЛЬНОГО ИССЛЕДОВАТЕЛЬСКОГО УНИВЕРСИТЕТА”ТИИИМСХ”

НАВОИЙСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ ИНСТИТУТ

МАНСУРОВ ДИЛШОД РАВИЛОВИЧ

АНАЛИЗ ОЦЕНОК МЕТОДАМИ СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ В СХЕМАХ НЕПОЛНЫХ НАБЛЮДЕНИЙ

05.01.07- Математическое моделирование. Численные методы и комплексы программ

АВТОРЕФЕРАТ ДИССЕРТАЦИИ ДОКТОРА ФИЛОСОФИИ (PhD)
ПО ФИЗИКО-МАТЕМАТИЧЕСКИМ НАУКАМ

НАВОИ – 2023

Тема диссертации доктора философии (PhD) по физико-математическим наукам зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за № B2022.4.PhD/FM825.

Диссертация выполнена в Навоийском государственном педагогическом институте.

Автореферат диссертации на трех языках (узбекский, русский, английский (резюме)) размещен на веб-странице Научного совета (https://www.ifar.uz) и на Информационно-образовательном портале «ZiyoNet» (www.ziyonet.uz).

Научный руководитель:	**Абдушукуров Абдурахим Ахмедович** доктор физико-математических наук, профессор
Официальные оппоненты:	**Эшматов Фарход Хасанович** доктор физико-математических наук, профессор
	Нурмухамедова Наргиза Сайдиллаевна доктор философии (PhD) по физико-математическим наукам, доцент
Ведущая организация:	**Ташкентский государственный транспортный университет**

Защита диссертации состоится «___» ___ 2023 года в ___ часов на заседании Научного совета DSc.03/31.03.2022.T/FM.10.04 при Институте фундаментальных и прикладных исследований Национального исследовательского университета "ТИИМСХ". (Адрес: 100000, г. Ташкент, улица Кори Ниязова 39, Тел.: +998 71 237-09-61; e-mail: info@ifar.uz)

С диссертацией можно ознакомиться в Информационно-ресурсном центре Института фундаментальных и прикладных исследований Национального исследовательского университета "ТИИИМСХ" (зарегистрирована за № ___). (Адрес: 100000, г. Ташкент, улица Кори Ниязова 39, Тел.: +998 71 237-09-61).

Автореферат диссертации разослан « ___ » ___ 2023 года.
(протокол рассылки № ___ от « ___ » ___ 2023 года).

Б.Ж.Ахмедов
Председатель Научного
совета по присуждению ученых
степеней, д.ф.-м.н., профессор

Э.Х.Каримбаев
Ученый секретарь Научного
совета по присуждению ученых
степеней, доктор философии (PhD) по ф.-м.н.

А.Р.Хаётов
Председатель Научного семинара
при Научном совете по присуждению ученых
степеней, д.ф.-м.н., профессор

ВВЕДЕНИЕ (аннотация диссертации доктора философии (PhD))

Актуальность и обоснование темы диссертационного исследования. Большая часть научных и прикладных исследований, проводимых во всем мире, посвящена изучению вопросов статистики. Как известно, основной задачей статистического исследования является обработка результатов наблюдения, принятие решений и представление их в удобной форме. Однако в жизни не всегда имеется полная информация. Иногда необходимо иметь дело с неполными наблюдениями и использовать их для вывода одного или нескольких неизвестных законов распределения. Мы можем видеть неполные наблюдения в исследованиях продолжительности жизни, страхования, контроле качества, демографии, почвоведении, астрономии и других областях.

Задачи математической статистики, в частности непараметрическая оценка цензурированных наблюдений, имеют специфические характеристики по сравнению с выборками с полными наблюдениями. Построение оценок на основе неполных выборок со свойствами, подобными эмпирической функции распределения, стало предметом исследований многих статистиков. Если обратить внимание на оценки этих авторов, основанные на неполных наблюдениях, то можно увидеть, что оценка имеет более сложный состав по сравнению с эмпирической функцией распределения. Поэтому необходимы нестандартные методы и специальные подходы, учитывающие специфику рассматриваемой модели. Однако, несмотря на такую сложность, многие практические задачи решаются с помощью неполных селективных моделей. Это требует изучения свойств величин численными методами, если не решать их аналитически. С этой целью данная диссертация также направлена на частичное решение существующих проблем путем сравнения существующих и используемых в настоящее время оценок, выделения их сильных и слабых сторон и разработки критериев согласия.

Данное диссертационное исследование в определенной степени служит решению задач, обозначенных в Указе Президента Республики Узбекистан Ш.М.Мирзиёева №-УП-4947 от 7 февраля 2017 года «О стратегии действий по дальнейшему развитию Республики Узбекистан», в постановлении №-ПП-2789 от 17 февраля 2017 года «О мерах по дальнейшему совершенствованию деятельности Академии наук, организации, управления и финансирования научно-исследовательской деятельности», №-ПП-2909 от 20 апреля 2017 года «О мерах по дальнейшему развитию системы высшего образования», №-ПП-3682 от 27 апреля 2018 года «О мерах по дальнейшему совершенствованию системы практического внедрения инновационных идей, технологий и проектов», № ПП-4708 от 7 мая 2020 года «О мерах по повышению качества образования и развитию научных исследований в области математики» и № ПП-4851 от 6 октября 2020 года «О мерах по дальнейшему совершенствованию системы образования в области информационных технологий, развитию и интеграции научных исследований с IT-индустрией», а также для других нормативно-правовых актов по данной деятельности.

Соответствие исследования приоритетным направлениям развития науки и технологий республики. Данное исследование выполнено в соответствии с приоритетном направлении развития науки и технологий республики IV. «Математика, механика и информатика».

Степень научной изученности проблемы. Отметим, что систематическое использование терминов «цензурирование» и «усечение» в статистике принадлежат А. Хальду и Б. Фридману. Позже, в работах над цензурированными выборками, в основном случайного цензурирования справа, изучалались такими учеными, как N.Breslow, M.Burke, G.Campbell, S.Csörgő, D.Dabrowska, B.Efron, A. Földes, L.Rejto, R.Gill, M.Hollander, E.Kaplan, P.Meier, R.Liu, K. S. Sagidullayev, R.S.Murodov, F.A.Abdukalikov. Это связано с тем, что случайные цензурированные выборки справа часто встречаются в природе, а также с тем, что их методологически легко объяснить. Это также связано с тем, что работы о цензурирования с правой стороны можно легко переместить и на случай с левой стороны. Известно, что в условиях полных наблюдений непараметрическая оценка функции распределения является эмпирической функцией распределения, а ее свойства изучались многими такими авторами, как Ю.Благовещенский, R.Gill, R.Dudley, A.Dvoretzky, W.Stute, J. Van Rayzin, P. Major, S.Csörgő. Поиск непараметрических оценок на основе цензурированных данных стала серьезной статистической проблемой из-за их широкого использования на практике. В результате исследования в 1958 году Каплан и Мейер первыми предложили свою оценку PL (множительную) для правосторонней модели случайной цензурирования, которая впоследствии была глубоко исследована многими авторами. Существуют различные модифицированные версии PL-оценки. Изучению и использованию этой оценки посвящены работы многих ученых. Это также подтверждается списком литературы в конце диссертации. Среди многочисленных ученых следует особо выделить таких авторов, как Ю.Беляева, Ю.Благовещенского, R. D. Gill, O. Aalen, M. Csörgő, S. Csörgő, N. Breslow, B. Efron, P. K. Andersen, M. D. Burke, L. Horváth, A. Földes, L.Rejto, D. M. Dabrowska, J. K. Ghorai, V. Susarla, J. Van Rayzin, P. Major, E. G. Phadia, W. Stute, J. Wellner, N. Veraverbeke. В Узбекистане следует выделить таких ученых, как В. И. Романовский, С. Х. Сирожиддинов, А. А. Гафуров, К. С. Сагидуллаев, Р. С. Муродов, Ф. А. Абдукаликов. В работах А. Абдушукурова предложены оценки другой-степенной структуры при случайном цензурировании справа. Позже (в работах Csörgő) эта оценка названа оценкой ACL (ACL-Abdushukurov – Cheng – Lin). В данном и последующих пунктах называем эту оценку степенной.

Вышеупомянутая оценка Каплана-Мейера является очень широко используемой оценкой на практике, и ее ковариатное представление было обобщено Бераном. Как видно из цитированной литературы, эти оценки тщательно изучены и внедрены на практике. Однако степенная оценка имеет несколько преимуществ по сравнению с оценкой Каплана-Майера. Его обобщение, ковариантные представление взгляд, практически не изучено.

Данное диссертационное исследование направлено на изучение и в некоторой степени сравнение вышеуказанных оценок с использованием статистического моделирования и численных методов.

Благодаря работам многих из упомянутых выше ученых, связанных с исследованиями цензурированных выборок, в настоящее время возникла особая отрасль математической статистики, называемая статистикой неполных наблюдений.

Цель исследования. Исследование и сравнение аналогов оценки степенной структуры при цензурирования с обоих сторон с существующими оценками, а также разработка и изучение критериев согласия для нее. Изучение существования оптимального параметра «ширины окна» для коваритного обобщения оценки степенной структуры и разработке адаптивного алгоритма её нахождения.

Задачи исследования. В соответствии с поставленной целью планируется решить следующие задачи:

изучение характеристики оценок, имеющихся в информативной модели со случайным цензурированием справа. Сравнение преимуществ и недостатков этих оценок (сравнение степенной оценки с оценкой Каплана-Мейера);

изучение зависимости оценок в информативной модели от неизвестных параметров цензурирования, а также параметрическое оценивание экспоненциального распределения;

создание критериев согласия для степенной оценки и изучение их свойств;

изучение свойств ковариатной оценки функции распределения, а также влияние выбора параметра «ширина окна» на свойства ковариатной оценки функции распределения;

разработка алгоритма нахождения оптимального параметра «ширина окна» для ковариатной оценки функции распределения и создание программного продукта с использованием предложенного алгоритма.

Объект исследования. Степенная оценка в информативной модели случайного цензурирования справа и ее естественные обобщения. Оценка Каплана-Мейера и ее обобщение.

Предмет исследования. Сравнение степенной оценки и оценок Каплана-Мейера в модели случайной цензурования справа и изучение свойств критериев согласия для них, а также в определении ковариатного представления оценок и их свойств.

Методология и методы исследования. Для решения поставленных задач использовались методы математической статистики, теории вероятностей, математического и статистического моделирования (программирования).

Научная новизна исследования заключается в следующем:

параметрически оценено экспоненциальное распределение генеральной совокупности в информативной модели случайного цензурирования с двух сторон;

в статистике Колмогорова при замене эмпирической функции распределения степенной оценкой с помощью численных методов было показано, что предельное распределение статистики не зависит от функции распределения и ее параметров, а если и зависит, то практически не имеет значения;

предлагается критерий согласия, основанный на степенной оценке;

показана и изучена зависимость статистических характеристик ковариатной оценки функции распределения от вида ядерных функций, размера выборки, плана эксперимента, количества контрольных точек, регрессии;

предлагается адаптивный алгоритм для определения оптимального значения параметра “ширины окна”, которое позволяет построить качественную полупараметрическую оценку на основе результатов экспериментальных исследований.

Практические результаты исследования состоят в следующем:

разработан полнофункциональный программный продукт, проверяющий критерии согласии для степенной оценки;

разработан адаптивный алгоритм выбора оптимального параметра «ширина окна»;

создано программное обеспечение, реализующее предложенные алгоритмы.

Достоверность результатов исследования обеспечиваются:

правильным использованием математического аппарата и методами статистического моделирования для разработки критериев согласия и исследования их статистических свойств;

соответствием результатов статистического моделирования ранее известным по теоретическим результатам;

строгостью математического мышления;

математической (теоретической) корректностью полученных результатов;

соответствием полученных результатов с частичными результатами, полученными другими авторами.

Научно-практическая значимость результатов исследования. Научная значимость результатов исследования заключается в том, что научно-практические результаты, полученные в работе, будут использованы в дальнейших исследованиях в зависимости от неполноты наблюдений. Также его можно использовать для решения практических задач с помощью разработанного программного продукта.

Внедрение результатов исследования. Результаты сравнения оценок в информативных и неинформативных моделях, полученные в диссертационной работе, и разработанный критерий согласия для оценки пропорциональности интенсивностей были использованы в фундаментальном проекте № Ф4-40 «Исследование асимптотических свойств индексированных интегральные эмпирические процессы в классе размерных функций» (Справка №04/11-566 от 6 февраля 2023 года Национального университета Узбекистана имени Мирзо Улугбека). Применение научных результатов позволило сравнить оценку частной модели пропорциональных интенсивностей Козиола-Грина с оценкой Каплана-Мейера;

программный продукт, формирующий искусственную выборку для информативных и неинформативных моделей и проверяющий с их помощью критерии согласия, а также ковариационное представление эмпирического процесса при применении оценок моделей с неполным наблюдением, были использованы в фундаментальном молодежном проекте Yo-F4-07 на тему “Статистическое оценивание и проверка гипотез с использованием копула функций” (Справка № 04/11-2875 от 13 мая 2023 года Национального университета Узбекистана имени Мирзо Улугбека). Применение научных результатов позволило проверить критерии о семействе функций распределения.

Апробация результатов исследования. Результаты данного исследования были обсуждены на 8 научно-практических конференциях, в том числе на 3 международных и 5 республиканских.

Публикация результатов исследования. По теме диссертации опубликованы 16 научных работ, из них 5 научные статьи, в том числе 2 в зарубежных и 3 в республиканских журналах, рекомендованных высшей аттестационной комиссией Республики Узбекистан для защиты докторских диссертаций. Получены 3 свидетельства о регистрации программных продуктов для ЭВМ.

Объём и структура исследования. Данная диссертация состоит из введения, 3 глав, заключения и списка литературы. Общий объем диссертации составляет 123 страниц, в том числе 46 рисунков и 21 таблицы. Список литературы состоит из 110 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснована актуальность и востребованность темы диссертации, определено соответствие исследования приоритетным направлениям развития науки и технологий республики, приведен обзор зарубежных научных исследований по теме диссертации, степень изученности проблемы, сформулированы цели и задачи, выявлены объект и предмет исследования, изложены научная новизна и практические результаты исследования, раскрыта теоретическая и практическая значимость полученных результатов, даны сведения о внедрении результатов исследования, об опубликованных работах и о структуре диссертации.

Первая глава диссертации **«Цензурированные выборки»** содержит основные определения и важные понятия, необходимые для полного освещения диссертационного исследования. Также описаны функции распределения и их свойства, выборки в виде продолжительности жизни и алгоритмы их искусственной генерации, на которые часто ссылаются в диссертационной работе.

Мы можем использовать различные оценки распределения генеральной совокупности вместо $F(t)$. Тогда нам нужно будет проверить, насколько близка оценка к распределению генеральной совокупности или какая из нескольких оценок лучше. С этой целью получим меру “качества” оценки следующим образом:

$$d(F_1,F_2)=\sup_{-\infty<t<\infty}\left|F_1(t)-F_2(t)\right|. \quad (1)$$

Для достижения требуемой точности статистических экспериментов необходимо повторять эксперименты несколько раз. Соответственно, количество повторений будет:

$$N\le\left[\frac{t_\gamma^2}{\varepsilon^2}\right]+1. \quad (2)$$

В данной работе взяты $\varepsilon=0,01$ и $t_\gamma=0,99$. Подставляя их в (2), вычислим, что минимальное количество повторений равно $N=16600$. Все результаты в диссертации были получены как среднее значение не менее 20000 экспериментов.

Исследователь может наблюдать наблюдаемый объект от начала эксперимента до его завершения. Наблюдения в таком виде называются полными наблюдениями (классическая статистика) и завершаются каким-то событием, называемым **«отказом»**, что означает конец эксперимента. Отказ-это субъективное понятие, которое вводится и описывается пользователем в ходе эксперимента. Примерами отказов могут служить поломка технического устройства, состояние банкротства по экономическим вопросам, состояние смерти пациента в медицине и ряд других подобных ситуаций.

Но на практике мы не всегда можем наблюдать объект от начала до конца эксперимента. То есть, некоторые объекты выходят из наблюдения до окончания эксперимента (не испытав «отказ»). Об этих объектах остается только информация до момента их выхода из наблюдения. Пример такой ситуации – когда пациент, еще не закончивший лечение, покидает клинику по каким-либо причинам (смена постоянного адреса, перевод в другую клинику и т.п.). При этом имеется информация с момента поступления пациента в клинику до момента его выписки из клиники. Но никакой информации после его ухода нет. В клинике есть только момент, когда пациент ушел. Такие наблюдения называются **цензурированными наблюдениями**. Проще говоря, объект исчезает из поля зрения наблюдателя (покидает клинику) или попадает под наблюдение через некоторое время после начала испытания (момент времени появления первых симптомов заболевания у пациента неизвестен) – это примеры цензурированных наблюдений.

Первый случай-это **цензурирование справа**, а второй-**цензурирование слева**. Поскольку правостороннее цензурирование широко используется в вопросах, связанных с продолжительностью жизни, а также потому, что ее можно легко обобщить и для левой стороны. В этой работе в основном рассматриваются модели правостороннего цензурирования.

Правоцензурированные выборки (наблюдениями в виде продолжительности жизни) определяются следующим образом:

$$C^n = \left\{ \left(Z_i, \delta_i \right), i = \overline{1,n} \right\} \tag{3}$$

где $Z_i = \min\left(T_i, Y_i\right)$, $T_i -$момент «отказа», $Y_i -$момент цензурирования (момент прекращения наблюдения за i-м объектом), $\delta_i = I\left(T_i \le Y_i\right) -$индикатор события.

Если наблюдались моменты «отказов» всех объектов в эксперименте ($Z_i = T_i$, $\delta_i = 1$), то такое наблюдение называется **полным наблюдением** (классическая статистика). Если T_i неизвестен из-за прекращения наблюдения в некоторый момент времени $Y_i \le T_i$ ($Z_i = Y_i$, $\delta_i = 0$), то говорят, что наблюдение C^n **цензурировано справа.**

Существует три основных типа цензурирования справа:

Цензурирование I типа. В этом случае все объекты наблюдаются до момента некоторого заранее определенного момента времени Y. Это означает, что все объекты, которые не подвергались «отказу» до момента времени Y, будут считаться цензурированными справа ($Z_i = Y$, $\delta_i = 0$).

Цензурирование II типа. При этом наблюдение за экспериментом ведется до тех пор, пока не произойдет $k\,(k < n)$ «отказ». Все остальные объекты считаются цензурированными, а их цензурирующий момент принимается за $Z_i = T_{(k)}$, $\delta_i = 0$. Здесь $T_{(k)} -$момент времени, когда был подвергнут цензурированию последний наблюдаемый $k -$объект.

Цензурирование III типа. Каждое наблюдение записывается как $Z_i = \min\left(T_i,\ Y_i\right)$. Здесь – $T_i -$момент времени «отказа» и $Y_i -$момент времени цензурирования, а их функции распределения равны $G\left(t\right)$ и $F\left(t\right)$ соответственно. T_i и Y_i являются независимыми случайными величинами. Этот вид цензурирования, в свою очередь, делится на **информативный** и **неинформативный** виды. Если функции распределения $F\left(t\right)$ и $G\left(t\right)$ функционально связаны друг с другом, цензурирующая модель называется информативной, в противном случае - неинформативной.

В данной работе мы использовали цензурирование III типа. Это связано с тем, что этот тип цензурирования распространен в медицине, биостатистике, инженерных исследованиях и других практических вопросах.

Ничего нельзя сказать о виде $F\left(t\right) -$функции по этой полной $T = \left\{T_1, T_2, \dots, T_n\right\}$ выборке. Но неизвестную функцию распределения $F\left(t\right)$ можно оценить с помощью эмпирической функции распределения $F_n^{э}\left(t\right)$ следующим образом:

$$F_n^{э}\left(t\right) = \frac{1}{n}\sum_{i=1}^{n} I\left(T_i \le t\right) \tag{4}$$

где $I(A)-$ индикатор события A: $I(A)=\begin{cases}1, \text{ если A произойдет,}\\ 0, \text{ если } \overline{\text{A}} \text{ произойдет.}\end{cases}$

Основной целью диссертационной работы является исследование оценок эмпирической функции распределения, аналогичной (4), которая оценивает функцию распределения с использованием цензурированных выборок. Это обсуждается во второй главе диссертации **«Оценка функции распределения с использованием цензурированных выборок»**:

Оценка Каплана-Мейера. PL-оценка «product-limit» для функции распределения F, предложенная Капланом и Мейером в 1958 г.:

$$F_n^{PL}(t)=\begin{cases}1-\prod\limits_{\{j:Z_{(j)}\le t\}}\left(1-\dfrac{\delta_{(j)}}{n-j+1}\right), & t\le Z_{(n)},\\ 1, & t>Z_{(n)},\ \ \delta_{(n)}=1,\\ \text{неопределена}, & t>Z_{(n)},\ \ \delta_{(n)}=0,\end{cases} \tag{5}$$

где $Z_{(1)}\le\ldots\le Z_{(n)}$ – упорядоченная статистика, $\delta_i=I(X_i\le Y_i)$.

Степенная оценка. Предположим, что $\{(X_k,L_k,Y_k),k\ge 1\}$ последовательность независимых действительных значений (X,L,Y) и

$$S^{(n)}=\{(Z_i,\Delta_i),i=1,\ldots,n\}$$

наблюдаемая выборка. Где $Z_i=\max\{L_i,\min\{X_i,Y_i\}\}$, $\Delta_i=\left(\delta_i^{(0)},\delta_i^{(1)},\delta_i^{(2)}\right)$, $\delta_i^{(0)}=I\left(\min(X_i,Y_i)<L_i\right)$, $\delta_i^{(1)}=I(L_i\le X_i<Y_i)$, $\delta_i^{(2)}=I(L_i\le Y_i<X_i)$ и $I(A)-$ индикатор событий A. Для описания модели в таком виде введем случайные величины Z_i и $V_i=\min(X_i,Y_i)$, определим их функции распределения через H и N, которые являются функциями распределения максимума и минимума соответственно. Легко понять, что

$$\begin{gathered}H(x)=K(x)N(t),\\ N(t)=1-(1-F(t))(1-G(t)),t\in\mathbb{R}^1.\end{gathered} \tag{6}$$

Теперь, когда в дополнение к равенству (6) найдены такие положительные неизвестные параметры θ и β, потребуем, чтобы для всех $t\in\mathbb{R}^1$ было выполнено следующее равенство:

$$\begin{cases}1-G(t)=(1-F(t))^{\theta},\\ K(t)=(N(t))^{\beta},\end{cases} \tag{7}$$

где параметр β определяет степень цензурирования с левой стороны, а θ с правой стороны. Стремление значений этих параметров к нулю означает ослабление цензурирования с соответствующей стороны. Специальная модель с двусторонней случайной цензурой такого типа была введена А. А.

Абдушукуровым. Из формул (6) и (7) легко понять, что методом подстановки находим следующее выражение для F:

$$1-F(t)=\left[1-\left(H(t)\right)^{\lambda}\right]^{\gamma}, t\in\mathbb{R}^1, \tag{8}$$

где $\lambda=\dfrac{1}{1+\beta}$ и $\gamma=\dfrac{1}{1+\theta}$, а стремление λ и γ к 1 означает ослабление цензурирования с соответствующей стороны. Находим оценки выражений $\left(H(x),\lambda,\gamma\right)$ в формуле (8) с помощью выборка $S^{(n)}$ и построим полупараметрическую оценку функции распределения F. Для этого воспользуемся характеристикой модели:

Теорема 1. *Для выполнения равенства* (7), *необходимо и достаточно, чтобы случайные величины* Z_i *и* Δ_i *были независимыми.*

На основании равенства (7) и результатов теоремы 1 можно оценить параметры λ и θ следующим образом:

$$\lambda_n=1-p_n^{(0)},\quad \gamma_n=p_n^{(1)}\left(1-p_n^{(0)}\right)^{-1} \tag{9}$$

где $p_n^{(m)}=\dfrac{1}{n}\sum_{i=1}^{n}\delta_i^{(m)}$, $m=0,1,2$.

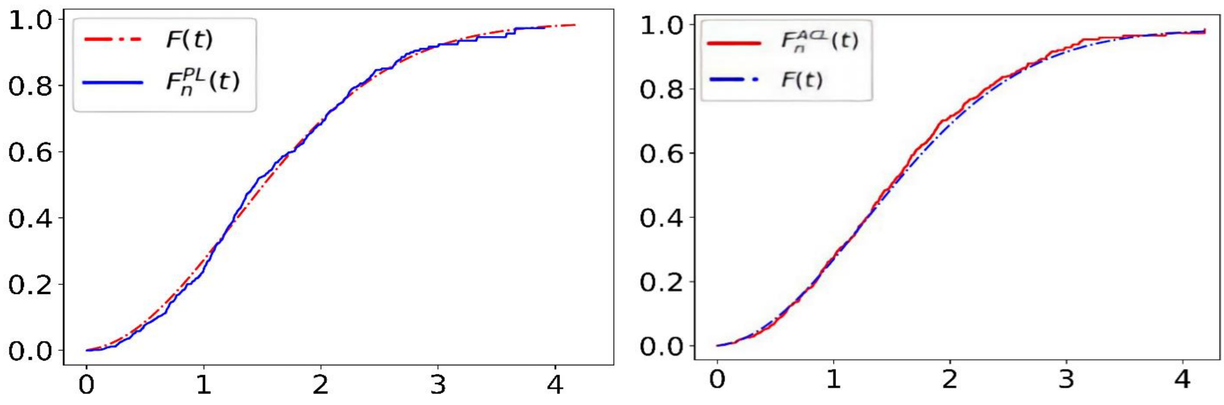
Подставляя (9) в выражение (8), получим следующую оценку для $F(t)$:

$$F_n^{ACL}(t)=1-\left[1-\left(H_n(t)\right)^{\lambda_n}\right]^{\gamma_n}, t\in\mathbb{R}^1. \tag{10}$$

Если в уравнении (7) $\beta=0$ или $\lambda=1$, то это означает отсутствие цензурирования с левой стороны. В этом случае оценка будет выглядеть так:

$$F_n(t)=1-\left[1-H_n(t)\right]^{\gamma_n}, t\in\mathbb{R}^1. \tag{11}$$

Используя язык программирования Python, построим график обеих оценок, выбирая модель цензурирования с правой стороны (Рисунок 1).



а) Оценка Каплана-Мейера　　　б) Степенная оценка

Рисунок 1. $F(t)$ – распределение Вейбулла, $n=500$, цензурирование 30%

Как видно из рисунка (рис. 1, а), б)), обе оценки хорошо аппроксимируют функцию распределения. Однако одним из основных недостатков оценки Каплана-Мейера является то, если в наибольшей выборке есть цензуририрование, оценка не может представлять функцию распределения справа от этой точки (рис. 1а)). В степенной оценке такого недостатка нет. Он определен по всей числовой оси. Теперь, чтобы выяснить, какая из оценок ближе к функции распределения, составим таблицу (таблица 1) из средних значений расстояния (1) между функцией распределения и оценкой 20000 опытов.

Таблица 1.

Отклонение оценки Каплана-Мейера от теоретической функции распределения ($F(t)$ – распределение Вейбулла)

Распределение Вейбулла ($c=2$)					
Средний уровень цензурирования	$n=100$	$n=200$	$n=500$	$n=1000$	$n=2000$
0%	0,08531	0,06062	0,03845	0,02729	0,01941
10 %	0,08704	0,06189	0,03940	0,02793	0,01973
20 %	0,09004	0,06393	0,04061	0,02876	0,02043
30 %	0,09591	0,06815	0,04323	0,03056	0,02174
40 %	0,10581	0,07545	0,04863	0,03451	0,02453
50 %	0,12317	0,09151	0,06073	0,04430	0,03249
60 %	0,15315	0,11967	0,08471	0,06535	0,05047
70 %	0,23079	0,18854	0,14359	0,11655	0,09443
80 %	0,36622	0,31974	0,26643	0,23165	0,20170

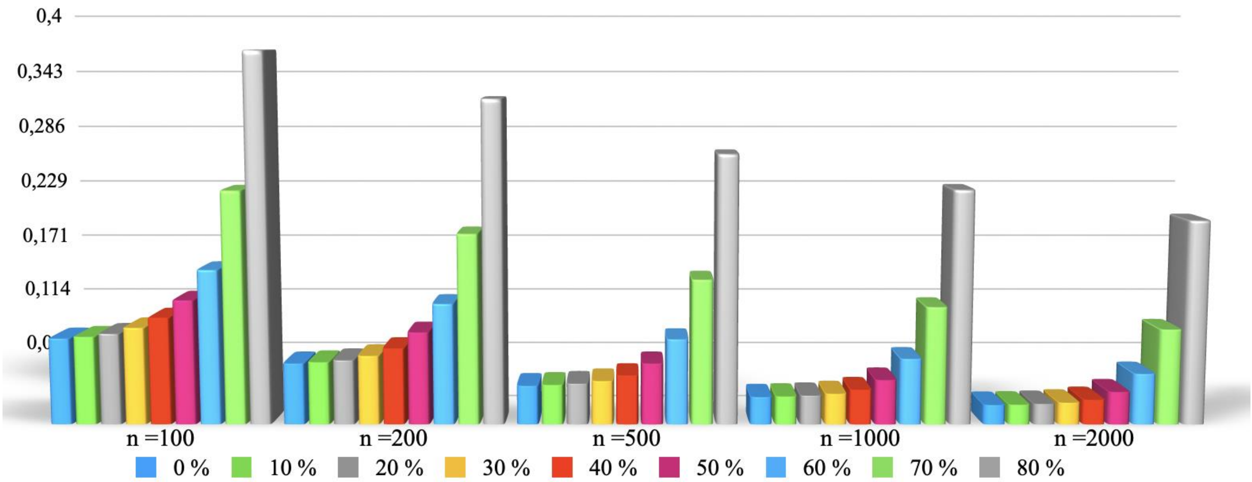


Рисунок 2. Диаграмма отклонения оценки Каплана-Мейера от теоретической функции распределения ($F(t)$ – распределение Вейбулла)

Таблица 2.

Отклонение степенной оценки от теоретической функции распределения ($F(t)$ – распределение Вейбулла)

Распределение Вейбулла ($c = 2$)					
Средний уровень цензурирования	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
0 %	0,08526	0,06053	0,03840	0,02726	0,01940
10 %	0,08569	0,06089	0,03896	0,02778	0,0195
20 %	0,08664	0,06127	0,03952	0,02803	0,01983
30 %	0,08752	0,06374	0,04087	0,02976	0,02095
40 %	0,09111	0,06467	0,04368	0,03271	0,02337
50 %	0,10931	0,07905	0,05118	0,03680	0,02629
60 %	0,15230	0,11517	0,08066	0,06104	0,04614
70 %	0,19078	0,15876	0,12329	0,10141	0,08323
80 %	0,22785	0,20286	0,17335	0,15289	0,13364

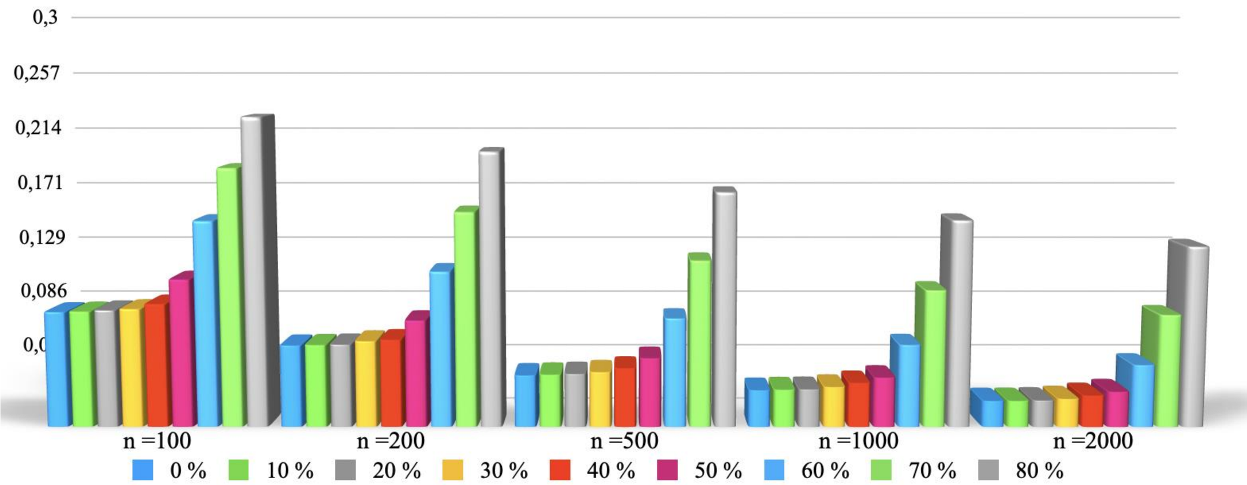


Рисунок 3. Диаграмма отклонения степенной оценки от теоретической функции распределения ($F(t)$ – распределение Вейбулла)

Таблицы 1 и 2, а также их диаграммы (рис.1 и рис. 2) показывают разницу между функцией распределения и оценкой при 9 различных степенях цензурирования для каждого размера выборки. По ним видно, что разница между оценкой и функцией распределения экспоненциально возрастает с увеличением степени цензурирования. Сравнивая каждый столбец и каждую строку в таблицах, можем уверенно сделать заключение о том, что степенная оценка в информативной модели ближе к функции распределения, чем оценка Каплана-Мейера.

Вышеуказанные сравнения были сделаны в информативной модели. В информативной модели было показано, что степенная оценка лучше, чем оценка Каплана-Мейера. Также сравнили оценки в неинформативной (случайное цензурирование) модели. Для этого использовали экспоненциальное распределение со стандартным параметром и момент времени цензурирования Гамма-распределение с желаемым параметром (найденным из уравнения $\int_{-\infty}^{+\infty}\left(1-G\left(t,\ \theta\right)\right)f\left(t\right)dt=1-r$), для того чтобы создать случайную цензурированную выборку и использовать их для создания следующей таблицы «расстояния» между функцией распределения и оценкой:

Таблица 3.

Отклонение оценки Каплана-Мейера от теоретической функции распределения (экспоненциальное распределение, цензурирующее момент времени, которое имеет гамма-распределение)

Экспоненциальное распределение $(\lambda=1)$					
Средний уровень цензурирования	$n=100$	$n=200$	$n=500$	$n=1000$	$n=2000$
0 %	0,08534	0,06105	0,03827	0,02737	0,01933
10 %	0,08718	0,05921	0,03897	0,02788	0,01970
20 %	0,08726	0,06213	0,03923	0,02807	0,01995
30 %	0,09448	0,07071	0,04418	0,03124	0,02241
40 %	0,10656	0,07779	0,05115	0,03683	0,02643
50 %	0,12372	0,09155	0,06136	0,04434	0,03217
60 %	0,14420	0,10884	0,07419	0,05488	0,04029

Таблица 4.

Отклонение оценки Каплана-Мейера от теоретической функции распределения (экспоненциальное распределение, цензурирующее момент времени, которое имеет гамма-распределение)

Экспоненциальное распределение $(\lambda=1)$					
Средний уровень цензурирования	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
0 %	0,08579	0,06115	0,03845	0,02726	0,01946
10 %	0,08621	0,06252	0,03812	0,02977	0,01953
20 %	0,08723	0,06759	0,04052	0,03036	0,01976
30 %	0,09285	0,07075	0,04287	0,03127	0,02188
40 %	0,09649	0,07148	0,04833	0,03587	0,02782
50 %	0,11017	0,07976	0,05133	0,03658	0,02938
60 %	0,13820	0,10474	0,07406	0,05446	0,04113

Внимательно изучив таблицы 3 и 4, можем сделать вывод, что в неинформативной (случайное цензурирование) модели степенная оценка ближе к функции распределения по сравнению с оценкой Каплана-Мейера, когда размер выборки мал, а уровень цензурирования высок.

Теперь в модели (7) рассмотрим случай, когда $X_i \sim F(x,\alpha) = 1 - e^{-\frac{x}{\alpha}},\ x \ge 0,\ \alpha > 0$, $Y_i \sim G(x,\alpha) = 1 - \left(1 - F(x,\alpha)\right)^{\theta}, \theta > 0$ и $L_i \sim K(x,\alpha) = \left(N(x,\alpha)\right)^{\beta}, \beta > 0$. Тогда согласно (8), имеем $1 - e^{-\frac{x}{\alpha}} = 1 - \left[1 - \left(H(x)\right)^{\lambda}\right]^{\gamma}$. Находя из этого уравнения $H(x)$ и применяя к нему метод максимального правдоподобия, получим полупараметрическую оценку параметра α:

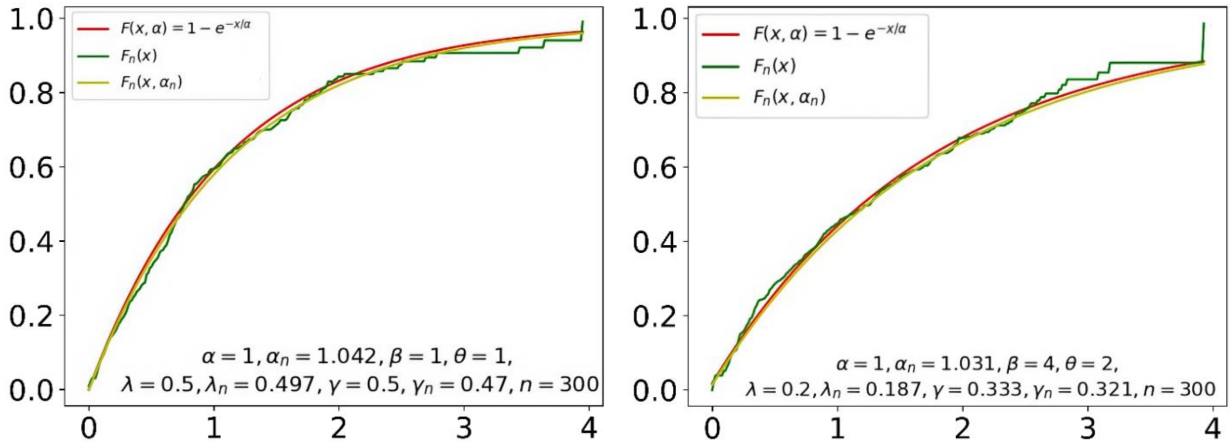
$$\alpha_n = \frac{1}{\lambda_n \gamma_n n} \cdot \sum_{i=1}^{n} Z_i \left(1 - \frac{1-\lambda_n}{\left(H_n(Z_i)\right)^{\lambda_n}}\right). \qquad (12)$$

Из формулы (12) для функции распределения $F(x,\alpha)$ получим оценку $F(x,\alpha_n)$. Теперь разработаем программу, дающую выборку по модели, и сравним между собой оценку, полученную с применением (10), и оценку, полученную с применением (12). Чтобы оценить, в какой степени они приближается к функции распределения, построим таблицу для среднего значения выражения (1) в эксперименте 20 000 опытов. Это рассматривается как “мера” отклонения оценки от функции распределения.

Таблица 5.

Отклонение полупараметрической оценки экспоненциального распределения от теоретической функции распределения в информативной модели

Экспоненциальное распределение ($\lambda = 1$)					
Средний уровень цензурирования	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
0 %	0,03041	0,02046	0,01400	0,00977	0,00681
10 %	0,03263	0,02160	0,01405	0,01002	0,00701
20 %	0,03353	0,02317	0,01443	0,01070	0,00717
30 %	0,03694	0,02502	0,01600	0,01083	0,00787
40 %	0,03938	0,02769	0,01740	0,01238	0,00874
50 %	0,04154	0,02815	0,01822	0,01366	0,00960
60 %	0,04657	0,03283	0,02056	0,01470	0,01029
70 %	0,05507	0,03903	0,02358	0,01696	0,01227
80 %	0,06493	0,04738	0,02953	0,02143	0,01494



а) Экспоненциальное распределение ($\beta = 1\theta = 1$)   б) Экспоненциальное распределение $\beta = 4, \theta = 2$

Рисунок 4. Исследование параметрического оценивания экспоненциального распределения в информативной модели с двусторонней случайного цензурирования

Из таблиц 2 и 5 и рисунков 4 и 5 можно сделать вывод, что $F(x, \alpha_n)$ оценка очень хорошая. Если функция распределения F имеет экспоненциальное распределение, то рекомендуется использовать полупараметрическую оценку, полученную с помощью (12). Потому что, как видно из вышесказанного, его чувствительность к цензурированию низка.

Известно, что по теореме Донскера $\left\{\sqrt{n}\cdot\left(F_n^{Э}(x) - F(x)\right), x \in R^1\right\}$ случайный процесс слабо приближается к броуновскому мосту $\mathbb{B}\left(F(x)\right)$ с нулевым средним и ковариацией $F\left(\min\left(x_1, x_2\right)\right) - F(x_1)F(x_2)$. Кроме того, можно привести следующую теорему:

Теорема 2 (А. Н. Колмогоров). *Если существует непрерывная функция распределения $F(t)$, то для всех $t \in \mathbb{R}$ уместно следующее равенство:*

$$\lim_{n\to\infty} P\left\{\sqrt{n}\sup_{u\in\mathbb{R}}\left|F(u) - F_n^{э}(u)\right| < t\right\} = K(t) \tag{13}$$

где $K(t) = \sum_{k=-\infty}^{\infty}(-1)^k e^{-2k^2t^2}$ – распределение Колмогорова, а результат предела не зависит от функции распределения $F(t)$. Эти результаты важны при построении доверительных интервалов для функции распределения $F(t)$ и при проверке критериев $H_0 : F = F_0$. Теперь нас интересует, как изменится предельное распределение, если вместо эмпирической функции распределения $F_n^{Э}(x)$ поставить оценку (10). Для этого в третьей главе диссертации **«Оценка информативной модели со случайной цензурой справа и ее ковариантное обобщение»** рассмотрен процесс $Q_n(x) = \sqrt{n}\cdot\left(F_n(x) - F(x)\right)$. Ковариационная структура этого процесса представлена как основной результат диссертации в виде следующей теоремы:

Теорема 3. Предположим, что $q>0$. Тогда последовательность случайных процессов $\{Q_n(x), x\in D\}$ слабо аппроксимирует центральный гауссовский процесс $\{A(x), x\in D\}$ и ее ковариация равна для $\forall x_1, x_2 \in D$:

$$\operatorname{cov}\{A(x_1), A(x_2)\} = (1-F(x_1))(1-F(x_2))\times$$
$$\times\{a(x_1)a(x_2)\left[H(\min(x_1,x_2)) - H(x_1)H(x_2)\right] + b(x_1)b(x_2)\times$$
$$\times p^{(0)}\left(1-p^{(0)}\right) + c(x_1)c(x_2)p^{(1)}\left(1-p^{(1)}\right) -$$
$$-2b(\min(x_1,x_2))c(\min(x_1,x_2))p^{(0)}p^{(1)}\},$$

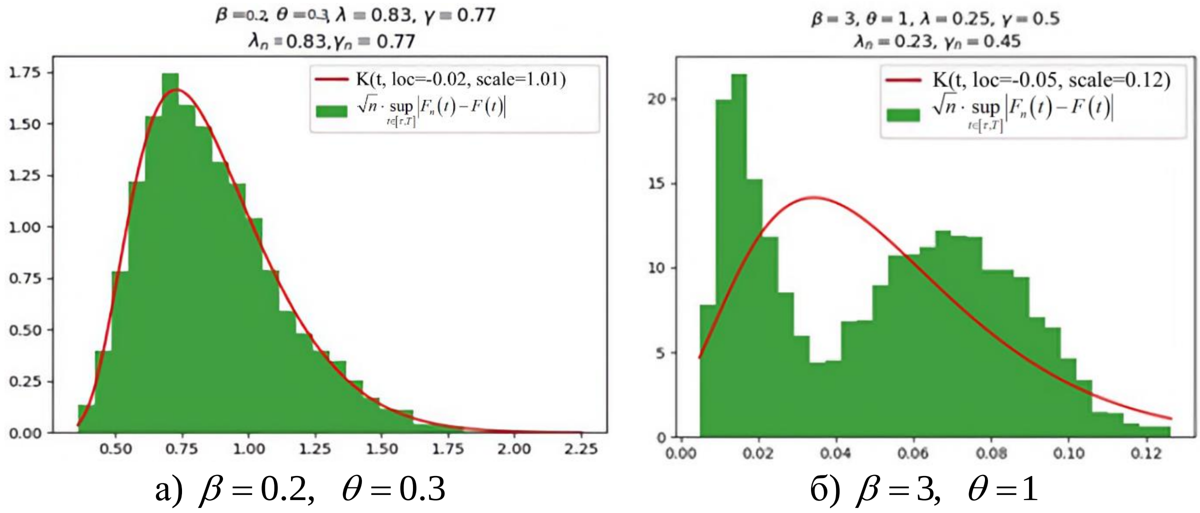
здесь

$$a(x) = p^{(1)}\left[(H(x))^{p^{(0)}} - H(x)\right]^{-1},$$

$$b(x) = -\left[\frac{p^{(1)}}{\left(1-p^{(0)}\right)}c(x) + \frac{a(x)}{\left(1-p^{(0)}\right)}H(x)\ln H(x)\right],$$

$$c(x) = -\frac{1}{\left(1-p^{(0)}\right)}\ln\left[1-(H(x))^{1-p^{(0)}}\right].$$

Как видно из приведенной выше теоремы, ковариационная форма имеет сложную структуру, поэтому ее свойства изучаются с помощью численных методов. Напоминаем, если предположить, что указанная выше ковариация не имеет цензурирования с обеих сторон ($\beta=\theta\equiv 0$, $\lambda=\gamma\equiv 1$, $p^{(0)}=p^{(2)}\equiv 0$, $p^{(1)}\equiv 1$, $H(x)\equiv F(x)$), то ее структура соответствует броуновскому мосту. Для численного изучения сравним распределение случайной величины $\sqrt{n}\cdot\sup\limits_{t\in[\tau,T]}\left|F_n(t)-F(t)\right|$ с распределением Колмогорова при разных степенях цензурирования.



а) $\beta=0.2,\quad \theta=0.3$ б) $\beta=3,\quad \theta=1$

Рисунок 5. Случайная величина $\sqrt{n}\cdot\sup\limits_{t\in[\tau,T]}\left|F_n(t)-F(t)\right|$, $n=5000$

Анализируя приведенные выше гистограммы (рис. 5), видим, что предельное распределение при значениях параметров $\beta,\theta \in [0,1]$ не отличаются от распределения Колмогорова. В других случаях, т. е. при повышении степени цензурирования, разница между ними становится существенной. Поэтому распределение Колмогорова нельзя использовать при высоком уровне цензурирования.

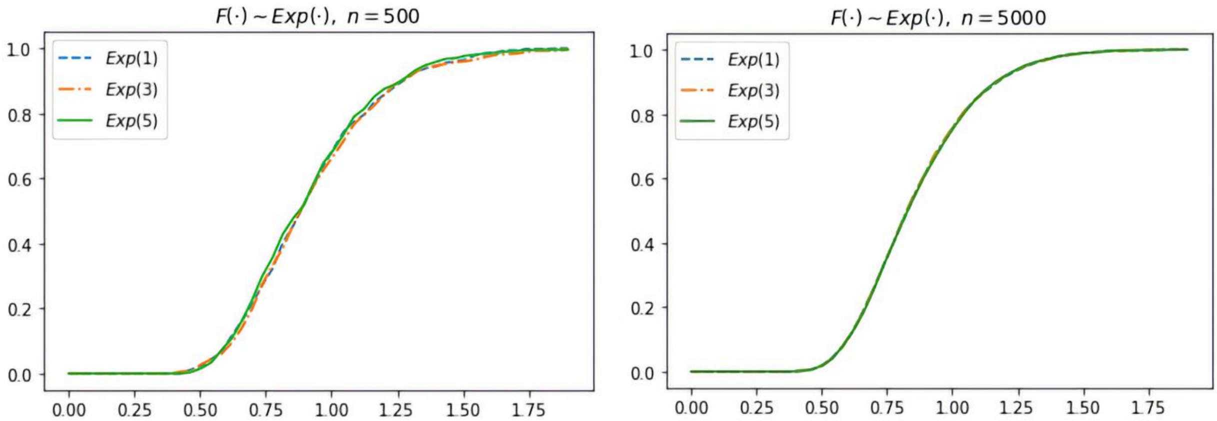
Чтобы построить критерии согласия, ввели следующие обозначения:

$$D_n^{ACL} = \sup_{t \le T} \left| F(t) - F_n^{ACL}(t) \right| = O\left(\sqrt{\frac{\ln \ln n}{n}} \right). \tag{14}$$

учитывая скорость стремления (14), генерируем случайную величину:

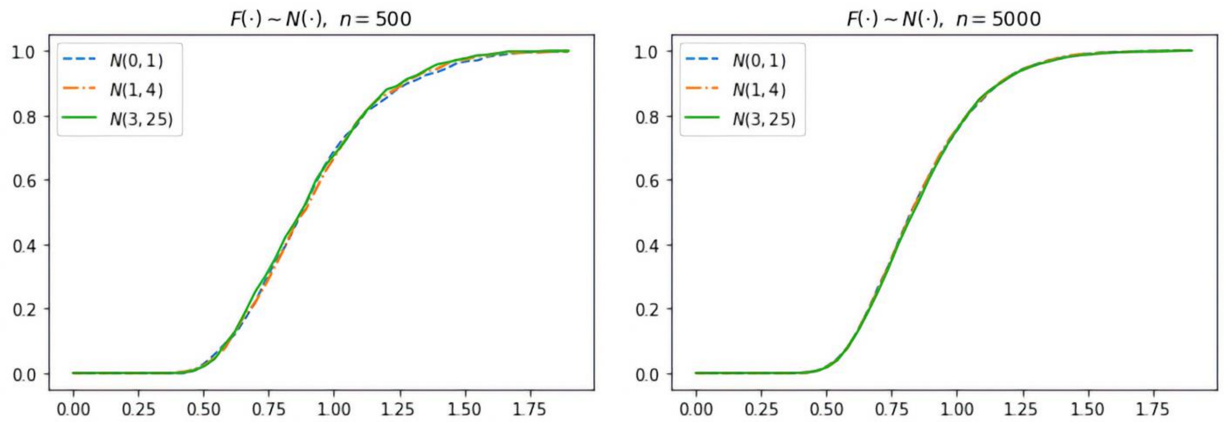
$$\sqrt{\frac{n}{\ln \ln n}} \cdot \sup_{t \le T} \left| F(t) - F_n^{ACL}(t) \right|. \tag{15}$$

Для того чтобы сделать вывод о предельном распределении случайной величины (15), разработаем программу, дающую выборку в виде (15), с помощью этой программы получим выборки для различных $F(t)$ и n. Сделаем выводы на основе полученных выборок. Чтобы проверить, зависит ли полученное предельное распределение (15) от параметров $F(t)$, в качестве $F(t)$ были выбраны экспоненциальное и нормальное распределения. Генерируем случайную величину в виде (15), размер которой n=500 (рис. 6а) и n=5000 (рис. 6б рис.) и построим их эмпирические функции распределения.



а) Объем выборки n=500 б) Объем выборки n=5000

Рисунок 6. Предельное распределение (15), когда $F(t)$ – распределена экспоненциально

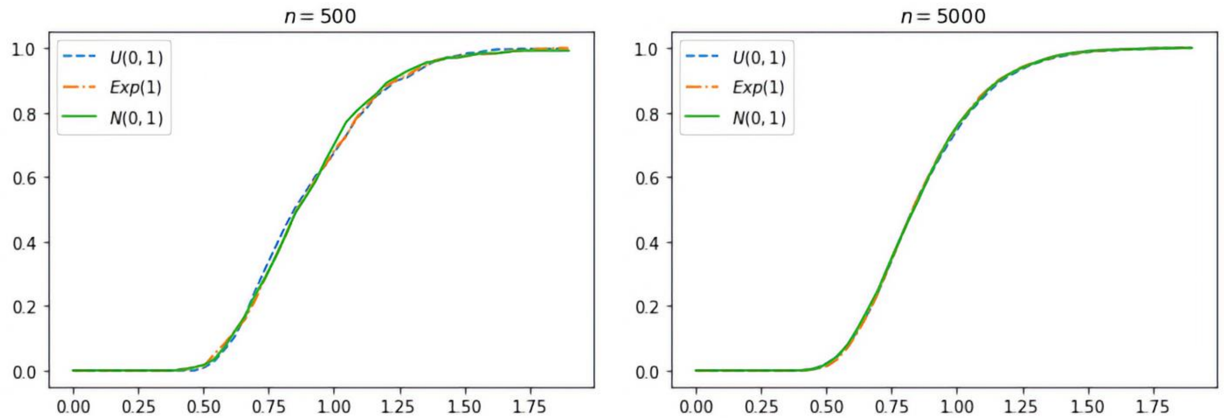


а) Объем выборки n=500 б) Объем выборки n=5000

Рисунок 7. Предельное распределение (15), когда $F(t)-$ нормально распределена

Теперь, обобщая рисунки 6 и 7, можно сделать вывод о том, что предельное распределение (15) не зависит от параметров $F(t)$. Следует также отметить, что все рисунки изображены в среднем на 50% цензурирования, если не указано иное.

Теперь давайте посмотрим на выбор различных функций распределения. Построим все функции распределения в параметре по умолчанию без ущерба для общности, поскольку они не зависят от параметров, как описано выше.



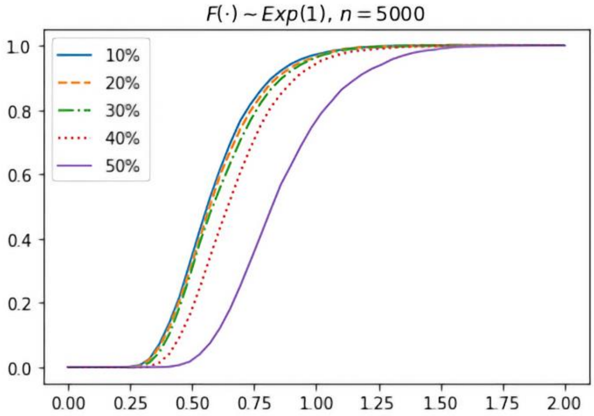
а) Объем выборки n=500 б) Объем выборки n=5000

Рисунок 8. Предельное распределение (15) с различными $F(t)-$ распределениями

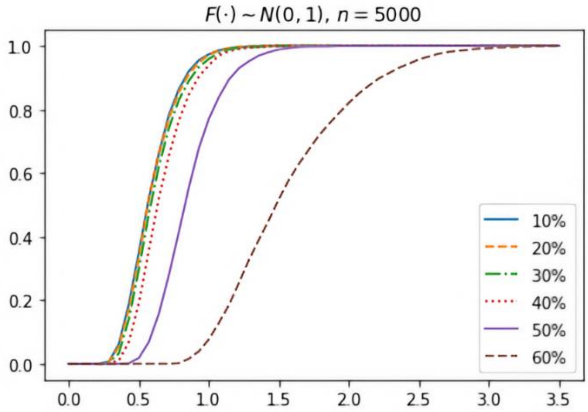
Из рисунков 8 и 9 видно, что функция распределения (15) становится абсолютно одинаковой, когда получим предельное распределение (15), тогда $F(t)$ будет равномерным, экспоненциальным и нормальным распределением. Судя по этим диаграммам, предельное распределение (15) не зависит от $F(t)$, а если и зависит, то практически не имеет значения. Особенно, это видно из (Рис. 8б), где размер выборки $n = 5000$.

Диаграммы показывают, что предельное распределение (15) не зависит от $F(t)$, а если и зависит, то практически не имеет значения. Особенно это видно из рисунок 9б, где размер выборки $n = 5000$.

Как отметили выше, все рисунки изображены без изменения степени цензурирования. Теперь, чтобы изучить зависимость предельного распределения от уровня цензурирования, построим предельное распределение для одного типа $F(t)$ (экспоненциального и нормального) с разными степенями цензурирования.



а) $F(t)$ – экспоненциальное распределение



б) $F(t)$ – нормальное распределение

Рисунок 9. Предельное распределение (15) в выборках с разной степеню цензурирования

Принимая во внимание, что на графиках могут быть ошибки масштаба, составим следующую таблицу с помощью программного обеспечения, чтобы прийти к окончательному выводу. В ней было построено предельное распределение в разных функциях распределения $F(t)$ и рассчитаны его значения при $t = 0,3;\ 0,5;\ 0,7;\ 1;\ 1,5$:

Таблица 6.

Значения предельного распределения в разных точках

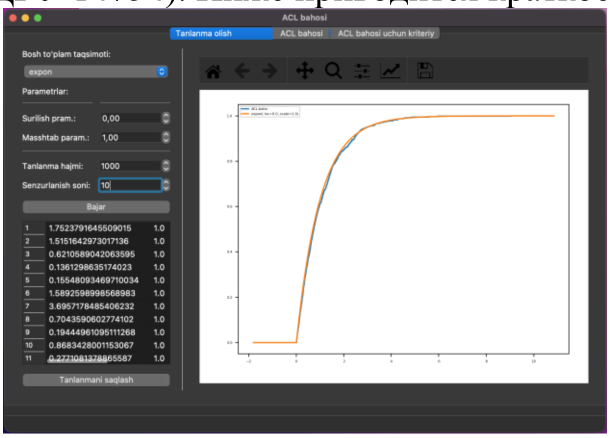
	0,3	0,5	0,7	1	1,5
N(1, 4)	0.0100	0.3455	0.7630	0.9749	0.9999
Exp(1)	0.0099	0.3449	0.7624	0.9747	0.9999
N(0, 1)	0.0100	0.3452	0.7621	0.9750	0.9999
Exp(1/4)	0.0100	0.3450	0.7623	0.9751	0.9999
N(3, 5)	0.0100	0.3454	0.7626	0.9751	0.9999
Exp(5)	0.0100	0.3450	0.7621	0.9747	0.9999
Exp(3)	0.0100	0.3456	0.7628	0.9749	0.9999

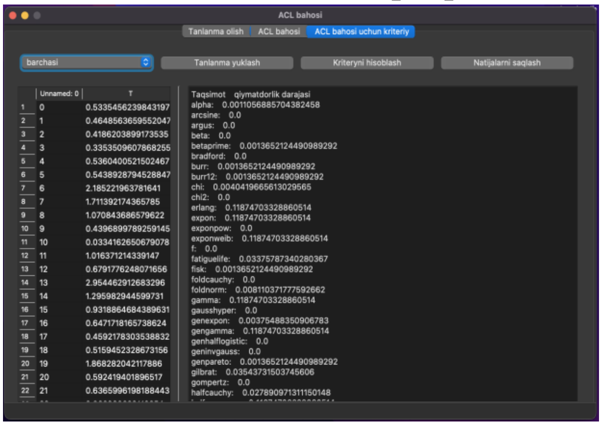
Сделаем окончательный вывод, предельное распределение (15) не зависит ни от параметров функции распределения $F(t)$, ни от самой себя. Но это будет сильно зависеть от степени цензурирования. Как только степень цензурирования достигнет 50% и более, будет видно резкое изменение распределения пределов. Однако тот факт, что предельное распределение сильно связано с цензурирированием, затрудняет нахождение предельного распределения. То есть необходимо ввести степень цензурирования как параметр предельного распределения. Более сложный состав оценки, чем эмпирическое распределение, не позволяет найти аналитический взгляд на предельное распределение.

С практической точки зрения такую ситуацию можно решить следующим образом:

1. Составить таблицы критических точек с очень высокой точностью с помощью программного продукта (для определенных степеней цензурирования).
2. В программном обеспечении найти критические точки с заданной точностью и проверить критерии.
3. Как видно из случайной величины $\sqrt{\dfrac{n}{\ln\ln n}}\cdot\sup\limits_{t\le T}\left|F(t)-F_n^{ACL}(t)\right|$, с увеличением степени цензурирования она сама увеличивается. Оценить скорость увеличения значения и найти функцию $\varphi(\theta)$ из представления $\varphi(\theta)\sqrt{\dfrac{n}{\ln\ln n}}\cdot\sup\limits_{t\le T}\left|F(t)-F_n^{ACL}(t)\right|$.

Все решения исследования представлены в данной работе. В частности, для решения второй ситуации получен программный продукт (№ ДГУ 11510, № ДГУ 14754). Ниже приводится краткое описание компонентов программы.





а) Раздел «Получение выборка» b) Раздел «Критерии оценки ACL»

Рисунок 10. Электронная программа анализа оценок правосторонней случайной цензурированной модели

Как видно из рисунка 10, программа состоит из трех окон. Первое окно – это создающее искусственную выборку для модели пропорциональной интенсивности с правосторонней случайной цензурирования. Этот раздел имеет 63 положительно определенных распределений. В следующем окне ("Оценка ACL") позволяет загрузить доступную выборку и выполнить на ее основе сравнения степенной оценки и оценки Каплана-Мейера и сохранить результаты. В следующем окне на основе введенной выборки в формате excel можно выбрать распределение генеральной совокупности и проверить критерии согласия (всего 63 типа распределения), а также просмотреть результаты критериев и сохранить результаты.

При работе с прикладными статистическими задачами приходится следить не только за цензурированными данными, но и за объектом под каким-то воздействием. Примерами таких факторов воздействия являются влажность, температура, давление, механическое напряжение в производственной сфере; в медицине можем указать количество лекарства, способ лечения, пол пациента, артериальное давление и возраст пациента. Такие факторы влияния называются **ковариатами**.

Обозначаем ковариаты $x=(x_1,x_2,\ldots,x_m)$ предполагая, что они находятся в возрастающем порядке $0\le x_1\le x_2\le\ldots\le x_m\le 1$ без ограничения общности. Очевидно, если n есть размер выборки, $m\le n$. В данной диссертационной работе ковариаты считаются независимыми от времени и $m=n$.

В компьютерном моделировании в качестве ковариат генерируют случайные величины, равномерно распределенные в сечении $[0,\ 1]$.

Полные выборки с ковариатами называются выборки, полученные следующим образом:

$$C_n=\{(T_1,x_1),(T_2,x_2),\ldots,(T_n,x_n)\},$$

где T_i – момент времени «отказа», x_i – ковариата.

Цензурированные справа данные с ковариатом определяются как**:**

$$C_n=\{(Z_1,\delta_1,x_1),(Z_2,\delta_2,x_2),\ldots,(Z_n,\delta_n,x_n)\},$$

где $Z_i=min(T_i,Y_i)$, T_i – момент времени «отказа», Y_i – момент времени цензурирования, $\delta_i=I(T_i\le Y_i)$ – индикатор события, а x_i – ковариата.

Следующей задачей диссертационной работы является исследование свойств следующей ковариационной оценки:

$$F_{xh}(t)=1-\left\{1-\left[H_{xh}(t)\right]^{\lambda_{xh}}\right\}^{\gamma_{xh}},t\ge 0. \tag{16}$$

где

$$\omega_{ni}(x;h_n)=\int_{x_{i-1}}^{x_i}\frac{1}{h_n}\pi\left(\frac{x-y}{h_n}\right)dy\Bigg/\int_{x_0}^{x_n}\frac{1}{h_n}\pi\left(\frac{x-y}{h_n}\right)dy, i=1,...,n,\ \sum_{i=1}^{n}\omega_{ni}(x;h_n)=1$$

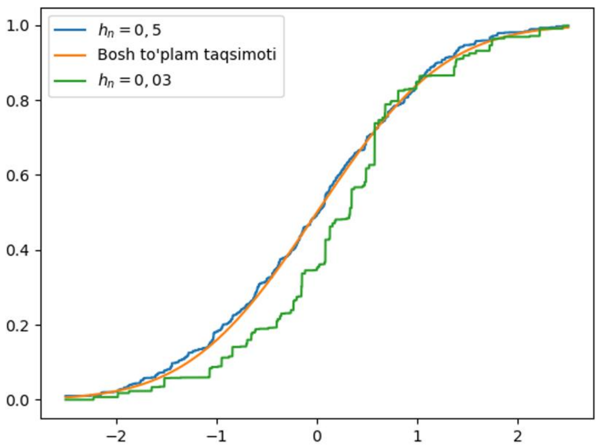
$$H_{xh}(t)=\sum_{i=1}^{n}I(\xi_i\le t)\omega_{ni}(x;h_n),\ \ P_{xh}^{(m)}=\sum_{i=1}^{n}\delta_i^{(m)}\omega_{ni}(x;h_n),m=0,1,2$$

$$\lambda_{xh}=1-p_{xh}^{(0)},\gamma_{xh}=p_{xh}^{(1)}\left(1-p_{xh}^{(0)}\right)^{-1}.$$

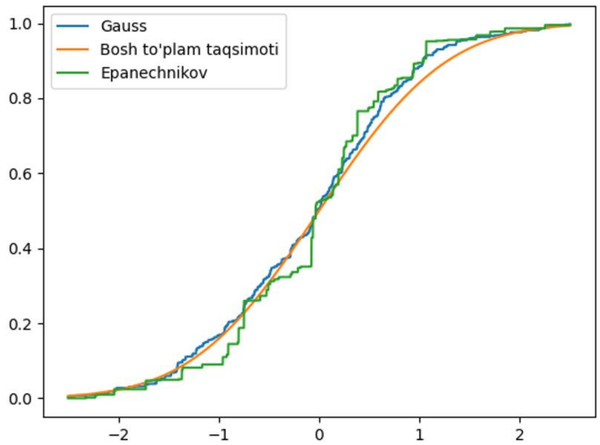
В частности, $\omega_{ni}(x;h_n)=\dfrac{1}{n}$ (или при отсутствии ковариаты) (16) становится простой эмпирической оценкой.

В данной диссертационной работе для вычисления весовой функции Гессера-Мюллера используются прямоугольная, треугольная, Епанечникова, квадратичная, трехвзвешенная, гауссовская, косинусная ядерные функции.

Теперь приступим к изучению свойств оценки. Сначала посторим оценку с разными ядра-функции и разной шириной окна.



a) В разной «ширине окна»



b) Различные ядра-функции

Рисунок 11. График ковариантной оценки с различной «шириной окна» и ядрами

Таблица 7.

Отклонение оценки от теоретического распределения

(модель Кокса $S(t)=\left(S_0(t)\right)^{r(x,\beta)}$, $S_0(t)$ – экспоненциальный, $r(x,\beta)=e^{\beta_0+\beta_1\log x}$, $\beta_0=1, \beta_1=2$, $n=1000$, цензурирования 10%**)**

Прямоугольная функция ядра							
h_n \ x	0,1	0,2	0,3	0,5	0,7	0,9	1
0,1	0,08368	0,08458	0,08422	0,08394	0,08334	0,08498	0,11643
0,3	0,06046	0,05330	0,05046	0,04932	0,04827	0,06094	0,06848
0,5	0,04897	0,04548	0,04189	0,03778	0,04285	0,04935	0,05378
0,7	0,04263	0,03994	0,03813	**0,03669**	0,03813	0,04194	0,04507
0,9	0,03803	**0,03643**	0,03774	0,03800	**0,03687**	0,03806	0,03982
1	**0,03647**	0,03820	**0,03669**	0,03829	0,03793	**0,03750**	0,03839
3	0,03783	0,03837	0,03804	0,03838	0,03855	0,03808	**0,03725**
5	0,03824	0,03845	0,03818	0,03793	0,03877	0,03882	0,03814

Внимательно изучив рисунок 11 и таблицу 7, можем заметить, что качество оценки сильно зависит от выбранного ядра-функции, ковариаты и выбранных параметров «ширины окна». Здесь выборки и ковариаты, которые получим при построении оценки, может не зависеть от нас. Но мы сами должны выбрать функцию ядра и параметры «ширины окна» при построении оценки. Итак, чтобы максимально улучшить оценку, пытаемся найти оптимальную функцию ядра и параметр «ширина окна». Внимательно наблюдая за каждым столбцом, можно сделать вывод, что существует оптимальная «ширина окна». В таблице они выделены жирным шрифтом. Присмотревшись внимательно к остальным столбцам, видим, что параметр «ширина окна» также сильно зависит от значения ковариаты.

Таким образом, после тестирования ковариационной оценки с различными ковариатами, разными ядрами и параметрами ширины окна, обнаружили, что во всех случаях оценка лучше всего, когда в качестве ядра функции использовалась ядра функция Гаусса. Теперь наша главная цель — подобрать оптимальный параметр ширины окна.

Существует несколько методов нахождения оптимального параметра ширины окна, наиболее важные из которых рассмотрены ниже:

– метод кросс-проверки, изучив все случаи и выбрав лучший из них;

– референт выбор;

– методы минимизации средней интегральной ошибки.

Теперь обсудим преимущества и недостатки каждого из вышеперечисленных методов. Первый способ можно найти параметр ширины окна с достаточной точностью. Однако этот метод требует огромных вычислительных усилий. Это может быть даже практически невозможно. Кроме того, в этом методе, если существует сильная ковариационная связь между продолжительностью жизни и ковариатами, то необходимо реконструировать метод с учетом этих связей. Учитывая вышеизложенное, данный метод не подходит для нахождения оптимального параметра ширины окна.

Второй референтный метод выбора приведен для оценки Берана:

$$h_{n,opt} = \frac{C}{\sqrt[5]{n}}.$$

Однако авторы не предоставили никакой информации о каком-либо алгоритме выбора параметра C. Сказано лишь, что параметр находится эмпирически. Этот метод не учитывает зависимость параметра ширины окна от ковариата.

Третий метод — средняя минимальная интегральная ошибка:

$$h_n = \hat{\sigma}\sqrt[5]{\frac{8\sqrt{\pi}R(K)}{2\mu_2(K)^2 n}} \tag{17}$$

где $\mu_2(K) = \int y^2 K(y)dy$; $R(K) = \int K^2(y)dy$ и $\hat{\sigma}$ – среднеквадратичное отклонение. Как видно в этом методе, вычисление функций $\mu_2(K)$ и $R(K)$ является очень сложной задачей.

В заключение отметим, что упомянутые выше методы не удовлетворяют нашим требованиям. Ниже предлагаем алгоритм нахождения оптимального параметра «ширина окна».

Для нахождения оптимального параметра ширины окна попытались минимизировать «расстояние» между обратной функцией распределения $F^{-1}(p)$ и моментом времени «отказа» Z_i. Из такого подхода ясно, что он подходит для любого значения ковариаты.

Итак, наша цель - найти параметр h – «ширина окна», который минимизирует значение суммы

$$\sum_{i=1}^{n} \delta_i \left| \hat{f}\left(F_{xh}\left(Z_i\right)\right) - Z_i \right| \qquad (18)$$

где $\hat{f}(\bullet) = \frac{1}{n}\sum_{i=1}^{n} \omega(\bullet) Z_i$ обратная функция функции ожидаемой продолжительности жизни, $F_{xh}(Z_i)$ – оценка функции распределения, $\omega(\bullet)$ – весовая функция, δ_i – индикатор цензурирования.

Здесь можем использовать различные весовые функции вместо $\omega(\bullet)$. В частности, в данной диссертации использовали весовую функцию Надарая-Ватсона (Nadaraya-Watson):

$$\omega_{ni}(x; a_n) = \frac{\sum_{i=1}^{n} Y_i \pi\left(\frac{x - x_i}{a_n}\right)}{\sum_{i=1}^{n} \pi\left(\frac{x - x_i}{a_n}\right)}.$$

При расчете приведенной выше весовой функции использовали формулу (17) в качестве параметра a_n – «ширина окна». Теперь, чтобы проверить, действительно ли найдена оптимальная «ширина окна», была построена модель оценки, и на основе этой модели были проведены следующие проверочные работы.

Таблица 8.

Оптимальный параметр «ширина окна» и отклонение оценки от теоретического распределения при использовании этого параметра

(модель Кокса $S(t) = \left(S_0(t)\right)^{r(x,\beta)}$, $S_0(t)$ – экспоненциальный, $r(x,\beta) = e^{\beta_0 + \beta_1 \log x}$, $\beta_0 = 1, \beta_1 = 2$, $n = 1000$, цензурирования 10%**)**

$K(\bullet)$ \ x		0,1	0,2	0,3	0,5	0,7	0,9	1
Прямоугольное ядро	$h_{n,opt.}$	0,9917	0,7404	2,9660	0,8954	0,7655	1,9410	2,9338
	D_n	0,0356	0,0360	0,0362	0,0363	0,0358	0,0365	0,0354

В приведенной таблице показаны оптимальный параметр «ширина окна» и «отклонение» оценки от функции распределения для каждой функции ядра, что в точности совпадает с результатами, представленными в таблице 7 выше. Итак, с помощью предложенного метода действительно смогли найти оптимальный параметр ширины окна.

ЗАКЛЮЧЕНИЕ

Основные результаты исследования, проведенного для диссертационной работы, написанной на соискание ученой степени доктора философии (PhD) по теме «Анализ оценок методами статистического моделирования в схемах неполных наблюдений», следующие:

1. имеющиеся оценки (Каплана-Мейера, степенная оценка, RR, PR) сравнивались в информативной модели со случайным цензурированием справа. Их достижения и недостатки были изучены в различных случаях. В информативной модели показаны преимущества степенной оценки по сравнению с оценкой Каплана-Мейера. Было обнаружено, что неинформативная модель степенной оценки имеет преимущества перед оценкой Каплана-Мейера, когда объем выборки мал и уровень цензурирования высок;

2. исследована степень зависимости оценки информативной модели с двусторонним случайным цензурированием от неизвестных параметров цензурирования. Для проведения научно-исследовательской работы разработан программный продукт искусственной выборки по информативной модели с двух сторон;

3. в информационной модели случайного цензурирования с двух сторон была разработана полупараметрическая оценка с экспоненциальным распределением генеральной совокупности, а также исследованы ее свойства. Показано, что эта оценка имеет низкую «чувствительность» к цензурированию. Также было рекомендовано использовать эту оценку, когда генеральная совокупность распределена экспоненциально (по сравнению с другими оценками);

4. изучены свойства критерия согласия Колмогорова для степенной оценки. На основании полученных результатов разработан критерий согласия и создан программный продукт для проверки критерия;

5. изучены свойства ковариатного оценивания функции распределения. Показано, что качество этой оценки сильно зависит от параметра «ширина окна», выбранной ядерной функции и ковариатами. Разработан алгоритм нахождения оптимального параметра «ширина окна», повышающий качество этой оценки.

SCIENTIFIC COUNCIL DSc.03/31.03.2022.T/FM.10.04 ON AWARD OF SCIENTIFIC DEGREE AT INSTITUTE OF FUNDAMENTAL AND APPLIED RESEARCH "TIIAME" NATIONAL RESEARCH UNIVERSITY

NAVOI STATE PEDAGOGICAL INSTITUTE

MANSUROV DILSHOD RAVILOVICH

ANALYSIS OF ESTIMATES USING STATISTICAL MODELING METHODS IN SCHEMES WITH INCOMPLETE OBSERVATIONS

05.01.07-Mathematical modeling. A set of numerical methods and programs

**ABSTRACT OF THESIS OF THE DOCTOR OF PHILOSOPHY (PHD)
ON PHYSICAL AND MATHEMATICAL SCIENCES**

NAVOI-2023

The theme of dissertation of doctor of philosophy (PhD) on physical and mathematical sciences was registered at the Supreme Attestation Commission at the Cabinet of Ministers of the Republic of Uzbekistan under number B2022.4.PhD/FM825.

Dissertation has been prepared at Navoi State Pedagogical Institute.

The abstract of the dissertation is posted in three languages (Uzbek, Russian, English (summary)) on the website https://www.ifar.uz and in the website of "ZiyoNet" Information and educational portal http://www.ziyonet.uz/.

Scientific supervisor:	**Abdushukurov Abdurakhim Akhmedovich** Doctor of Physical and Mathematical Sciences, Professor
Official opponents:	**Eshmatov Farkhod Khasanovich** Doctor of Physical and Mathematical Sciences, Professor
	Nurmukhamedova Nargiza Saydillayevna PhD in Physical and Mathematical Sciences, Docent
Leading organization:	**Tashkent State transport University**

The defense of the dissertation will be held on "___" __________ 2022 at ___ in the meeting of the Scientific Council No. DSc.03/31.03.2022.T/FM.10.04 at the Institute of Fundamental and Applied Research under the National Research University "TIIAME". (Address: 100000, Tashkent city, Qori Niyazov Street 39, Institute of Fundamental and Applied Research, Hall 108; tel.: 71 237-09-61.; e-mail: info@ifar.uz)

The doctoral (PhD) dissertation can be looked through at the Information Resource Center of the Institute of Fundamental and Applied Research under the National Research University "TIIAME" (registered under № ___). (Address: 100000, Tashkent city, 39 Qori Niyazov str., Institute of Fundamental and Applied Research, hall 108; ph.: 71 237-09-61).

The Abstract of the dissertation was distributed on « ___ » ___ 2023 year
(Registry record № ___ dated "___"_________, 2022)

B. J. Ahmedov
Chairman of Scientific Council
on award of scientific degrees,
D.Ph.-M.S., professor

E.Kh. Karimbaev
Scientific Secretary of Scientific
Council on Award of Scientific
Degrees PhD

A. R. Hayotov
Chairman of Scientific seminar under
Scientific Council on award of scientific degrees,
D.Ph.-M.S., professor

INTRODUCTION (Annotation of PhD dissertation)

The aim of the research is to study and compare the analogy of power structure estimate with censoring from both sides with existing estimations, as well as to develop and study the goodness of fit criteria. To study the existence of the optimal parameter “bandwidth” for the covariate generalization of power structure estimator, what parameters it depends on, and to develop an adaptive algorithm for finding it.

Object of the research. The estimator power-type of the informative model of right random censorship and its natural generalizations. Kaplan-Meier estimation and its generalizations.

The scientific novelty of the research is as follows:

firstly, the exponential distribution of the population has been parametrically estimated in the informative model of random censorship from both sides;

firstly, for the first time in Kolmogorov's statistics, when the empirical distribution function was replaced by a power-type estimate using numerical methods, it was shown that the limiting distribution of statistics does not depend on the distribution function and its parameters, and if it does, it practically does not matter;

the goodness of fit criteria based on the power-type estimator is proposed;

the dependence of statistical characteristics of estimator with covariate of the distribution function on the type of kernel functions, sample size, experimental design, number of control points, and regression is shown and studied;

the adaptive algorithm is proposed for determining the optimal value of the “bandwidth” parameter, which allows constructing a high-quality semi-parametric estimate based on the results of experimental research.

Implementation of research results. The results of the comparison of estimates in informative and non-informative models, obtained in the dissertation, and developed goodness of fit criteria for assessing the proportionality of intensities were used in the fundamental project No. F4-40 “Study of asymptotic properties of indexed integral empirical processes in the class of dimensional functions” (Certificate No. 04/11-566 of February 6, 2023, of the National University of Uzbekistan named after Mirzo Ulugbek). The application of scientific results made it possible to compare the estimate of the private model of proportional intensities of Koziol-Green with the Kaplan-Meier estimate;

a software product that forms an artificial sample for informative and non-informative models and checks with their help the goodness of fit criteria, as well as the covariance representation of the empirical process when applying estimates of models with incomplete observation, were used in the fundamental youth project Yo-F4-07 “Statistical estimation and hypothesis testing using copula functions” (Certificate No. 04/11-2875 of May 13, 2023, of the National University of Uzbekistan named after Mirzo Ulugbek). The application of scientific results made it possible to test the criteria for a family of distribution functions.

Volume and structure of the dissertation. This dissertation consists of an introduction, 3 chapters, and a conclusion. The total volume of the dissertation is 122 pages, including 46 figures and 21 tables. The bibliography contains 110 references.

E'LON QILINGAN ISHLAR RO'YXATI
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I bo'lim (часть 1; part 1)

1. Abdushukurov A.A., Mansurov D.R. Estimating the Distribution Function Using Parametric Methods in Informative Model of Random Censorship from Both Sides // Communications in Computer and Information Science / editors A.Dudin, A.Nazarov, A.Moiseyev, Cham: Springer Nature Switzerland, 2023. Vol. 1803. – P. 225-237. DOI https://doi.org/10.1007/978-3-031-32990-6_19 (3. Scopus. IF=0,209).
2. Abdushukurov A.A., Bozorov S.B., Mansurov D.R. Estimation of Distribution Function Based on Presmoothed Relative-Risk Function // Applied Mathematics, 2022. – № 02 (13). – P. 191-204. DOI https://doi.org/10.4236/am.2022.132015 (35.CrossRef. IF=0.58).
3. Abdushukurov A.A., Mansurov D.R. Asymptotic Results For Empirical Processes In Informative Model Of Random Censorship From Both Sides // Bulletin of National University of Uzbekistan: Mathematics and Natural Sciences. – Tashkent, 2021. – № 1 (4). – P. 78-88. URL https://uzjournals.edu.uz/mns_nuu/vol4/iss1/7 (01.00.00. № 8).
4. Абдушукуров А.А., Мансуров Д.Р. Критерий согласия для проверки модели пропорциональности интенсивностей // Проблемы вычислительной и прикладной математики, 2022. – № 5 (43). – С. 166-178. URL https://elibrary.ru/edjmpv (01.00.00. № 9).
5. Абдушукуров А.А., Мансуров Д.Р. Асимптотические результаты для последовательных эмпирических процессов в информативных моделях неполных наблюдений // "Ilm sarchashmalari" ilmiy-nazariy, metodik jurnal. – Urganch, 2020. – № 11. – B. 3-7 (01.00.00. № 12).

II bo'lim (часть 2; part 2)

6. Мансуров Д.Р. Последовательные эмпирические процессы в информативных моделях неполных наблюдений / Материалы Международной научно-практической онлайн-конференции "Теории функций одного и многих комплексных переменных". – Нукус, 2020. 26-28 ноября. – С. 165-168.
7. Abdushukurov A.A., Mansurov D.R. Parametric estimation of exponential distribution in informative model of random censorship from both sides / XXXVI International Seminar on Stability Problems for Stochastic Models Petrozavodsk. – Karelia, Russia, 2021. June 21-25. – P. 193-194.
8. Mansurov D.R. Studies estimates of the distribution function in informative modelof random censorship from both sides / Республиканской научной конференции с участием зарубежных ученых "Сарымсаковские чтения". – Ташкент, Узбекистан, 2021. 16-18 сентября. – С. 242-243.
9. Abdushukurov A.A., Mansurov D.R. On estimates of the distribution function in informative model of random censorship from both sides / Informational

technologies and mathematical modelling (ITMM-2021). 01-05 декабря 2021 года. Национальный исследовательский Томский государственный университет. – Томск, 2022. – С. 274-279.

10. Abdushukurov A.A., Mansurov D.R. A Study of The Kaplan–Meier Estimator and Its Dependence on The Degree of Censorship / Collection Materials of The Republican Scientific and Practical Conference “Theoretical Foundations and Applied Problems of Modern Mathematics II”. – Andijan, Uzbekistan, 2022. March 28. – P. 163-166.

11. Mansurov D.R., Bozorov S.B. On Survival Function Estimation in Partially Dependent Informative Random Censorship / Abstracts of The Conference of “Young Scientists Mathematics, Mechanics and Intellectual Technologies”. – Tashkent, 2022. 21-22 April. – P. 17-18.

12. Mansurov D.R., Bozorov S.B. Полупараметрическое оценивание условной функции выживания в информативной регрессионной модели случайного цензурирования с двух сторон / “Matematika va informatika fanlarini oʻqitishning muammolari va ularning yechimlari” mavzusidagi Respublika miqyosidagi ilmiy-amaliy anjuman materiallari. – Nukus, 2022. 21-22-aprel. – B. 110-114.

13. Mansurov D.R. Proporsional intensivlik modeli bahosi / “Yosh matematiklarning yangi teoremalari – 2022” nomli ilmiy anjumani tezislari toʻplami. – Namangan, 2022. 13-14-may. – B. 237-239.

14. Mansurov D.R. Ikki tomonlama tasodifiy senzurali informativ model baholarini modellashtiruvchi elektron dastur. Elektron hisoblash mashinalari uchun yaratilgan dasturning rasmiy roʻyxatdan oʻtkazilganligi toʻgʻrisidagi guvohnoma (№ DGU 11510, 18.06.2021-y.).

15. Abdushukurov A.A., Mansurov D.R. Toʻliq boʻlmagan kuzatilmali sxemalarda statistik modellashtirish metodlari yordamida baholarni tahlil qiluvchi elektron dastur. Elektron hisoblash mashinalari uchun yaratilgan dasturning rasmiy roʻyxatdan oʻtkazilganligi toʻgʻrisidagi guvohnoma (№ DGU 14754, 25.02.2022-y.).

16. Mansurov D.R. Toʻliq boʻlmagan tanlanmalar yordamida taqsimot funksiyalarini baholovchi va ushbu baholar yordamida muvofiqlik mezonlarini tekshiruvchi dastur. Elektron hisoblash mashinalari uchun yaratilgan dasturning rasmiy roʻyxatdan oʻtkazilganligi toʻgʻrisidagi guvohnoma (№ DGU 24964, 24.05.2023-y.).

Avtoreferat «O‘zMU xabarlari» jurnali tahririyatida tahrirdan o‘tkazilib, o‘zbek, rus va ingliz tillaridagi matnlar o‘zaro muvofiqlashtirildi.

Bosmaxona litsenziyasi:

9338

Bichimi: 84x60 $^{1}/_{16}$. «Times New Roman» garniturasi.
Raqamli bosma usulda bosildi.
Shartli bosma tabog‘i: 3,75. Adadi 100 dona. Buyurtma № 43/23.

Guvohnoma № 851684.
«Tipograff» MCHJ bosmaxonasida chop etilgan.
Bosmaxona manzili: 100011, Toshkent sh., Beruniy ko‘chasi, 83-uy.