Київський національний університет імені Тараса Шевченка

МЕТОДИЧНІ ВКАЗІВКИ
до практичних робіт з дисципліни
“Моделі виживання”

для студентів механіко-математичного факультету

Видавничо-поліграфічний центр
'Київський університет'
2019

Рецензенти:
д-р фiз.-мат.наук, проф. Г.М.Шевченко,
д-р фiз.-мат.наук, проф. Є.О.Лєбєдєв

Рекомендовано до друку вченою радою механiко-математичного факультету
протокол N2 вiд 15.10.2019 року

МЕТОДИЧНI ВКАЗIВКИ до самостiйних робiт з дисциплiни “Моделi виживання ймовiрностей” / Упорядник: В.В. Голомозий- К., Видавничо-полiграфiчний центр 'Київський унiверситет', 2019 - 44 с.

Навчальне видання

МЕТОДИЧНI ВКАЗIВКИ
до самостiйних робiт з дисциплiни
“Моделi виживання ”

для студентiв механiко-математичного факультету

Упорядники

Голомозий Вiталiй Вiкторович

Зміст

Виконання завдань

Методичні вказівки до самостійних робіт з спеціального курсу "Моделі виживання " призначені для студентів I курсу магістратури механіко-математичного факультету спеціальностей "Математика", "Статистика", Актуарна та фінансова математика.

Зміст та структура матеріалу відповідає програмі курсу та вимогам кредитно - модульної системи організації навчального процесу.

Розділи посібника містять завдання самостійних робіт, що входять до складу змістовного модулю курсу: I. Моделі виживання

Завдання робіт з кожної теми розбито на три групи:
A – задачі для аудиторної роботи,
B – задачі самостійних робіт,
задачі.

Аудиторні завдання з розділу **A** виконуються під час практичних занять. Позааудиторні самостійні завдання з розділів **B** виконуються студентами самостійно на підставі прослуханих лекцій, отриманих у аудиторії, задач розв'язаних в аудиторії, та поточних консультацій з теоретичного матеріалу для самостійних робіт.

Якість виконання вказаних завдань контролюється викладачами, оцінюється у балах, та враховується згідно кредитно-модульній системі при оцінюванні рівня знань студентів.

Змістовно матеріал посібника розбитий на теми відповідно до модуля.

1. Марківські моделі

1. Ланцюги Маркова.

Призначення: Оволодіння поняттями Марківської властивості та ланцюга Маркова. Навчитися класифікувати ланцюги Маркова (за періодичністі, незвідністю). Навчитися обчислювати стаціонарі ймовірінсі розподіли ланцюгів Маркова, та граничні ймовірності переходу.

2. Стрибкоподібні процеси Маркова.

Призначення: Оволодіти поняттями стрибкоподібних ланцюгів Маркова. Навчитися записувати і розв'язувати диференціальні рівняння Колмогорова, досліджувати ймовірності знаходження ланцюга в певному стані та ймовірності стрибків.

2. Моделі виживання

3. Базова актуарна модель

Призначення: Оволодіти основними поняттями та позначеннями прийнятими в актуарній математиці. Ознайомитися з найпростішими моделями тривалості життя.

4. Оцінювання розподілу тривалості життя.

Призначення: Ознайомитися з оцінками Каплана-Мейєра та Нельсона-Аалена. Навчитися обчислювати ці оцінки.

5. Регресійна модель Кокса.

Призначення: Ознайомитися з регресійною моделлю Кокса та поняттям пропорційних ризиків.

Література

1. *Карташов М.В.* Імовірність, процеси, статистика - Київ, ВПЦ Київський університет, 2007.

2. *Скороход А.В.* Елементи теорії ймовірностей та випадкові процеси - Київ, Вища школа, 1975.

3. *Bowers N., Gerber H., Hickman J., Jones D., Nesbitt C.* Actuarial mathematics, 2nd edition, 1997.

1. Марківські моделі

1. Ланцюги Маркова

Література: [1,с.162-169]

Теоретичні відомості

1.1.Рівняння Чепмена-Колмогорова.

Ми визначимо *Ланцюг Маркова* – як процес Маркова у дискретному часі зі скінченною або зліченною множиною станів S. Це означає, що ланцюг Маркова – це послідовність випадкових величин $X_0, X_1, ..., X_n, ...$ із властивістю

$$P(X_n = j|X_0 = i_0, X_1 = i_1, ..., X_{m-1} = i_{i-1}, X_m = i) = P(X_n = j|X_m = i) \tag{1.1}$$

для всіх $n > m$ і $i_0, i_1, ..., i_{m-1}, i, j \in S$. Властивість (1.1) називається Марківською і означає, що знаючи теперішній стан процесу $X_m = i$, для обчислення розподілу ймовірностей майбутніх станів процесу не потрібне знання минулого. Ймовірність $P(X_n = j|X_m = i) = p_{ij}^{(m,n)}, m < n$ називають ймовірністю переходу.

Теорема 1.1. Ймовірність переходу задовольняє рівняння Чепмена-Колмогорова

$$p_{ij}^{(m,n)} = \sum_{k\in S} p_{ik}^{(m,l)} p_{kj}^{(l,n)}, \qquad \forall i, j \in S, \qquad \forall m < l < n.$$

1.2. Однорідні ланцюги Маркова.

Розглянемо ланцюг Маркова, для якого ймовірність переходу за один крок не залежить від часу:

$$p_{ij}^{(n,n+l)} = p_{ik}. \tag{1.2}$$

У цьому випадку ми говоримо, що ланцюг однорідний за часом. Із (1.2) легко вивести, що загальна ймовірність переходу залежить лише від різниці між моментами часу

$$P(X_{l+m} = j|X_m = i) = p_{ij}^{(l)}. \tag{1.3}$$

Ймовірність із (1.3) будемо назвати ймовірністю переходу за l кроків. Для однорідних за часом ланцюгів Маркова рівняння Чепмена-Колмогорова має вигляд

$$p_{ij}^{(n-m)} = \sum_{k\in S} p_{ik}^{(l-m)} p_{kj}^{(n-l)}, \qquad m < l < n.$$

Якщо ввести *матрицю переходів* $P = \{p_{ij}\}_{i,j \in S}$, тоді за індукцією із рівняння Чепмена-Колмогорова одержимо $p_{ij}^{(l)} = (P^l)_{ij}$, де P^l–це l-та степінь матриці P. Матриця переходів P– це $N \times M$ матриця, де N це число станів в S (можливо нескінченне). Для всіх $i \in S$ маємо умову $\sum_{j \in S} p_{ik} = 1$, отже сума елементів кожного рядка матриці переходів дорівнює одиниці. За індукцією одержимо співвідношення $\sum_{j \in S} p_{ij}^{(l)} = 1, \forall i \in S$.

Часто потрібно намалювати *граф переходів* ланцюга Маркова. Це діаграма, у якій кожен стан із S зображається вершиною графа і стрілкою з'єднують вершину i із вершиною j, якщо $p_{ij} > 0$, що означає можливість переходу зі стану i в стан j за один крок. Ймовірність p_{ij} зображається над відповідною стрілкою.

1.3. Приклад.

Проста модель системи знижок за відсутності позовів (No Claim Discunt Policy, NCD)

Система знижок за відсутності позовів у страхуванні автомобілів полягає у тому, що величина страхових премій залежить від історії позовів клієнтів. Розглянемо дві прості моделі на основі ланцюгів Маркова, а також різні можливості покращення цих моделей.

Компанія по стархуванню авто надає клієнтам такі знижки: 0% знижки (стан 0), 25% знижки (стан 1), 50% знижки (стан 2). Якщо протягом року позовів не було, то клієнт переходить до наступного вищого стану у наступному році (або залишается у стані з найвищим боунсом). Аналогічно, якщо протягом року були позови, то наступного року клієнт переходить до сусіднього нижчого стану (або залишается без знижки). За цих умова система знижок для клієнта утворює ланцюг Маркова із простором станів $S = \{0, 1, 2\}$, і якщо ймовірність не мати позовів протягом року дорівнює 3/4, то матриця переходів має вигляд

$$P = \begin{pmatrix} 1/4 & 3/4 & 0 \\ 1/4 & 0 & 3/4 \\ 0 & 1/4 & 3/4 \end{pmatrix}$$

Ймовірність одержати максимальну знижку у $n+3$-ому році за умови, що у n-ому році знижок не було, дорівнює $p_{02}^{(3)} = (P^3)_{02} = 9/16$.

У випадку неоднорідності за часом цієї моделі, ймовірність аварії може залежати від часу, відображаючи зміну умов руху з часом. У цьому випадку матриця переходів матиме вигляд:

$$P = \begin{pmatrix} P_{00}(t) & P_{01}(t) & P_{02}(t) \\ P_{10}(t) & P_{11}(t) & P_{12}(t) \\ P_{20}(t) & P_{21}(t) & P_{22}(t) \end{pmatrix}$$

1.4. Стаціонарні розподіли.

Будемо говорити, що $\pi_j, j \in S$– це *стаціонарний (або інваріантний) розподіл ланцюга Маркова* з матрицею переходу $P = \{p_{ij}\}$, якщо для $\forall j \in S$:

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}, \tag{1.4}$$

$$\pi_j \geqslant 0, \sum_{j \in S} \pi_j = 1.$$

Зауважимо, що (1.4) можна записати у вигляді $\pi = \pi P$, де $\pi = (\pi_1, ..., \pi_n, ...)$.

Якщо взяти розподіл π у якості початкового розподілу $P(X_0 = i) = \pi_i$, тоді

$$P(X_1 = j) = \sum_{i \in S} P(X_1 = j | X_0 = i) P(X_0 = i) = \sum_{i \in S} p_{ij} \pi_i = \pi_j, \forall j \in S.$$

За індукцією будемо мати $P(X_n = j) = \pi_j, \forall j \in S, \forall n$. Тому ланцюг з початковим розподілом π буде стаціонарним процесом.

У загальному випадку ланцюг Маркова не обов'язково має стаціонарний розподіл, а якщо і має, то не обов'язково єдиний.

Теорема 1.4.1. Якщо ланцюг Маркова має скінченний простір станів, то існує принаймні один стаціонарний розподіл.

Приклад

Знайдемо стаціонарний розподіл для NCD моделі з пункту 1.3. Із (1.4) маємо

$$\begin{cases} \pi_0 = 1/4\pi_0 + 1/4\pi_1, \\ \pi_1 = 3/4\pi_0 + 1/4\pi_2, \\ \pi_2 = 3/4\pi_1 + 3/4\pi_2. \end{cases}$$

Дана система є виродженою, і має лише два лінійно незалежних рівняння. Але стаціонарний розподіл, є ймовірнісним, тому до системи слід додати рівняння: $\pi_1 + \pi_2 + \pi_3 = 1$.

Розв'язавши систему, отримаємо що Стаціонарний розподіл даного ланцюга Маркова має вигляд $\pi = (1/13, 3/13, 9/13)$.

Питання єдиності стаціонарного розподілу більш складне і ми розглянемо його для *незвідних* ланцюгів Маркова, тобто таких, для яких у будь-який стан $j \in S$ можна потрапити із будь-якого іншого стану $i \in S$. Таким чином ланцюг Маркова буде незвідним, якщо для $\forall i, j \in S$ $\exists n$ таке, що $p_{ij}^{(n)} > 0$. Цю властивість можна виявити із графа переходів. Поглинаючий стан часто зустрічається у реальних системах, наприклад у задачах про банкрутство.

Теорема 1.4.2. Незвідний ланцюг Маркова зі скінченим простором станів має єдиний стаціонарний розподіл ймовірностей.

A

1. Учасники програми медичного страхування поділяються на дві групи – вкладники та бенефіціари. Якщо вкладник серйозно захворів, він перетворюється у бенефіціара. Ймовірність такої події рівна 0.1. Ймовірність того, що хвороба буде продовжуватися у наступний період складає 0.2. За правилами програми, будь-який учасник, що є бенефіціаром протягом трьох періодів часу поспіль, обов'язково має стати вкладником у наступному періоді. Якщо до цього моменту він все ще хворіє, він все одно повинен стати вкладником на наступний період.

а) Для даної медичної схеми побудувати Ланцюг Маркова з дискретним часом, ввівши за потреби різні класи вкладників та бенефіціарів (пропонується представити 5 класів)
Зобразити ланцюг переходів
Записати матрицю переходів
б) Чи є даний ланцюг Маркова незвідним, періодичним?
в) Знайти стаціонарний розподіл, якщо він існує
г) Визначити частку бенефіціарів серед всіх учасників медичного страхування

2. Задано однорідний ланцюг Маркова з простором станів S={1,2,3} та матрицею переходів:

$$P = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 3/4 & 1/4 & 0 \end{bmatrix} \tag{1.5}$$

а) Обчислити матрицю переходу за 3 кроки. Для кожного початкового стану обчислити ймовірність того, що ланцюг буде у стані 3 у період часу $n = 3$, якщо:
а.1) Ланцюг у стані 1 у період часу 0
а.2) Ланцюг у стані 1 у період часу 0 та у стані 2 у період часу 1
а.3) Ймовірності знаходитися у стані 1, 2, 3 у період часу 0 є 14/31, 9/31 та 8/31 відповідно
б) Як зміниться відповідь до запитань a), b), c) , якщо $n = 300$?

3. Деяка компанія оцінює кредитоспроможність різних фірм кожного кварталу. Рейтингова шкала у порядку спадання якості від A до D. Історичні дані підтверджують, що кредитний рейтинг типової фірми є ланцюгом Маркова з наступною матрицею переходів для деякого параметра μ:

$$P = \begin{bmatrix} 1-\mu-\mu^2 & \mu & \mu^2 & 0 \\ \mu & 1-2\mu-\mu^2 & \mu & \mu^2 \\ \mu^2 & \mu & 1-2\mu-\mu^2 & \mu \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1.6}$$

а) Намалювати графік переходу станів із одного в інший
б) Визначити межі, у яких лежить параметр μ
в) Чи є даний ланцюг Маркова незвідним, періодичним?
г) Знайти стаціонарний розподіл, якщо він існує, і визначити чи є він єдиним
д) Для значення $\mu = 1$, обрахувати ймовірність того, що компанія, яка отримала певний рейтинг (див. пункти д.1-д.4) у третьому кварталі, знаходиться у стані дефолту (рейтинг D):
д.1) У випадку, коли рейтинг компанії у першому кварталі рівний A
д.2) У випадку, коли рейтинг компанії у першому кварталі рівний B
д.3) У випадку, коли рейтинг компанії у першому кварталі рівний C
д.4) У випадку, коли рейтинг компанії у першому кварталі рівний D

4. Компанія займається страхуванням двигуна автомобіля і має п'ятирівневу систему дисконтів. На різних рівнях застраховані сплачують наступну частку вартості поліса: Рівень 1 - 60%,Рівень 2 - 70%,Рівень 3 - 80%,Рівень 4 - 90%,Рівень 5 - 100%. Застрахований рухається у межах від першого до п'ятого рівнів у залежності від страхових позовів у попередньому році. Для кожного власника страхового полісу, кількість позовів за рік має розподіл Пуассона з середнім 0.25. Для тих, хто на початку попереднього року мав рейтинг 2,3,4,5:

- Якщо протягом попереднього року не було жодного позову, застрахований рухається вниз на один рівень (наприклад з 4 до 3)
- Якщо відбувся один позов протягом попереднього року, застрахований рухається вгору на один рівень (крім випадка, якщо він вже має рівень 5)

Для тих, хто мав рівень 1 на початку перед попереднього року, застосовується акція захисту від втрати одного рівня, якщо протягом попереднього року був випадок лише одного позову. Якщо було зроблено 2 позови – застрахований рухається на один рівень вгору (від 1 до 2). Якщо було зроблено 3 і більше позовів – застрахований отримує рівень 5.

а) Визначити матрицю переходів для системи отримання дисконту
б) Застрахований має рівень 3 у перший рік страхування. Порахувати ймовірність того, що на початку третього року дії полісу, застрахований матиме рівень 1 (рівень 2)
в) Визначити умови за яких ймовірність знаходитися у деякому стані після n років збігається до якогось рівня, який не залежить від початкового стану
Перевірити коректність умов у даному випадку
г) Визначити граничну ймовірність того, що застрахований буде мати рівень 1.

5. Учасники процесу страхування від інвалідності є активними (A), тимчасово фізично неспроможними (T), постійно фізично неспроможними (P) або мертвими (D). Ті, хто знаходиться у стані T або P мають право на отри-

мання дивідендів від страхування. З метою аналізу стан кожного учасника визначається 1 квітня кожного року. Було визначено, що типовий учасник даного страхування змінює свій стан відповідно до ланцюга Маркова з дискретним часом та матрицею переходу:

$$A = \begin{bmatrix} 0.75 & 0.1 & 0.05 & 0.1 \\ 0.5 & 0.3 & 0.1 & 0.1 \\ 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1.7}$$

а) Намалювати графік переходів зі стану в стан та стаціонарний розподіл ймовірностей
б) Обчислити середню тривалість знаходження у стані фізично неспроможного (у роках)
в) Обчислити ймовірність, що учасник, який спочатку був повністю здоровий через три роки буде знаходитися у стані P або T
г) Обчислити ймовірність того, що здоровий учасник ніколи не отримає страхової премії за даної схеми страхування

6. Страхувальник автомобільних двигунів має наступну схему дисконтів за безаварійне водіння:
Рівень 1 - Знижка 0%,Рівень 1 - Знижка 25%,Рівень 1 - Знижка 40%,Рівень 1 - Знижка 50%.
Правила переходу від одного рівня до іншого наступні:
- Якщо застрахований протягом року жодного разу не звернувся до страхової компанії, він рухається вверх на один рівень або залишається на 4-ому
- У випадку 1-го чи більше позовів протягом року застрахований отримує рейтинг на один менше, якщо у рік перед цим, жодного позову не було. Отримує рейтинг на два менше від того що був у іншому випадку
Для деякого застрахованого імовірність жодного позову протягом року рівна 0.8.

а) Нехай $X(t)$ – деякий рівень застрахованого у році t. Пояснити чому $\{X(t)\}_{t=1}^{\infty}$ не є ланцюгом Маркова.
б) Збільшуючи кількість станів визначити новий стохастичний процес $\{Y(t)\}_{t=1}^{\infty}$, який буде Марковським та такий що $Y(t)$ описує рівень застрахованого у році t
в) Написати матрицю переходів для ланцюга Маркова $\{Y(t)\}_{t=1}^{\infty}$ г) Обчислити довгострокову імовірність того, що застрахований має дисконт рівня 3.

2. Стрибкоподібні процеси Маркова

Література: [1,с.170-174]

Теоретичні відомості

2.1.Рівняння Чепмена-Колмогорова.

Процес Маркова $X_t, t \geqslant 0$ у неперервному часі та з дискретним (скінченним або зліченним) простором станів S називається *стрибкоподібним процесом Маркова.*

У цьому підрозділі будемо розглядати однорідні за часом процеси, для яких ймовірність переходу $P(X_t = j|X_s = i)$ залежить лише від різниці $t-s$. Ймовірність переходу однорідного стрибкоподібного процесу Маркова $p_{ij}(t) = P(X_t = j|X_0 = i)$ задовольняє рівняння Чепмена-Колмогорова

$$p_{ij}(t+s) = \sum_{k\in S} p_{ik}(s)p_{kj}(t), \qquad \forall s, t > 0. \tag{2.1}$$

2.2.Матриця переходів.

Позначивши через $P(t) = p_{ij}(t)$ матрицю переходів, рівняння (2.1) можна переписати у вигляді $P(t+s) = P(s)P(t), \forall s, t > 0$. Якщо ми знаємо матрицю переходів $P(t)$ і початковий розподіл $q_i = P(X_0 = i)$, то ми можемо визначити загальні ймовірності для процесу X_t, використовуючи Марківську властивість. Наприклад, нехай $0 < t_1 < t_2 < ... < t_n$, тоді

$$P(X_0 = i, X_{t_1} = j_1, X_{t_2} = j_2, ..., X_{t_n} = j_n,) = q_i p_{ij_1}(t_1)p_{j_1j_2}(t_2-t_1)...p_{j_{n-1}j_n}(t_n-t_{n-1}),$$

де $p_{j_kj_{k+1}}(t_{k+1} - t_k) = P(X_{t_{k+1}} = j_{k+1}|X_{t_k} = j_k)$, і

$$P(X_{t_1} = j_1, X_{t_2} = j_2, ..., X_{t_n} = j_n,) = \sum_{i\in S} q_i p_{ij_1}(t_1)p_{j_1j_2}(t_2-t_1)...p_{j_{n-1}j_n}(t_n-t_{n-1}).$$

Ми будемо вважати, що функції $p_{ij}(t)$ неперервно диференційовні.

2.3.Диференціальні рівняння Колмогорова.

Зауваживши, що

$$p_{ij}(0) = \delta_{ij} = \begin{cases} 0, & \text{якщо } i \neq , \\ 1, & \text{якщо} i = j. \end{cases} \tag{2.2}$$

із умови диференційовності одержимо існування величин

$$\sigma_{ij} = \frac{d}{dt}p_{ij}(t)\Big|_{t=0} = \lim_{h\to 0}\frac{p_{ij}(h) - p_{ij}(0)}{h}.$$

Еквівалентно, наступне співвідношення виконується при $h \to 0, (h > 0)$:

$$p_{ij} = \begin{cases} \sigma_{ij}h + o(h) & \text{якщо } i \neq j, \\ 1 + \sigma_{ii}h + o(h) & \text{якщо } i = j. \end{cases} \tag{2.3}$$

Із першого рядка (2.3) випливає, що ймовірність переходу із стану i у стан j протягом довільного малого інтервалу часу $[s, s+h]$ пропорційна h, тому величина σ_{ij} називається *інтенсивністю переходів.* Інтенсивність переходів повністю характеризує розподіл стрибкоподібного процесу Маркова. Дійстно, із рівняння (2.1) підставивши $t = h$ і $s = t$, одержимо

$$p_{ij}(t+h) = \sum_{k\in S} p_{ik}(t)p_{kj}(h) = \sum_{k\neq j} p_{ik}(t)\sigma_{kj}h + o(h) + p_{ij}(t)(1+\sigma_{jj}h+o(h)) =$$

$$= p_{ij}(t) + h\sum_{k\in S} p_{ik}(t)\sigma_{kj} + o(h).$$

Звідси при $h \longrightarrow 0$ одержимо диференціальне рівняння

$$\frac{d}{dt}p_{ij}(t) = \sum_{k\in S} p_{ik}(t)\sigma_{kj}, \qquad \forall i,j \tag{2.4}$$

Це *система прямих рівнянь Колмогорова.*У матричній формі одержимо $\frac{d}{dt}P(t) = P(t)A$, де матриця $A = \{\sigma_{ij}\}$ називається *генеруючою матрицею* стрибкоподібного процесу Маркова.
Якщо простір старів S скінченний, то (2.4) дає при кожному фіксованому i скінченну лінійну систему диференціальних рівнянь (індекс i входить через початкову умову (2.2)). Отже, для заданих інтенсивностей переходів σ_{ij}, система (2.4) має єдиний розв'язок, що задовольняє умову (2.2). Із цього випливає, що марківську модель можна визначити через її інтенсивності переходів σ_{ij}.
Підставивши $s = h$ в (2.1) і повторивши викладки, одержимо інщу систему рівнянь

$$\frac{\mathrm{d}}{\mathrm{d}t}P(t) = AP(t),$$

яка називається *системою оебрнених рівнянь Колмогорова*.
При "звичайних"умовах пряма і обернена системи еквівалентні. Це, зокрема виконується, коли інтенсивності переходів обмежені $\sup_{i,j}|\sigma_{ij}| < \infty \forall t \geqslant 0$. Однак, коли ця умова не виконується, то система обернених рівнянь є більш важливою.
Зауважимо, що із (2.3) випливає, що $\sigma_{ij} \geqslant 0$ для $i \neq j$, але $\sigma_{ii} \leqslant 0$. Дійсно, диференціюючи рівність $\sum_{j\in S} p_{ij}(t) = 1$ відносно t при $t = 0$ одержимо $\sigma_{ii} = -\sum_{i\neq j}\sigma_{ij}$. Отже сума по кожному рядку матриці A дорівнює нулю.

2.4.Процес Пуассона.

Розглянемо стрибкоподібний процес Маркова із простором станів $S = \{0, 1, 2\ldots\}$ та інтенсивностями переходів

$$\sigma_{ij} = \begin{cases} -\lambda, & \text{якщо } j = i, \\ \lambda, & \text{якщо} j = i+1, \\ 0 & \text{у інших випадках.} \end{cases}$$

генеруюча матриця A у рівняннях Колмогорова така:

$$\begin{pmatrix} -\lambda & \lambda & 0 & \dots & \dots & \dots & 0 \\ 0 & -\lambda & \lambda & 0 & \dots & \dots & 0 \\ 0 & 0 & -\lambda & \lambda & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Система прямих рівнянь має вигляд

$$\left\{ \begin{array}{l} p'_{i0}(t) = -\lambda p_{i0}(t), \\ p'_{ij}(t) = \lambda p_{ij-1}(t) - p_{ij}(t), j > 0 \end{array} \right\}.$$

яка співпадає із (1.3) і (1.4).
Цікаво розглянути систему обернених рівнянь

$$p'_{ij}(t) = -\lambda p_{ij}(t) + \lambda p_{i+1,j}(t),$$

яка має той самий розв'язок, хоча прямі рівняння мають інший вигляд.

2.4.Час перебування (holding times).

Покажемо, що час перебування у стані для однорідного стрибкоподібного процесу Маркова має показниковий розподіл. Розглянемо $T_0 = \inf\{t : X_t \neq X_0\}$ – це час до першої зміни стану.
Теорема. T_0 для однорідного стрибкоподібного процесу Маркова з інтенсивностями переходів $\sigma_i j$ має розподіл

$$P(T_0 > t | X_0 = i) = e^{-\lambda_{t} t},$$

де $\lambda_i = -\sigma_{ii} = \sum_{j \neq i} \sigma_{ij}$.

У випадку процесу Пуассона моменти стрибків несуть всю інформацію про процес. Для загальних стрибкоподібних процесів Маркова ми також повинні вказати стан у який відбувається стрибок. Виявляється, що це досить просто: *ймовірність стрибка із $X_0 = i$ у стан не залежить від часу перебування T_0. Щоб перевірити це, розглянемо для $j \neq i$*:

$$\begin{aligned} P(X_{t+h} = j, t < T_0 \leqslant t+h | X_0 = i) &= P(X_{t+h} = j, T_0 > t | X_0 = i) = \\ = P(X_{t+h} = j | X_0 = i, T_0 > t) P(T_0 > t | X_0 = i) &= \\ = P(X_{t+h} = j | x_s = i, 0 \leqslant s \leqslant t) e^{-\lambda_i t} &= \\ = P(X_{t+h} = j | X_t = i) e^{-\lambda_i t} &= p_{ij}(h) e^{-\lambda_i t}. \end{aligned}$$

Далі, поділивши на h і поклавши $h \to 0$, одержимо, що сумісний розподіл/щільність X_{T_0} і T_0, при умові $X_0 = i$,дорівнює $\sigma_{ij} e^{-\lambda_i t}$. Іншими словами, це добуток щільністі часу перебування $\lambda_i e^{-\lambda_i t}$ та величини σ_{ij}/λ_i. Це доводить зразу два результати: ймовірнсть стрибка із стану i у стан j дорівнює

$$P(X_{T_0} = j | X_0 = i) = \frac{\sigma_{ij}}{\lambda_i} \qquad (j \neq i)$$

та X_{T_0} не залежить від T_0.
Із Марківської властивості випливає, що для наступних стрибків поведінка буде ідентичною: після того, як деякий стан j було досягнуто, час який процес проведе у цьому стані матиме показниковий розподіл з параметром λ_j, а потім процес зробить стрибок у стан $k \neq$ із ймовірністю σ_{jk}/λ_j.
Зауважимо, що середній час перебування у стані j дорівнює $1/\lambda_j$. Це важливо пам'ятати, при вободі значень інтенсивності переходу.

2.5.Ланцюг стрибків.

Якщо стрибкоподібний процес Маркова вивчати лише у моменти його переходів, то одержимо процес $\{\hat{X}_n : n = 0, 1, 2 \ldots\}$ $\hat{X}_n = X_{T_0+T_1+\cdots+T_n}$, який називають *ланцюгом стрибків* пов'язаним з X. Попередній аналіз показав, що $\hat{X}_n$ не залежить від T_n і всього , що відбулось до моменту T_{n-1}. У дійсності, розподіл $\hat{X}_n$ залежить лише від $\hat{X}_{n-1}$. Таким чином ланцюг стрибків є ланцюгом Маркова. Єдине чим відрізняється ланцюг стрибків від стандартного ланцюга Маркова, це ситуація, коли стрибкоподібний процес $\{X_t : t \geqslant 0\}$ попадає у поглинаючий стан. З цього моменту не відбувається подальших преходів а отже це означає, що час зупинився для ланцюга стрибків. Для того, щоб застосувати теорію ланцюгів Маркова до ланцюга стрибків потрібно вважати, що для поглинаючого стану переходи продовжують відбуватись, але ланцюг залишається після кожного переходу у тому ж самому стані.
На питання стосовно послідовності станів, у які відбулися переходи стрибкоподібного процесу Маркова, такі як "Яка ймовірність того, що процес попав у стан i_0 до того, як він попав у поглинаючий стан?"або "Чи попадав процес у стан j нескінченну кількість раз?"можна відповісти використовуючи ланцюг стрибків, бо обидва процеси мають однакові траєкторії у просторі станів. Таким чином у цих випадках можна використовувати теорію линцюгів Маркова. На питання стосовно часу перебування у стані можна віповісти лише використовуючи теорію стрибкоподібних процесв Маркова.

А

1. Пацієнти, які приїжджають у відділення травматології та невідкладної допомоги (стан А), у середньому чекають годину поки молодший лікар визначить, чи потребують вони стаціонарного лікування (І), амбулаторного лікування (О) або подальшого обстеження (F). Тільки один з десяти новоприбулих залишається для стаціонарного лікування, для п'ятьох з десяти призначається амбулаторне лікування. Якщо така потреба виникає, подальше обстеження займає в середньому 3 години, після чого 50% пацієнтів виписують (D), 25% призначається амбулаторне лікування, і 25% залишають для стаціонарного лікування. Амбулаторне лікування займає в середньому 2 години, а стаціонарне – 60 годин. Обидва типи лікування закінчується виписуванням пацієнта. Припускається, що для моделювання просування пацієнта у межах системи можна використати однорідний

марковський процес із станами A,I,O,F та D з кінцевою метою зменшити середній час, проведений пацієнтами у лікарні.

а) Запишіть матрицю переходів такої $\{\mu_{ij} : i, j = A, I, O, F, D\}$ моделі.
б) Обчисліть частку пацієнтів, які врешті-решт отримують стаціонарне лікування.
в) Виведіть вираз для ймовірності того, що пацієнт, який прибув у час t=0:
в.1) ще не є класифікованим молодшим лікарем у час t
в.2) проходить додаткове обстеження у час t
г) Нехай e_i позначає математичне сподівання часу до виписування для пацієнта, який зараз знаходиться у стані i
г.1) Поясніть словами, чому e_i задовольняє наступне рівняння:

$$e_i = \frac{1}{\lambda_i} + \sum_{j \notin \{i,D\}} \frac{\mu_{ij}}{\lambda_i} e_i$$

де

$$\lambda_i = \sum_j \mu_{ij}$$

г.2) Обчисліть математичне очікування сукупного часу до виписування пацієнта, який щойно прибув

2. Розглянемо однорідний стрибкоподібний марківський процес $\{X(t) : t \geqslant 0\}$ з двома станами, які позначаються 0 та 1, та двома переходами

$$\sigma_{01} = \lambda$$

та

$$\sigma_{10} = \mu$$

а) Запишіть пряме рівняння Колмогорова для ймовірності $P_{0,0}(t)$, що X знаходиться в стані 0 у час t, знаючи що він починає зі стану 0
б) Покажіть що:

$$P_{0,0}(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)}$$

в) Нехай O_t позначає сукупний час, витрачений у стані 0 до часу t, що може бути виражене як $O_t = \int_0^t I_s \, ds$, де $I_s = \begin{cases} 1, & \text{якщо X(s) = 0} \\ 0, & \text{в іншому випадку} \end{cases}$.
Використовуючи результат у частині (ii), виведіть вираз для $E[O_t|X(0) = 0]$, математичного сподівання часу перебування у стані 0 до часу t для неперервного марковського ланцюга з двома станами, що починається у стані 0

3. Неперервна марківська модель хвороби та смерті має чотири стани: H (здоровий), S (хворий), T (невиліковно хворий) та D (мертвий). Зі здорового стану можна перейти у стани S та D, кожний з інтенсивністю переходів 0.05 на рік. Хвора людина виліковується з майже напевно кожного року; також можливі переходи до станів D та T, кожен з інтенсивністю 0.1

на рік. Зі стану невиліковно хворого можливий перехід тільки до стану D з інтенсивністю переходу 0.4 на рік.

а) Намалюйте граф переходів для цього процесу. б) Визначте P(t)=$\{p_{ij}(t)$: $i, j\ \in \{H, S, T, D\}\}$, де $p_{ij}(t)$ позначає ймовірність бути у стані j у час t, знаючи, що особа була в стані i у час 0. Запишіть пряме рівняння Колмогорова, яке задовільняє матриця $P(t)$, упевнившись, що ви вказали всі елементи генеруючої матриці
в) Обчисліть ймовірність бути здоровим протягом 10 років підряд, знаючи що особа зараз здорова
г) Нехай d_j позначає ймовірність, що людина, яка зараз перебуває у стані j, ніколи не буде невиліковно хворою. Розглядаючи перший перехід від стану , покажіть, що $d_H = \frac{1}{2} + \frac{1}{2}\ d_S$, і схожим чином виведіть $d_S = \frac{1}{12} + \frac{5}{6}$ d_H. Звідси оцініть d_H та d_S

4. Розглянемо спрощену неперервну модель кредитного рейтингу компаній, яку описано нижче. Усього існує три ранги, які може отримати компанія, A, B та D; при цьому можливі наступні переходи:

- з A в B з інтенсивністю 4α
- з B в A з інтенсивністю α
- з B в D з інтенсивністю 3α

Запишіть прямі рівняння Колмогорова у матричній формі для цієї задачі та покажіть, що матриця переходів $P(t) = P(0, t)$, яка наведена нижче, дійсно є її розв'язком

$$P(t) = \begin{bmatrix} \frac{1}{2}e^{-2\alpha t} + \frac{1}{2}e^{-6\alpha t} & e^{-2\alpha t} + e^{-6\alpha t} & 1 - \frac{3}{2}e^{-2\alpha t} + \frac{1}{2}e^{-6\alpha t} \\ \frac{1}{4}e^{-2\alpha t} + \frac{1}{4}e^{-6\alpha t} & \frac{1}{2}e^{-2\alpha t} + \frac{1}{2}e^{-6\alpha t} & 1 - \frac{3}{4}e^{-2\alpha t} + \frac{1}{4}e^{-6\alpha t} \\ 0 & 0 & 1 \end{bmatrix} \tag{2.5}$$

Знайдіть час t, після якого компанія, що почала зі стану A має більшу ймовірність бути у стані D, ніж у стані A.

5. Програма допомоги з непрацездатності у неперервному часі моделюється стрибкоподібним марківським процесом зі станами A (активний), T (тимчасово непрацездатний), P (постійна непрацездатність) та D (мертвий). Інтенсивності переходів наведено нижче:
A $\to$ T: 3λ; A $\to$ P: λ; A $\to$ D: α;
T $\to$ A: 5λ; T $\to$ P: 2λ; T $\to$ D: α; P $\to$ D: 2α;

а) Запишіть генеруючу матрицю процесу
б) Обчисліть ймовірність того, що процес, який розпочався у стані A, до часу t не побуває ні в стані T, ні в стані P
в) Запишіть у матричному вигляді зворотні диференційні рівняння Колмогорова та використайте їх для виведення диференційного рівняння для

$p_{PD}(t)$, тобто ймовірності того, що член програми в стані P у час 0 буде в стані D у час t
г) Розв'яжіть рівняння для $p_{PD}(t)$ - ймовірності, що початково постійно непрацездатний член програми, буде мертвим до моменту часу t

6. Каса на залізничному вокзалі обладнана тільки одним квитковим автоматом, що використовується пасажирами для купівлі квитків. Автомат має схильність ламатися, і тоді його потрібно лагодити. І час до поломки, і час, необхідний для налагодження автомату, мають експоненційний розподіл. Нехай $P_{1i}(t), i = 0, 1$, позначає ймовірність, що в час $t,\ t > 0$ у касі працює i квиткових автоматів, причому квиткові автомати працюють у час $t = 0$.

а) Виведіть прямі рівняння Колмогорова для $P_{1i}(t), i = 0, 1$ відносно:
а1) σ, де $\frac{1}{\sigma}$ - середній час до поломки автомата
а2) ρ, де $\frac{1}{\rho}$ - середній час, необхідний для налагодження автомата
б) Покажіть, що:

$$P_{10}(t) = \frac{\sigma}{\sigma + \rho}(1 - e^{-(\sigma+\rho)t})$$

та виведіть значення для P_{11}(t)
в) Керуючий станцією роздумує над встановленням ще одного подібного автомату, незважаючи на те, що тільки одна ремонтна команда має змогу обслуговувати автомати навіть у випадку, якщо обидва вони вийдуть з ладу одночасно. Припускаючи, що другий квитковий автомат встановлено і він працює незалежно від першого:

в.1) Запишіть матрицю переходів стрибкоподібного марківського процесу X_t, який рахує кількість працюючих квиткових автоматів у час t
в.2) Виведіть прямі диференційні рівняння Колмогорова для $_{1i}(t)$,i=0,1,2, ймовірності, що i автоматів працюють
в.3) Знаючи, що для певного t, p_0(t) = $\frac{2\sigma^2}{2\sigma^2+2\rho\sigma+\rho^2}$, p_1(t) = $\frac{2\rho\sigma}{2\sigma^2+2\rho\sigma+\rho^2}$, та p_2(t) = $\frac{2\rho^2}{2\sigma^2+2\rho\sigma+\rho^2}$
Покажіть, що $\frac{d}{dt}p_i(t) = 0$ для i=0,1,2

7. Кількість членів пенсійної схеми (N(t)), які отримують пенсію у час t, може змінюватися у два способи:
- збільшується на 1, коли член схеми досягає пенсійного віку;
- зменшується на 1, коли член на пенсії помирає.
Припустимо, що вихід на пенсію відбувається відповідно до пуасонівського процесу з частотою λ, а також що кожен член на пенсії, незалежно від інших, має ймовірність померти в інтервалі часу $(t, t + dt)$.

а) Поясніть чому за таких припущень N(t) є стрибкоподібним марківським процесом
б) Запишіть переходи для N(t)

в) Використовуючи позначення $p_n(t)$ = P(N(t)=n), отримайте диференціальне рівняння, яке задовольняється $p_n(t)$
г) Перевірте, що один з розв'язків у завданні в) виглядає так:

$$p_n(t) = \frac{1}{n!} e^{-\frac{\lambda}{\mu}} (\frac{\lambda}{\mu})^n$$

8. Телефонний call-центр отримує дзвінки від клієнтів із середньою частотою 0.5 на хвилину. Кожен дзвінок має випадкову тривалість, яка розподілена експоненційно з математичним сподіванням 3 хвилини, незалежно від кількості чи тривалості інших дзвінків. Для роботи зі дзвінками призначено два оператори. Якщо дзвінок надходить, коли обидва оператори зайняті, то дзвінок ставлять на очікування, крім випадку, якщо на очікування вже передано два дзвінки – тоді новий дзвінок губиться. Коли дзвінок закінчується, один з дзвінків на очікуванні відразу передається до оператора, який звільнився.

а)Визначте п'ять станів, які необхідні, щоб ця система могла бути змодельована як стрибкоподібний марківський процес
б) Намалюйте діаграму переходів для цієї системи
в) Запишіть матрицю переходів для процесу

9. Транспортні засоби у певній країні повинні бути перевірені кожен рік для придатності для їзди на дорогах. В одному з центрів перевірки транспортних засобів, водії чекають у середньому 15 хвилин поки почнеться перевірка їхніх авто. Перевірка займає у середньому 20 хвилин. Після перевірки, 80% транспортних засобів визнаються придатними, і водіям дозволено поїхати додому. 15% автомобілів класифікуються як такі що мають "незначні пошкодження"; такі транспортні засоби потребують у середньому 30 хвилин ремонтних робіт поки водій може поїхати додому. Останні 5% транспортних засобів класифікуються як такі що мають "значні пошкодження" і потребують у середньому 3 години ремонтних робіт, поки водій може поїхати додому.

Неперервна марківська модель зі станами W (очікує перевірки), A (на перевірці), M (проводиться незначний ремонт), S (проводиться значний ремонт) та H (на шляху додому) може бути використана для моделювання роботу центра перевірки транспортних засобів.

а) Поясніть, які припущення необхідно зробити щодо розподілу часу перебування в кожному стані
б) Запишіть генеруючу матрицю для цього процесу
в) Використайте прямі рівняння Колмогорова для того, щоб записати диференційні рівняння, які задовольняються p_{WM}(t) та p_{WA}(t). г) Перевірте, що:

$$p_{WA}(t) = 4e^{\frac{-t}{20}} - 4e^{\frac{-t}{15}}$$

де t вимірюється у хвилинах, $t \geq 0$
) Виведіть вираз для p_{WM}(t) для $t \geq 0$ д) Нехай T_i буде очікуваною кіль-

кістю часу (у хвилинах) поки автомобіль можна забрати з центру, знаючи, що процес перевірки наразі в стані i.
д.1)Поясніть, чому $T_W = 15 + T_A$
д.2) Виведіть відповідні рівняння для T_A, T_M та T_S д.3) Обчисліть T_W

10. Сімейний стан розглядається з використанням наступного однорідного неперервного стрибкоподібного марківського процесу:
- інтенсивність переходу з незаміжнього до одруженого дорівнює 0.1 на рік
- інтенсивність переходу з незаміжнього до одруженого дорівнює 0.1 на рік
- смертність еквівалентна інтенсивності 0.025 на рік, незалежно від сімейного стану

Простір станів для цього процесу включає: ніколи не був одруженим (NM), одружений (M), вдівець (W), розлучений (DIV), мертвий (D).
P_x – ймовірність того, що особа, яка зараз перебуває в стані x, ніколи не була вдівцем і до смерті вже не буде в цьому стані.

а) Побудуйте діаграму переходів між п'ятьма станами
б Покажіть за допомогою загальних міркувань або іншим чином, що P_{NM} дорівнює P_{DIV}
в) Покажіть, що:

$$P_{NM} = \frac{1}{5} + \frac{4}{5}P_M$$
$$P_M = \frac{1}{4} + \frac{1}{2}P_{DIV}$$

г) Обчисліть ймовірність того, що особа не стане вдівцем, знаючи, що зараз вона знаходиться в стані NM
д) Запропонуйте два способи, які можуть зробити цю модель більш реалістичною

11. Страховий поліс покриває ремонт пральної машини і дозволяє максимум 3 заяви протягом року покриття.
Ймовірність поломки пральної машини має експоненційний розподіл з наступними (приведеними до річних) частотами, λ:

$$\lambda = \begin{cases} \frac{1}{10}, & \text{якщо машина раніше не ламалася} \\ \frac{1}{5}, & \text{якщо машина до цього ламалася один раз} \\ \frac{1}{4}, & \text{якщо машина до цього ламалася два рази або більше} \end{cases}.$$

Як тільки виникає поломка, на місце висилають інженера. Можна припустити, що ремонт відбувається миттєво і що машину завжди можна полагодити. На момент початку року (час t=0), машина ніколи не ламалася.
$P_i(t)$ – ймовірність того, що машина мала і поломок до часу t.

а) Намалюйте схему переходів для цього процесу, який визначається кількістю поломок, які трапилися до часу t
б) Запишіть рівняння Колмогорова, яким слідують P_0'(t), P_1'(t) та P_2'(t)
в) Запишіть вираз для P_0(t)

г) Покажіть, що:

$$P_1(t) = e^{\frac{-t}{10}} - e^{\frac{-t}{5}}$$

д) Виведіть вираз для P_2(t)
е) Обчисліть очікувану кількість заяв за полісом

12. Стрибкоподібний марківський процес X_t із простором станів $S = \{0, 1, 2, \ldots, N\}$ має наступні інтенсивності переходів:
$\sigma_{ii} = -\lambda$ для $0 \leq i \leq N$
$\sigma_{ii+1} = \lambda$ для $0 \leq i \leq N-1$
$\sigma_{ij} = 0$ для $0 \leq i \leq N-1$

а) Запишіть генеруючу матрицю та прямі рівняння Колмогорова (у формі компонент), які пов'язані з цим процесом
б) Перевірте, що для $0 \leq i \leq N-1$ та для всіх $j \geq i$, функція

$$p_{ij}(t) = e^{-\lambda t}\frac{\lambda t^{j-i}}{(j-1)!}$$

є розв'язком прямих рівнянь в завданні 1. в) Визначте розподіли часів чекання, пов'язані з цим стрибкоподібним процесом

3. Моделі виживання.

3. Базові актуарні моделі

Теоретичні відомості

3.1.Проста модель виживання (Модель 1).

3.1.1. Основним припущенням у простій математичній моделі виживання є те, що тривалість майбутнього життя особи ("життя"у актуарних роботах) є невідомим заздалегіть. Далі ми розглядаємо тривалість життя у межах від 0 до 100 років. Природно вважати тривалість майбутнього життя випадковою величиною.

3.1.2. Припущення. Тривалість майбутнього життя новонародженої особи є випадковою величиною T, яка має неперервний розподіл на інтервалі $[0, \omega]$, де $0 < \omega < \infty$.

3.1.3. Максимальний вік ω називається *граничним віком*. У практичних роботах типові значення ω змінюються у межах 100-200 років. Можливість прожити довше за ω відкидаєтся для зручності і простоти.

3.1.4. Означення. $F(t) = P(T \leqslant t)$ - це функція розподілу, а $S(t) = P(T > t) = 1 - F(t)$ - це функція виживання для T.

3.1.5. Нам часто доведеться мати справу з віком большім за нульовий. Щоб враховувати це, позначимо через T_x тривалість майбутнього життя після віку x, особи, яка дожила до віку x, $0 \leqslant x \leqslant \omega$. Зауважимо, що $T_0 = T$.

3.1.6. Означення. Позначимо через $F_x(t) = P(T_x \leqslant t), (0 \leqslant x \leqslant \omega)$ функцію розподілу T_x, а через $S_x(t) = P(T_x > t) = 1 - F_x(t)$ функцію виживання для T_x.

3.1.7. Для узгодженості з T, функція розподілу випадкової величини $T_x(0 \leqslant x > \omega)$ повинна задовільняти наступному співвідношенню

$$F_x(t) = P(T_x \leqslant t) = P(T \leqslant x + t | T > x) = \frac{F(x+t) - F(x)}{S(x)}$$

3.1.8. Зараз ми введемо позначення, які використовують актуарії для ймовірностей смерті та виживання:

$$_tq_x = F_x(t) \quad \text{i} \quad _tp_x = 1 - {_tq_x} = S_x(t).$$

У більшості актуарних робіт використовується одиниця часу в один рік. У цьому випадку, коли $t = 1$ ми будемо відкидати індекс t у цих ймовірностях. Так ми позначимо

$$q_x = {_1q_x} \quad \text{i} \quad p_x = {_1p_x}.$$

q_x і $_tq_x$ називаються *інтенсивностями смертності*.

3.1.9. Величина, яка грає центральну роль у моделях виживання- це *сила смертності*, яка більше відома у статистиці, як інтенсивність ризику. Силу смертності у віці $x(0 \leqslant x < \omega)$ будемо позначати через μ_x і визначимо так

$$\mu_x = \lim_{h \to 0+} \frac{P(T \leqslant x + h | T > x)}{h}.$$

Завжди будемо вважати, що ця границя завжди існує.

3.1.10. Дуже важливою є інтерпритація μ_x. Маємо $P(T \leqslant x + h | T > x) = F_x(h) = {_hq_x}$. Для малих h можна покласти $_hq_x \cong \mu_x$. Іншими словами ймовірність смерті на короткому проміжку часу h після віку x приблизно пропорційна h з коефіцієнтом пропорційності рівним μ_x.

3.1.11. Для $x \geqslant 0$ і $t > 0$ ми можемо визначити силу смертності μ_{x+t} двома шляхами:

$$(1) \quad \mu_{x+t} = \lim_{h \to 0+} \frac{P(T \leqslant x + t + h | T > x + t)}{h},$$

$$\mu_{x+t} = \lim_{h \to 0+} \frac{P(T_x \leqslant t + h | T_x > t)}{h}.$$

Легко показати, що ці означення еквівалентні. Ми будемо часто використовувати μ_{x+t} для фіксованого віку x та $0 \leqslant t < \omega - x$.

3.1.12. Із означення $S_x(t)$ випливає важливе співвідношення.

$$S_x(t) = P(T_x > t) = P(T > x+t | T > x) = \frac{P(T > x+t)}{P(T > x)} = \frac{S(x+t)}{S(x)},$$

яке можна переписати у актуарних позначеннях так

$${}_tp_x = \frac{{}_{x+t}p_0}{{}_xp_0}.$$

Таким чином, для довільного віку x і для $s > 0, t > 0$ маємо

$${}_{s+t}p_x = \frac{{}_{x+s+t}p_0}{{}_xp_0} = \frac{{}_{x+s}p_0}{{}_xp_0} \times \frac{{}_{x+s+t}p_0}{{}_{x+s}p_0} = {}_sp_x \times {}_sp_{x+t}.$$

Аналогічно ${}_{s+t}p_x = {}_tp_x \times {}_sp_{x+t}$. Словами: ймовірність прожити час $s+t$ після віку x дорівнює або (i) добутку ймовірності прожити час s після віку x і ймовірності прожити ще час t, або (ii) добутку прожити час t після віку x і ймовірності прожити ще час s.

3.1.13. Функцією розподілу T_x за означенням є $F_x(t)$. Нам буде потрібна щільність розподілу T_x. Позначимо цю щільність розподілу через $f_x(t)$ і нагадаємо, що $f_x(t) = \frac{d}{dt}F_x(t)$. Тому

$$f_x(t) = \frac{d}{dt}P(T_x \leqslant t) = \lim_{h\to 0+} \frac{P(T_x \leqslant t+h) - P(T_x \leqslant t)}{h} =$$
$$= \lim_{h\to 0+} \frac{P(T \leqslant x+t+h | T > x) - P(T \leqslant x+t | T > x)}{h} =$$
$$= \lim_{h\to 0+} \frac{P(T \leqslant x+t+h) - P(T \leqslant x) - (P(T \leqslant x+t) - P(T \leqslant x))}{h} =$$
$$= \lim_{h\to 0+} \frac{P(T \leqslant x+t+h) - P(T \leqslant x+t)}{S(x)h}.$$

Помноживши і поділивши цей вираз на $S(x+t)$ одержимо

$$f_x(t) = \frac{S(x+t)}{S(x)} \times \lim_{h\to 0+} \frac{P(T \leqslant x+t+h) - P(T \leqslant x+t)}{S(x+t)h} =$$
$$= S_x(t) \times \lim_{h\to 0+} \frac{P(T \leqslant x+t+h | T > x+t)}{h} = S_x(t) \times \mu_{x+t},$$

або в актуарних позначеннях для фіксованого віку з проміжку $[0, \omega]$

$$f_x(t) = {}_tp_x \times \mu_{x+t}, \quad (0 \leqslant t < \omega - x).$$

Це один із наважливіших результатів теорії моделей виживання.

3.1.14. Таким чином ми маємо у моделі: T_x– випадкову тривалість майбутнього життя після віку x; ця випадкова величина за припущенням є неперервною, та приймає значення з проміжка $[0, \omega - x]$; функція розподілу цієї випадкової величини $F_x(t) = {}_tq_x$ і щільність розподілу $f_x(t) = {}_tp_x\mu_{x+t}$.

Сила смертності наближенно визначається співвідношенням: ${}_hq_x \cong h \cdot \mu_x$ (для малих h). Функція виживання $S_x(t) = {}_tp_x$ задовольняє співвідношенню ${}_{s+t}p_x = {}_sp_x \cdot {}_tp_{x+s} = {}_tp_x \cdot {}_sp_{x+t}$ для довільних $s > 0, t > 0$.

3.1.15. q_x називається *початковою інтенсивністю смертності*, тому що це ймовірність того, що особа, яка жива у віці x(початковий час) помре до віку $x + 1$. Як альтернативу часто використовують, особливо у демографії, *центральну інтенсивність смертності* m_x, яка визначається так

$$m_x = \frac{d_x}{\int_0^1 l_{x+t}dt} = \frac{q_x}{\int_0^1 {}_tp_x dt}.$$

Тут l_x– це число осіб, які досягли віку x; $d_x = l_x - l_{x+1}$– це число осіб, які померли протягом періоду $(x, x + 1)$. Знаменник $\int_0^1 {}_tp_x dt$ розуміється, як очікуваний час проведений живими особами віку x протягом вікового інтервалу між x і $x + 1$.
m_x буває корисною, коли потрібно передбачити число сметрей при заданій кількості живих осіб у кожній віковій групі. Ця величина є однією із основних компонент у передбаченні кількості населення, і на практиці вікові групи часто визначаються періодом часу тривалішим ніж 1 рік (тому означення m_x потрібно відповідно модифікувати).
Історично величину m_x оцінюють статистикою вигляду

$$\frac{\text{Число смертей}}{\text{Загальний час виживання під ризиком}},$$

яка називається "інтенсивністю події-експозиції". Останнім часом ця статистика використовується більше для оцінки сили смертності ніж для оцінки m_x, тому що у цьому випадку існує строге обґрунтування у рамках моделі. Однак, якщо m_{x+t} є стала між віком x і віком $x + 1$, тоді

$$m_x = \frac{q_x}{\int_0^1 {}_tp_x dt} = \frac{\int_0^1 {}_tp_x \cdot \mu dt}{\int_0^1 {}_tp_x dt} = \mu.$$

3.2.Повна і урізана очікувана тривалість залишку життя.

3.2.1. *Повна очікувана тривалість залишку життя* у віці x визначається як $E[T_x]$ і позначається через $\mathring{e}_x$. За означенням

$$\mathring{e}_x = \int\limits_0^{\omega-x} t \cdot {}_tp_x \cdot \mu_{x+t} dt = \int\limits_0^{\omega-x} t \cdot \Big(-\frac{\partial}{\partial t} {}_tp_x\Big) dt = -(t \cdot {}_tp_x)|_0^{\omega-x} + \int\limits_0^{\omega-x} {}_tp_x dt = \int\limits_0^{\omega-x} {}_tp_x dt$$

3.2.2. Щоб визначити урізану очікувану тривалість залишку життя, спочатку визначимі K_x – урізану майбутню тривалітсь життя особи віку x. За означенням $K_x = [T_x]$, де $[x]$– це ціла частина числа x. Очевидно, що K_x– це дискретна випадкова величина, яка приймає цілі значення $0, 1, \ldots, [\omega - x]$.

Розподіл ймовірностей K_x легко знайти, використавші означення пидрозділу 1.

$$P(K_x = k) = P(k \leqslant T_x < k+1) = P(k < T_x \leqslant k+1) = {}_kp_x \cdot q_{x+k}. \quad (*)$$

Зауважимо, що зміна нерівностей у $(*)$ вимагає певних припущень стосовно T_x. Достатньо припустити, що $F_x(t)$ неперервна за t.

3.2.3. Визначимо *урізану очікувану тривалість залишку життя* $e_x = E[K_x]$. Тому

$$e_x = \sum_{k=0}^{[\omega-x]} k \cdot {}_kp_x \cdot q_{x+k} = \sum_{k=1}^{[\omega-x]} {}_kp_x \cdot q_{x+k} + \sum_{k=2}^{[\omega-x]} {}_kp_x \cdot q_{x+k} + \ldots + {}_{[\omega-x]}p_x \cdot q_{x+[\omega-x]} =$$
$$= \sum_{k=1}^{[\omega-x]} \sum_{j=k}^{[\omega-x]} {}_jp_x \cdot q_{x+j} = \sum_{k=1}^{[\omega-x]} \sum_{j=k}^{[\omega-x]} P(K_x = j) = \sum_{k=1}^{[\omega-x]} P(K_x \geqslant k) =$$
$$= \sum_{k=1}^{[\omega-x]} P(T_x > k) = \sum_{k=1}^{[\omega-x]} {}_kp_x.$$

3.2.4. Ми маємо дві прості формули

$$\mathring{e}_x = \int_0^{\omega-x} {}_tp_x dt \quad \text{i} \quad e_x = \sum_{k=1}^{[\omega-x]} {}_kp_x.$$

Ці величини пов'язані наближеним співвідношенням $\mathring{e}_x \cong e_x + 1/2$. Щоб довести це визначимо випадкову величину $J_x = T_x - K_x$. Наближено $E[J_x] \approx 1/2$, крім того $E[T_x] = E[J_x] + E[K_x]$. Тому $\mathring{e}_x \cong e_x + 1/2$.

3.2.5. Легко записати вирази для дисперсії повної і урізаної тривалості життя:

$$\mathrm{Var}[T_x] = \int_0^{\omega-x} t^2 \cdot {}_tp_x \cdot \mu_{x+t} dt - \mathring{e}_x^2, \quad \mathrm{Var}[K_x] = \sum_0^{[\omega-x]} k^2 \cdot {}_kp_x \cdot q_{x+k} - e_x^2,$$

але ці вирази не можна значно спростити, як вирази для середніх.

3.2.6. Очікувану тривалість життя часто використовують як міру стандартів життя і рівня охорони здоров'я у даній країні.

3.3.Деякі важливі формули.

3.3.1. У цьому підрозділі ми навведемо дві важливі формули, одну для ${}_tq_x$ і одну для ${}_tp_x$. Перша є наслідком того, що $f_x(t) = f_x(t) = {}_tp_x \cdot \mu_{x+t}$:

$${}_tq_x = F_x(t) = \int_0^t f_x(s) ds = \int_0^t {}_sp_x \cdot \mu_{x+s} ds.$$

Ця формула має просту інтерпретацію. Для кожного $s \in [0,t]$ підінтегральна функція є добутком ${}_sp_x$ – ймовірності вижити до віку $x+s$ і величини $\mu_{x+s} \cdot ds$, яка наближено дорівнює ${}_{ds}q_{x+s}$ – ймовірності померти зразу після віку $x+s$. Оскількі неможливо померти больше ніж один раз, то ми просто додаємо ці взаємовиключні ймовірності, а у границі маємо інтеграл.
3.3.2. Формулу для ${}_tp_x$ одержимо із диференціального рівняння. Маємо

$$\frac{\partial}{\partial s}{}_sp_x = \frac{\partial}{\partial s}(1 - {}_sq_x) = -\frac{\partial}{\partial s}{}_sq_x = -f_x(t) = -{}_sp_x \cdot \mu_{x+s}.$$

Звідси

$$\frac{\frac{\partial}{\partial s}{}_sp_x}{{}_sp_x} = -\mu_{x+s}, \quad {}_0p_x = 1.$$

Розв'язавши це рівняння одержимо

$${}_tp_x = \exp\Big\{ -\int\limits_0^t \mu_{x+s} ds\Big\}.$$

3.4.Прості закони смертності.

3.4.1. Часто буває корисною можливість виразити q_x або μ_x у вигляді простих математичних функцій віку x. Два таких найпростіших "закони"мають вигляд:
Закон Гомперца $\mu_x = Bc^x$;
Закон Мейкхема $\mu_x = A + Bc^x$.
3.4.2. Закон Гомперца– це показникова функція, і часто вона є розумним припущенням для середнього і старшого віку. Закон Мейкхема містить сталий додаток, який іноді інтерпритується, як можливість сметрі у результаті нещасного випадку, незалежно вік віку.
3.4.3. Якщо виконується закон Гомперца, то параметри B і c можна визначити по заданим значенням μ_x для двох довільних значень віку. Якщо виконується закон Мейкхема, то параметри A, B і c можна визначити по заданим значенням μ_x для трьох значень віку.
3.4.4. Ймовірність виживання ${}_tp_x$ можна знайти використовуючи співвідношення

$${}_tp_x = \exp\Big\{ -\int\limits_0^t \mu_{x+s} ds\Big\}.$$

Наприклад, для закону Гомперца маємо

$${}_tp_x = g^{c^x(c^t-1)}, \quad \text{де } g = \exp\Big\{\frac{-B}{\ln(c)}\Big\}.$$

А у випадку закону Мейкхема, маємо

$${}_tp_x = s^t \cdot g^{c^x(c^t-1)}, \quad \text{де } g = \exp\Big\{\frac{-B}{\ln(c)}\Big\} \quad \text{і } s = \exp\{-A\}.$$

A

1. Детально поясніть, що означає ${}_{3|5}q_{50}$. Запишіть інтегральний вираз для ${}_{3|5}q_{50}$ відносно коефіцієнтів смертності тільки для відповідного вікового інтервалу

2. Покажіть, що у віці x, якщо $0 \leq a < b \leq 1$, то

$$ {}_{b-a}q_{x+a} = 1 - \frac{{}_bp_x}{{}_ap_x} $$

Таким же чином, або в інший спосіб, покажіть, що за умови рівномірного розподілу смертності на інтервалі (x,x+1), виконано:

$$ {}_{b-a}q_{x+a} = \frac{(b-1)q_x}{1-aq_x} $$

3. Знаючи, що $p_x = 0.9$, обчисліть ${}_{0.5}q_x$ та ${}_{0.5}q_{x+0.5}$, використовуючи наступні припущення щодо смертності між віком x та x+1:
а. Припущення щодо рівномірного розподілу смертності.
б. Припущення Балдуччі.
Прокоментуйте, наскільки доречним є кожне з цих припущень

4. Поясніть, що мається на увазі під припущенням про рівномірний розподіл смертності між віком x та x+1
Використовуючи припущення про рівномірний розподіл смертності та ELT15 (Чоловіки) таблицю смертності, обчисліть ${}_{0.25}p_{75}$ та ${}_{0.25}p_{75.5}$

5. Вкажіть вікові інтервали, для яких закон Ґомпертца є доречною моделлю для людської смертності
Покажіть, що за законом Ґомпертца, ймовірність дожиття від віку x до віку x+t дорівнює:

$$ (e^{-\frac{B}{logc}})^{e^x(e^t-1)} $$

6. Нехай T_x позначає майбутню тривалість життя особи, яка зараз перебуває у віці x. Запишіть функцію щільності для T_x
Використовуючи відповіді до попереднього завдання, покажіть що:

$$ \frac{d}{ds} log_s(p_x) = -\mu_{x+s} $$

$$ (e^{-\frac{B}{logc}})^{e^x(e^t-1)} $$

Нехай інтенсивність смертності дорівнює:
$60 < x \leq 70$ - 0.01
$70 < x \leq 80$ - 0.015
$x > 80$ - 0.025

Обчисліть ймовірність того, що особа у віці 65 років (рівно) помре між 80 та 83 роками життя

7. Дані, що подані нижче, походять з національної таблиці дожиття по поколіннях розвинутої країни:
Вік x = 80 - l_x = 22933
Вік x = 81 - l_x = 22010

1. Оцініть ${}_{0.5}p_{80}$, припускаючи, що:
   a) смертність між віком 80 та 81 розподілена рівномірно
   b) інтенсивність смертності не змінюється між віком 80 та 81

2. Поясніть, чому дві оцінки в завданні 1. відрізняються. Зазначте, яку оцінку ви би використали, і чому

8. Виведіть вираз відносно ${}_tp_x$ для повного математичного сподівання кількості тих, хто дожив до віку x
Для певного населення інтенсивність смертності постійна і складає 0.02 для всіх вікових категорій. Для особи, якій зараз 25 років, обчисліть математичне сподівання тривалості життя.

9. Покажіть, що, у випадку коли інтенсивність смертності $\mu_(x+t)$, $0 \leq t \leq 1$ дорівнює

$$\mu_(x+t) = \frac{q_x}{1 - tq_x}$$

смертність між віком та $+1$ є рівномірно розподіленою.

4. Оцінка розподілу тривалості життя.

4. Оцінка розподілу тривалості життя.

Література: [1,с.194-197]

Теоретичні відомості

4.1.Оцінка Каплана-Мейера (оцінка добуток-границя).

4.1.1 У цьому підрозділі ми побудуємо емпіричну функцію розподілу, яка допускає цензурування.

4.1.2 Будемо розглядати тривалість життя як функцію часу t, без врахування початкового віку x. Одержані результати можна буде застосувати для новонароджених, для тих хто досяг віку x, або для тих осіб, які мали спільні властивості в момент часу $t = 0$, наприклад мали однаковий діагноз медичного стану. Медичні дослідження часто грунтуються на моменті постановки діагнозу, або на моменті початку лікування, і якщо вік пацієнта враховується прі аналізі, то лише як пояснювальна змінна у регресійній моделі.

4.1.3 Припустимо, що ми спостерігаємо за популяцією з N осіб при наявності неінформативного цензурування, і припустимо, що ми виявили m смертей. Нехай $t_1 < t_2 < \ldots < t_k$–це впорядковані моменти часу, в які спостерігалась смерть. Ми не припускаємо, що $k = m$, тому в один і той же момент часу можна спостерігати більше ніж одну смерть. Припустимі, що d_j смертей спостерігали в момент $t_j (1 \leqslant j \leqslant k)$, тому $d_1 + d_2 + \ldots + d_k = m$. Спостереження інших $N - m$ осіб цензуровано (тобто ми не будемо намагатись продовжувати спостереження за цими особами у нашому дослідженні). Припустимо, що c_j осіб цензуровано протягом проміжку часу між t_j і $t_{j+1} (0 \leqslant l \leqslant k)$, де $t_0 = 0, t_{k+1} = +\infty$, щоб допускати цезурування спостережень після останнього моменту спостереження сметрі t_k. Тому $c_0 + c_1 + \ldots + c_k = N - m$. Отже c_j–це число осіб, які випали зі спостереження між моментами t_j і t_{j+1} з причин інших ніж смерть. Будемо вважати, що всі c_j цензуровані спостереження попадають у вікритий інтервал (t_j, t_{j+1}). Якщо особа цензурується у той самий час, що і спостереження сметрі іншої особи, то вважаємо, що смерть настала раніше. Нехай $t_{j1}, t_{j2}, \ldots, t_{jc_j}$– це моменти часу(не обов'язково різні) із інтервалу (t_j, t_{j+1}), у які спостереження були цезуровані. Зручно позначити через n_j число осіб, які вижили у групі зирику до моменту t_j^-, тобто якраз до j-го моменту спостереження смерті.
Для одержання функції правдоподібності цих спостережень і не роблячи ніяких припущень стосовно виду функції розподілу $F(x)$, будемо вважати, що:

(a) ймовірність того, що смерть трапиться у момент часу t_j дорівнює $F(t_j) - F(t_j^-)$;

(b) ймовірність того, що особа доживе до моменту її цензурування t_{jl} дорівнює $1 - F(t_{jl})$ при неінформативному цензуруванні.

Сметрі та цензурування незалежні, тому загальна функція правдоподібності має вигляд

$$L = \prod_{j=1}^{k} (F(t_j) - F(t_j^-))^{d_j} \prod_{j=0}^{k} \prod_{l=1}^{c_j} (1 - F(t_{jl})).$$

Будемо шукати функцію $F(t)$, яка максимізує L, при єдиній умові, що вона є функцією розподілу. Оскільки функція розподілу є неспадна, а $t_{j1}, t_{j2}, \ldots, t_{jc_j} \in (t_j, t_{j+1})$, то кожен множник $1 - F(t_{jl})$ максимізується, якщо $F(t_{jl}) = F(t_j)$ для всіх t_{jl}. Таким чином $F(t) = F(t_j), \forall t \in [t_j, t_{j+1})$. Прі цьому для всіх $j : F(t_j) > F(t_j^-)$, бо інакше $L = 0$. Таким чином будь-яка оцінка максимальної правдоподібності для $F(t)$ буде східчастою функцією із стрибками у момент спостереження сметрі.

4.1.4 Тепер зручно поширити на дискретні розподіли означення сили смертності(або ризику), яке було раніше введено для неперервних розподілів. Припустимо, що $F(t)$– це дискретний розподіл із стрибками в точках $t_j, \ldots, t_k$. Покладемо

$$\lambda_j = P(T = t_j | T \geqslant t_j), \quad 1 \leqslant j \leqslant k.$$

Цю величину називають дискретною функцією ризику. Символ λ_j використовується, щоб не плутати зі звичайною силою смертності. Отримаємо

$$L = \prod_{j=1}^{k} (\lambda_j)^{d_j} (1 - \lambda_j)^{n_j - d_j}.$$

4.1.5Розглянувши $\ln L$ і дослідивши цю функцію на максимум одержимо оцінку максимальної правдоподібності

$$\hat{\lambda}_j = \frac{d_j}{n_j}, \quad 1 \leqslant j \leqslant k.$$

А використавши властивість інваріантності оцінок максимальної правдоподібності одержимо із $(*)$ оцінку

$$\hat{F}(t) = 1 - \prod_{t_j \leqslant t} (1 - \hat{\lambda}_j).$$

4.1.6 Одержана оцінка– це *оцінка Каплана-Мейера* або *оцінка добуток-границя*. Її можна розглядати з різних точок зору:

(a) Вивчаючи ймовірність сметрі протягом малих вікових інтервалів, ми можемо вибирати розбиття часової осі так як нам хочеться. Зручний вибір– мати дуже малі часові нтервали, які містять кожен момент t_j (достатньо малі, щоб виключити моменти цензурування t_{jl}) і довші часові інтервали, які містять тільки цензуровані спостереження. Єдина інформація, яку ми одержимо із останніх інтервалів, це те, що протягом них не було смертей. Тому можна вважати функцію $F(t)$ сталою на цих інтервалах. В той же час малі інтервали дають біноміальну оцінку тривалостей життів, що спостерігались.

(b) Як альтернативний підхід, ми можемо вибирати все більш тонке розбиття осі часу і оцінювати $1-F(t)$ як добуток ймовірностей виживання для кожного інтервалу. Тоді із даного означення дискретної сили смертності, ми одержимо оцінку Каплана-Мейера, якщо крок розбиття спрямувати до нуля. Це є причиною назви "оцінка добуток-границя яка іноді використовується.

4.1.7 Тільки ті особи із групи ризику, для яких проводілось спостереження за тривалістю життя $\{t_j\}$ дають внесок у оцінку. Тому не обов'язково починати спостереження всіх осіб в один і той же час або вік. Оцінка буде мати місце і для даних зрізаних зліва, прі умові, що зрізання є неінформативним, в тому сенсі, що початок спостереження у визначений час або вік не залежить вік залишку майбутнього життя.

4.2.Порівняння розподілів тривалості життя.

4.2.1 Оцінка Каплана-Мейера часто використовується для порівняння розподілів тривалості двох або більше популяцій, наприклад прі порівнянні методів лікування. Тому важливо знати її статистичні властивості. Має місце наближена формула (формула Грінвуда) для дисперсії $\tilde{F}(t)$:

$$\mathrm{Var}[\tilde{F}(t)] \approx (1-\hat{F}(t))^2 \sum_{t_j \leqslant t} \frac{d_j}{n_j(n_j-d_j)}$$

прийнятна для більшості t, але яка може применшувати дисперсію у хвості розподілу.

4.3.Оцінка Нельсона-Аалена.

4.3.1 Альтернативний непараметричний підхід – це *оцінка інтегрального ризику*

$$\Lambda = \int_0^t \mu_s ds + \sum_{t_j \leqslant t} \lambda_j ,$$

де інтеграл відповідає неперервній частині розподілу, а сума– дискретній частині. Оскільки методологія була розроблема статистиками, то широко використовується термін "інтегральний ризик в той час як термін "інтегральна сила смертності"майже не використовується.

4.3.2 Оцінка Нельсона-Аалена інтегрального ризику має вигляд

$$\hat{\lambda}_t = \sum_{t_j \leqslant t} \frac{d_j}{n_j}.$$

4.3.3 Оцінку Каплана-Мейера можна наблизити через $\hat{\lambda}_t$:

$$\hat{F}_t = 1 - \prod_{t_j \leqslant t} \Big(1 - \frac{d_j}{n_j}\Big) \approx 1 - \exp\Big\{ - \sum_{t_j \leqslant t} \frac{d_j}{n_j} \Big\} = 1 - \exp\{-\hat{\lambda}_t\}.$$

4.3.4 У відповідності до формули Грінвуда для дисперсії оцінки Каплана-Мейера, має місце формула для дисперсії оцінки Нельсона-Аалена:

$$\mathrm{Var}[\tilde{\lambda}_t] \approx \sum_{t_j \leqslant t} \frac{d_j(n_j - d_j)}{n_j^3}.$$

А

1. Протягом періоду з 1 січня 1994 року по 31 грудня 1996 року в тюрмі ведеться спостереження за людьми похилого віку. Реєструється тривалість знаходження під вартою (з точністю до кількості місяців) тих, хто помер за даний період, був звільнений з тюрми або все ще ув'язнений на 31 грудня 1996 року.

Зареєстровані дані (в місяцях) наступні:

$$\begin{array}{cccccc} 6^{\dagger} & 6 & 6 & 6 & 7 & 9^{\dagger} \\ 10^{\dagger} & 10 & 11^{\dagger} & 13 & 16 & 17^{\dagger} \\ 20 & 23^{\dagger} & & & & \end{array}$$

де † вказує на тих, хто був звільнений з тюрми або все ще під вартою на 31 грудня 1996 року.

(i) Вкажіть вид(и) цензурування, присутні в цих даних.

(ii) Знайдіть оцінку добуток-границя функції виживання (оцінка Каплана-Мейера), *S(t)*, де t - тривалість знаходження під вартою.

(iii) Вкажіть припущення, що лежать в основі оцінки (ii), та поясніть, як кожне з припущень може бути застосоване до цих даних.

2. Наступні дані відносяться до групи з 12 пацієнтів, яким була проведена операція з метою виправити стан, що загрожував життю, момент 0 є початком дослідження:

№ пацієнта	*Момент операції (в тижнях)*	*Момент закінчення спостереження (в тижнях)*	*Причина закінчення спостереження*
1	0	120	цензурування
2	0	68	смерть
3	0	40	смерть
4	4	120	цензурування
5	5	35	цензурування
6	10	40	смерть
7	20	120	цензурування
8	44	115	смерть
9	50	90	смерть
10	63	98	смерть
11	70	120	смерть
12	80	110	смерть

Ви можете зробити висновок, що цензурування було неіформативним у разі виживання пацієнта.

(i) Обчисліть оцінку Нельсона-Аалена інтегральної функції ризику, $\Lambda(t)$, де t - час, що пройшов з моменту операції.

(ii) Використовуючи результат частини (i), знайдіть оцінку функції виживання для пацієнтів, яким була проведена операція.

(iii) Оцініть ймовірність того, що пацієнт буде жити протягом не менш ніж 70 тижнів псля операції.

3. Компанія, що займається страхуванням життя, провела дослідження смертності на вибірці незалежних застрахованих віком від 40 до 45 років. Застраховані спостерігалися від моменту їх сорокового дня народження до моменту смерті, або до моменту вибуття з дослідження (живими), або до їх сорок п'ятого дня нарождення (який би з цих моментів не настав першим).

(i) Зазначте види цензурування, використані в дослідженні.

(ii) У таблиці нижче представлена частина отриманих під час дослідження даних, а саме результати спостереження 20 застахованих, що брали в ньому участь. Використовуючи ці дані, знайдіть оцінку Каплана-Мейера функції виживання. Визначте довірчий інтервал з рівнем довіри 95% для вашої оцінки.

№ пацієнта	*Вік останнього спостереження (роки і місяці)*	*Результат спостереження*
1	40 6	помер
2	40 6	вибув
3	41 0	помер
4	41 0	помер
5	41 6	вибув
6	42 3	помер
7	42 3	вибув
8	42 3	помер
9	43 0	вибув
10	43 3	вибув
11	43 3	помер
12	44 3	вибув
13	44 6	вибув
14	44 9	вибув
15	45 0	помер
16	45 0	вижив
17	45 0	вижив
18	45 0	вижив
19	45 0	вижив
20	45 0	вижив

(iii) Зобразіть на відповідним чином маркованому графіку отриману оцінку Каплана-Мейера та границі довірчого інтервалу з рівнем довіри 95%.

4. У клінічному дослідженні спостерігали 50 пацієнтів, що проходили лікування із застосуванням нових ліків. Наступні дані демонструють кількість повних місяців з моменту початку лікування до моменту закінчення дослідження для тих пацієнтів, які померли або вибули з дослідження до кінця дворічного періоду.

Смерті	6, 6, 12, 15, 20, 20, 23
Вибуття	1, 3, 5, 8, 10, 18

(i) Знайдіть оцінку Нельсона-Аалена інтегральної функції ризику, Λ_t.

(ii) Звідси, або інакшим способом, оцініть імовірність того, що пацієнт проживе що найменше 18 місяців після початку лікування.

5. Виробник комп'ютерних чіпів намагається оцінити робочу тривалість життя своєї продукції. Для цього досліджують 20 чіпів проданих з 1 січня 1997 року. З покупцем кожного чіпу зв'язуються кожні три місяці до 1 січня 2002 року та перевіряють, чи працює чіп належним чином. Результати представлені нижче (тут B означає кінець дослідження через некоректне функціонування чіпу, O - з іншої причини):

Дата закінчення дослідження	*Причина*
1 квітня 1997 року	В
1 липня 1997 року	В
1 жовтня 1997 року	О
1 жовтня 1997 року	О
1 жовтня 1997 року	О
1 січня 1998 року	В
1 січня 1998 року	В
1 липня 1998 року	О
1 липня 1998 року	О
1 жовтня 1998 року	О
1 липня 1999 року	О
1 липня 1999 року	О
1 липня 1999 року	О
1 жовтня 1999 року	В
1 квітня 2000 року	О
1 липня 2000 року	О
1 жовтня 2000 року	В
1 січня 2001 року	В
1 січня 2002 року	О
1 січня 2002 року	О

(i) Поясніть, як метод відстеження продукції даного виборника використовує цензурування у дослідженні.

(ii) Знайдіть оцінку Нельсона-Аалена інтегральної функції ризику для цих чіпів. Сформулюйте будь-які припущення, які ви можете зробити про точний час припинення функціонування кожного чіпу.

(iii) Використайте цей результат, щоб наблизити оцінку Каплана-Мейера функції виживання, та схематично зобразіть цю фукнцію.

(iv) Запропонуйте два аспекти, в яких рівень виходів з ладу компютерних чіпів має схожі риси з типовими моделями людської смертноті.

6. Протягом двох років випробування нового лікування велося спостереження 50 пацієнтів після того, як вони отримували цей тип лікування 1 липня 2000 року. Для тих пацієнтів, які померли або залишили дослідження до 30 червня 2002 року, був зареєстрований час, протягом якого вони брали участь у дослідженні. Деталі представлені нижче.

Час участі в дослідженні (у місяцях) для пацієнтів, які:

Померли	*Залишили дослідження*
4	2
6	5
6	7
8	9
11	10
16	13
22	15
	18

(i) Знайдіть оцінку Каплана-Мейера функції виживання, $S(t)$.

(ii) Схематично зобразіть ризик, $h(t)$, що випливає з оцінки Каплана-Мейера функції $S(t)$.

7. Було проведено випробування тривалості життя нових батарейок за допомогою розміщення 1000 батарейок у 1000 пухнастих електронних іграшках. Іграшки були включені та залишені працювати протягом 24 годин. Дослідник повертався кожну годину, щоб зафіксувати кількість іграшок, які припинили працювати. Після ретельного аналізу виявилося, що частина іграшок перестала працювати через механічні причини, а інша частина - не була увімкнена з самого початку. Деякі іграшки все ще працювали після закінчення 24 годин.

Поясніть присутність (або відсутність) наступних видів цензурування у даному дослідженні:

(а) цензурування зліва

(б) інтервальне цензурування

(в) цензурування типу I

(г) неінформативне цензурування

8. Основним видом діяльності певної благодійної організації є тренування собак-поводирів та забезпечення ними сліпих людей. Дана організація хотіла б оцінити тривалість життя собак-поводирів і для цього вирішила спостерігати 100 собак, що народилися 1990 року. З обраних 100 собак 81 собака все ще була на службі, досягнувши віку 10 років. Із собак, що залишилися, деякі померли, деякі залишили службу з інших причин.

Вік, у який собаки залишали службу через смерть або з інших причин, представлений у наступній таблиці:

Вік (роки)	*Кількість смертей*	*Кількість припинень служби з інших причин*
0.25		2
0.5	1	
1		3
2.25		3
3.5		2
4	2	
4.5		1
6	1	
7.5		2
9		2

(i) Знайдіть оцінку Нельсона-Аалена інтегрального ризику для собак-поводирів.

9. У клінічному дослідженні 100 пацієнтам була введена нова вакцина від тяжкої, але виліковної хвороби, потім за цими пацієнтами спостерігали впродовж 2 років. Були отримані наступні результати:

Вибули з дослідження: 13 пацієнтів

Захворіли під час дослідження: 8 пацієнтів

Не були хворі на момент закінчення дослідження: 79 пацієнтів

Час (у місяцях), коли пацієнти вибували з дослідження або починали хворіти:

Вибуття	1	3	3	4	4	5	7	7	9	10	11	13	15
Початок хвороби	2	2	2	6	6	8	8	17					

(i) На основі цих даних знайдіть оцінку Каплана-Мейера функції виживання.

10. Було проведено дослідження смертності 12 виведених в лабораторії комах. Вони спостерігалися з моменту народження до моменту смерті або до моменту закінчення дослідження; в останньому випадку комах, що були живими на цей момент, вважали відцензурованими.

Дана таблиця демонструє оцінку Каплана-Мейера функції виживання, побудовану за даними, отриманими під час дослідження:

t (*тижні*)	*S(t)*
$0 \leqslant t < 1$	1.0000
$1 \leqslant t < 3$	0.9167
$3 \leqslant t < 6$	0.7130
$6 \leqslant t$	0.4278

(i) Знайдіть кількість комах, що померли у віці від 3 до 6 тижнів.

(ii) Знайдіть кількість комах, що були відцензуровані.

5. Регресійна модель Кокса.

5. Регресійна модель Кокса

Література: [1,с.198-204]

Теоретичні відомості

5.1.Модель Кокса.

Модель Кокса пропонує таку форму функції ризиків для i-ї особи

$$\lambda(t; z_i) = \lambda_0(t) \exp(\beta z_i^T),$$

тут T означає транспортування вектора, і притримуючись усталених у статистиці позначень, ми позначаємо ризик через λ замість μ. β –це $1 \times p$ вектор *регресійних параметрів*. Оскільки βz_i^T – це скалярний добуток, то кожний фактор із z_i в функцію ризику входить мультиплікативно. $\lambda_0(t)$ називається *базовим ризиком*. У представленій простій моделі лише $\lambda_0(t)$ залежить від часу, але можна також використовувати коваріати залежні від часу.

У моделі Кокса функції ризику ризних осіб із векторами коваріатів z_1 і z_2 пропорційні протягом будь-якого часу

$$\frac{\lambda(t; z_1)}{\lambda(t; z_2)} = \frac{\exp(\beta z_1^T)}{\exp(\beta z_2^T)}$$

тому цю модель називають *моделлю пропорційних ризиків*. Для моделі із пропорціними ризиками можна взяти $\lambda(t; z_i) = \lambda_0(t)g(z_i)$, де $g(z_i)$ – це довільна функція z. Однак модель Кокса забеспечує додатність інтенсивності ризику і дає лінійну модель для логарифму ризику, що дуже зручно і в теорії і на практиці.

Зручність і корисність цієї моделі випливає із того, що загальний вигляд функції ризику всіх осіб визначається базовим ризиком $\lambda_0(t)$, в той час як експоненційна частина враховує відмінність між особами. Таким чином, якщо ми не цікавимося точною формою функції ризику, а більше цікавимося впливом коваріатів(факторів), то ми можемо не заважати на $\lambda_0(t)$ і оцінювати β на основі даних, не зважаючи на форму базового ризику.

Тому цей підхід нахивається *напів-параметричним*. Корисність і гнучкість моделі Кокса зробили її домінуючою в літературі з аналізу виживання. І ця модель є першою, до якої звертаються статистики при аналізі даних виживання.

5.2.Часткова функція правдоподібності.

Для того, щоб оцінити β потрібно максимізувати наступну часткову функцію правдоподібності. Нехай $R(t_j)$ позначає множину осіб, які знаходяться у групі ризику протягом спостереження за тривалістю життя j-ї особи. Припустимо, що відбувається тільки одна смерть у момент t_j, тобто $d_j = 1, (1 \leqslant j \leqslant k)$. Тому ймовірність того, що помре j-та особа за умови, що в момент t_j помре рівно одна особа із $R(t_j)$ дорівнює

$$\frac{\lambda(t_j; z_j)}{\sum\limits_{i \in R(T_j)} \lambda(t_j; z_i)} = \frac{\lambda_0(t_j) \exp(\beta z_j^T)}{\lambda_0(t_j) \sum\limits_{i \in R(t_j)} \exp(\beta z_i^T)} = \frac{\exp(\beta z_j^T)}{\sum\limits_{i \in R(t_j)} \exp(\beta z_i^T)}.$$

Ми тут використали співвідношення ${}_h q_t \approx \lambda(t)h$.
Таким чином часткова функція правдоподібності буде мати вигляд

$$L(\beta) = \prod_{j=1}^{k} \frac{\exp(\beta z_j^T)}{\sum\limits_{i \in R(t_j)} \exp(\beta z_i^T)}.$$

Ми бачимо, що часткова функція правдоподібності залежить лише від порядку, в якому спостерігається сметрь. Назва "часткова"функція правдоподібності виникає тому, що це частина повної функції правдоподібності, яка включає моменти часу в які спостерігалася смерть, а те, що спостерігалось між моментами сметрі, відкидається. Максимізація $L(\beta)$ проводиться чисельно, і більшість статистичних пакетів містять процедуру підгонки моделі Кокса.
На практиці можуть існувати зв'язки між даними, а саме

(a) $d_j = 1$;

(b) деякі спостереження цензуруються протягом періоду спостереження за тривалістю життя.

У випадку (b) вважають, що цензурування відбулось не в момент t_j, а зразу після нього, тому цензуровані в момент t_j особи включається до групи ризику $R(t_j)$. У випадку (a) обчислення L досить складні, бо потрібно перебрати всі можливі комбінації d_j сметрей із множини $R(t_j)$ – групи ризику в момент t_j. Тому використовують наближення Бреслова(Breslov)

$$L(\beta)=\prod_{j=1}^{k}\frac{\exp(\beta s_j^T)}{\left(\sum\limits_{i\in R(t_j)}\exp(\beta z_i^T)\right)^{d_j}},$$

де s_j– це сума векторів коваріат z всіх d_j осіб, які спостерігались до моменту їх сметрі в момент t_j.
Оцінка одержана із максимізації часткової функції правдоподібності має такі ж асимптотичні властивості, як звичайна оцінка максимальної правдоподібності: вона асимптотично нормальна і незміщена, її асимптотична кореляційна матриця оцінюється матрицею оберненою до інформаційної матриці.
Функція

$$u(\beta)=\left(\frac{\partial\ln L(\beta)}{\partial\beta_1},\ldots,\frac{\partial\ln L(\beta)}{\partial\beta_p}\right)$$

називається *функцією міри ефективності*. Розв'язок рівняння $u(\hat{\beta})=0$ дає оцінку максимальної правдоподібності $\hat{\beta}$.
Інформаційна матриця, що спостерігається, $I(\hat{\beta})$ має вигляд

$$\left\{I(\hat{\beta})_{ij}\right\}=\left\{-\frac{\partial^2\ln L(\beta)}{\partial\beta_i\partial\beta_j}\right\}\Bigg|_{\beta=\hat{\beta}},\quad 1\leqslant i,j\leqslant p.$$

отже, кореляційна матриця $C=\{\mathrm{cov}(\tilde{\beta}_i,\tilde{\beta}_j)\}$ асимптотично (при $n\to\infty, n$– об'єм вибірки) співпадає з $I^{-1}(\hat{\beta})$.
Корисною рисою більшості комп'ютерних пакетів для підгонки моделі Кокса є те, що інформаційна матриця обчислена для $\hat{\beta}$ зазвичай має вигляд добутку процесів, які підганяються (це використовується у алгоритмі Ньютона-Рафсона). Тому є доступними стандартні похибки компонент $\hat{\beta}$. Це буває корисним при підгонці конкретної моделі.

A

1. X є випадковою величиною, значення якої - час, що проходить від дати трансплантації нирки до смерті.

(i) Виразіть рівень ризику та інтегральну функцію ризику для тривалості часу від трансплантації до смерті *x* в термінах ймовірнстей виживання.

(ii) Нехай рівень ризику для тривалості часу від трансплантації до смерті *x* заданий наступним чином:

$$h(x)=\alpha\lambda x^{\alpha-1}$$

Знайдіть вираз для інтегрального ризику, *H(x)*.

(iii) Рівень ризику *h(x)* варіюється у пацієнтів із трансплантацією так, що

$$\alpha = \alpha_0 + \alpha_1 z_1$$

$$\lambda = \lambda_1 z_1 + \lambda_2 z_2$$

де $\alpha_0, \alpha_1, \lambda_1, \lambda_2$ константи; z_1 вік пацієнта на момент трансплантації і z_2 стать пацієнта: $z_2 = 0 = i$, $z_2 = 1 = i$.

Покажіть, що у загальному випадку ці ризики не є пропорційними, а якщо $\alpha_1 = 0$, то вони пропорційні.

2. Нехай за даними смертності власників полісу страхування життя було підібрано наступну регресійну модель пропорційних ризиків:

$$h_i(t) = h_0(t) exp\{0.01(x_i - 30) + 0.2y_i - 0.05z_i\},$$

де:
$h_i(t)$ - функція ризику клієнта *i* в момент *t*,
$h_0(t)$ - базова функція ризику в момент *t*,
x_i - вік клієнта *i* на момент придбання полісу,
$y_i = 1$, якщо клієнт *i* курить, 0 - якщо ні,
$z_i = 1$, якщо клієнт *i* жіночої статі, 0 - якщо чоловічої.

(i) Опишіть клас клієнтів, до яких застосовується базова функція ризику.

(ii) Що дана модель (якщо вона є вірною) може сказати про функцію виживання чоловіка, що палить, який придбав поліс у 30 років, по відношенню до функції виживання жінки, що палить, яка придбала поліс у 40 років?

(iii) Що дана модель (якщо вона є вірною) може сказати про функцію виживання жінки, що палить, яка придбала поліс у 30 років, по відношенню до функції виживання чоловіка, що не палить, який придбав поліс у 40 років?

3. Наступна таблиця містить дані про невелику вибірку працівників фабрики. У ній записана кількість місяців безперервної роботи до першої відсутності на робочому місці. Спостереження промарковані "+"демонструють кількість місяців безперервної роботи працівників, які пішли з фабрики, та за цей час жодного разу не пропустили роботу.

Працівники-чоловіки	6+	11	13+	15	16+	19+	20	
Працівники-жінки	2+	4	7	8+	10+	12+	17	21+

Потрібно підібрати модель пропорційних ризиків Кокса

$$\lambda(t|x) = \lambda_0(t) exp\{\beta x\},$$

де *t* - час до першої відсутності на роботі, $\lambda_0(t)$ - базовий ризик та *x*=0 для чоловіків, *x*=1 для жінок.

(i) Продемонструйте, що часткова логарифмічна правдоподібність може бути записана наступним чином:

$$l(\beta) = 3\beta - 4log_e(1 + e^{\beta}) - 2log_e(2 + e^{\beta}) + c,$$

де c є константою, яка не залежить від β.

(ii) Знайдіть оцінку часткової найбільної правдоподібності для β.

(iii) Знайдіть асимптотичну стандартну похибку цієї оцінки.

(iv) Перевірте гіпотезу про те, що у працівників жіночої статі частота перших відсутностей на роботі вища, ніж у працівників чоловічої статі. Поясніть кроки свого розв'язку та зробіть висновки.

4. Серед чоловіків було проведено дослідження факторів ризику смертності від раку легенів. Метою дослідження було встановити, чи є новий вид лікування ефективним для підтримки життя. Було сформовано дві групи пацієнтів. Перша група отримала "нове"лікування, друга - "загальноприйняте". Бралися до уваги також фактори загального стану здоров'я на момент діагностики (що реєструвалися як "здатний про себе піклуватися"або "не здатний про себе піклуватися") і тип пухлини ("велика", "плоскоклітинна", "мала", "аденоїдна").

Була оцінена модель Кокса пропорційних ризиків для ризику смерті. Наступна таблиця демонструє частину результатів:

Коваріат	*Параметр*	*Стандартна похибка*
Загальний стан здоров'я на момент дослідження		
Здатний про себе піклуватися	-0.60	0.05
Не здатний про себе піклуватися	0.00	
Лікування		
Нове	0.25	0.25
Існуюче	0.00	
Тип пухлини		
Велика	0.00	
Плоскоклітинна	-0.40	0.28
Мала	0.45	0.26
Аденоїдна	0.75	0.28

(i) Визначаючи всі терміни, які використовуються, запишіть загальний вираз для моделі пропорційних ризиків Кокса в термінах коваріатів, пов'язних з ними параметрів та базової функції ризику.

(ii) В контексті розглянутого вище дослідження, опишіть клас чоловіків, до яких застосовується базова функція ризику.

(iii) Порівняйте нове лікування з попереднім. Воно збільшує шанси на виживання, зменшує їх, чи зробити висновок про його ефективність неможливо? Поясніть свою відповідь.

(iv) Знадіть пропорцію, за якою ризик смерті чоловіка, що має пухлину типу "аденоїдна"та на момент дослідження "здатен про себе піклуватися", є більшим, ніж ризик смерті чоловіка з пухлиною типу "велика", який на момент дослідження "не здатен про себе піклуватися".

5. Потрібно застосувати модель пропорційних ризиків Кокса для моделювання частоти, з якою студенти лишають професію, не отримавши кваліфікації. Припускаючи, що вони залишаються в професії, студенти кваліфікуються через три роки після того, як до неї приєднались. Ризик у моделі залежить від часу, *t*, з моменту приднання до професії та трьох коваріатів. У таблиці нижче представлені коваріати, їх категорії та підібрані параметри:

Коваріат	*Ймовірність*	*Параметр*
Роботодавець	великий	0
	малий	0.4
Напрямок навчання	без напрямку	0.3
	наука	-0.1
	мистецтво	0.2
	інше	0
Місцезнаходження	Лондон	0
	інші регіони Великої Британії	-0.3
	зарубіж	0.4

(i) Визначаючи всі терміни, які використовуєте, запишіть вираз для функції ризику даної моделі.

(ii) Опишіть клас студентів, які швидше за все кваліфікуються, та клас студентів, чия кваліфікація найменш імовірна.

(iii) Студент, що пробув у професії рік, перейшов від "малого"роботодавця до "великого". Знайдіть імовірність того, що цей студент кваліфікується у великого роботодавця, P_L, в термінах імовірності того, що він кваліфікувався у малого роботодавця, P_S, якщо б залишився у нього, за інших однакових умов.

6. Виконайте наступні завдання:

(i) Випадкова величина *X* - час до того, як відбудеться деяка подія. Запишіть функцію ризику *h(t)* того, що подія відбудеться у момент часу *t*, в термінах імовірностей, пов'язаних з випадковою величиною *X*, та поясніть означення цієї функції.

(ii) Дослідіть ефект двох факторів, Z_1 та Z_2, на час до моменту, коли відбувається певна подія. Хтось пропонує використати модель пропорційних ризиків. Поясніть, що розуміють під терміном “модель пропорційних ризиків”.

(iii) Розподіл Вейбула має функцію виживання, що задається формулою:

$$S(t) = exp\{(-\lambda t)^{\alpha}\},$$

де α і λ є параметрами.

Покажіть, що обираючи λ, яке залежить тільки від Z_1 та Z_2, розподіл Вейбула може бути використаний як модель пропорційних ризиків для дослідження ефекту Z_1 та Z_2 на час до моменту, коли відбувається певна подія.

7. Виконайте наступні завдання:

(i) Запишіть рівняння моделі пропорційних ризиків Кокса, в якій функція ризику залежить від часу *t* та вектору коваріатів *z*. Визначте всі інші терміни, які будете використовувати.

(ii) Поясніть, чому модель Кокса іноді описують як “напівпараметричну”.