Статистичні методи теорії ризику

© Ямненко Р.Є

КНУ імені Тараса Шевченка
механіко-математичний факультет
магістратура "Актуарна та фінансова математика"

I семестр 2012

Модель колективного ризику. Розподіл сумарної величини виплат за портфелем та його характеристики

Розглянемо сумарну величину S позовів $X_1, \ldots, X_N$, які виникли за певний короткий проміжок часу у деякому страховому портфелі. Будемо припускати, що:

- кількість позовів не впливає на величини індивідуальних позовів;
- на величину окремого позову ніяк не впливають величини інших позовів;
- розподіл величини індивідуальних позовів не змінюється протягом терміну дії поліса.

Тобто $S = X_1 + \cdots + X_N$, де X_i – н.о.р.в.в. та N не залежить від $\{X_i\}_{i \geq 1}$, зокрема $S = 0$, коли $N = 0$.

Позначимо через $G(x)$ та $F(x)$ функції розподілу в.в. S та X_i відповідно.

$$G(x) = \mathrm{P}(S < x), \quad F(x) = \mathrm{P}(X_i < x),\ 1 \leq i \leq N.$$

Для зручності часто припускатимемо існування щільності у $F(x)$, яку позначатимемо $f(x)$.

Нехай також $m_k = \mathrm{E}X_i^k$ та $\mu_k = \mathrm{E}(X_i - m_1)^k$ – відповідно k-й нецентральний та центральний моменти X_i. Крім того, вважатимемо, що величина позовів не може бути від'ємною, тобто $F(0) = \mathrm{P}(X_j < 0) = 0$.

Отримаємо загальні формули для функції розподілу, математичного сподівання, дисперсії та генератриси моментів в.в. S.

Функція розподілу

Розглянемо подію $\{S < x\}$ для довільного $x \geq 0$. Вона має місце, якщо відбувається лише одна з таких несумісних подій:

- $\{S < x$ та $N = 0\}$, що означає відсутність позовів;
- $\{S < x$ та $N = 1\}$, тобто надійшов один позов, величина якого менша x;
- $\{S < x$ та $N = 2\}$, надійшли два позови, сумарна величина яких менша x;

  $\vdots$
- $\{S < x$ та $N = n\}$, надійшли n позовів, сумарна величина яких менша x;

  $\vdots$

$$\{S < x\} = \bigcup_{n=0}^{\infty} \{S < x, N = n\},$$

$$G(x) = \mathrm{P}(S < x) = \sum_{n=0}^{\infty} \mathrm{P}(S < x, N = n)$$

$$= \sum_{n=0}^{\infty} \mathrm{P}(S < x | N = n) \mathrm{P}(N = n) = \sum_{n=0}^{\infty} F^{n*}(x) \mathrm{P}(N = n),$$

оскільки при $N = n$ сумарні виплати S є сумою фіксованої кількості n в.в. $\{X_i\}_{i=1}^{n}$, а тому

$$\mathrm{P}(S < x | N = n) = F^{n*}(x),$$

де $F^{n*}(x)$ є n-кратною згорткою функції розподілу $F(x)$.

Загальний вираз для ф-ї розподілу S отримано без жодних припущень щодо розподілів в.в. N та X_i.

Моменти

Для знаходження моментів в.в. S використаємо властивості умовних математичних сподівань. Зокрема, щоб визначити $\mathrm{E}S$, використаємо формулу повного математичного сподівання $\mathrm{E}S = \mathrm{E}(\mathrm{E}(S|N))$. Підрахуємо спочатку

$$\mathrm{E}(S|N = n) = \sum_{i=1}^{n} \mathrm{E}X_i = nm_1.$$

Отже,

$$\mathrm{E}(S|N) = Nm_1,$$

тоді

$$\mathrm{E}S = \mathrm{E}(Nm_1) = m_1\mathrm{E}N = \mathrm{E}X_i\mathrm{E}N. \tag{1}$$

Таким чином, математичне сподівання сумарних виплат дорівнює добутку середньої очікуваної кількості позовів та математичного сподівання величини індивідуального позову.

Щоб знайти вираз для $\mathsf{D}S$, застосуємо формулу повної дисперсії $\mathsf{D}S = \mathsf{E}(\mathsf{D}(S|N)) + \mathsf{D}(\mathsf{E}(S|N))$. Тоді, пам'ятаючи, що величини X_i позовів незалежні, маємо

$$D(S|N = n) = \mathsf{D}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathsf{D}X_i = n\mu_2.$$

Звідси маємо, що

$$\mathsf{D}(S|N) = N\mu_2.$$

Отже,

$$\mathsf{D}S = \mathsf{E}(N\mu_2) + \mathsf{D}(Nm_1) = \mu_2\mathsf{E}N + m_1^2\mathsf{D}N = \mathsf{D}X_i\mathsf{E}N + \mathsf{E}X_i^2\mathsf{D}N, \tag{2}$$

тобто дисперсія S визначається через перші два моменти в.в. N та X_i.

Генератриса розподілу

Нехай $G_N(t)$ та $G_X(t)$ – генератриси розподілів в.в. N та X_i відповідно. Визначимо вигляд генератриси розподілу S. За формулою повної ймовірності маємо

$$
\begin{aligned}
G_S(t) &= \mathrm{E}t^S = \sum_{n=0}^{\infty} \mathrm{E}(t^S | N = n)\mathrm{P}(N = n) \\
&= \sum_{n=0}^{\infty} \mathrm{E}(t^{X_1+\cdots+X_n})\mathrm{P}(N = n) = \sum_{n=0}^{\infty} \prod_{i=1}^{n} \mathrm{E}(t^{X_i})\mathrm{P}(N = n) \\
&= \sum_{n=0}^{\infty} (G_X(t))^n \mathrm{P}(N = n) = G_N(G_X(t)).
\end{aligned}
$$

Генератриса моментів

За формулою повного мат. сподівання

$$M_S(t) = \mathrm{E}\exp\{tS\} = \mathrm{E}(\mathrm{E}(\exp\{tS\}|N)).$$

Нехай $M_X(t)$ – генератриса моментів в.в. X_i. З того, що $\{X_i\}$ – незалежні в.в., випливає, що

$$\mathrm{E}(\exp\{tS\}|N=n) = \mathrm{E}\exp\left\{t\sum_{i=1}^{n} X_i\right\} = \mathrm{E}\prod_{i=1}^{n}\exp\{tX_i\}$$

$$= \prod_{i=1}^{n}\mathrm{E}\exp\{tX_i\} = \prod_{i=1}^{n} M_X(t) = (M_X(t))^n.$$

Таким чином,

$$\mathrm{E}(\exp\{tS\}|N) = (M_X(t))^N$$

$$M_S(t) = \mathrm{E}(M_X(t))^N = \mathrm{E}\exp\{N\ln M_X(t)\} = M_N(\ln M_X(t)). \quad (3)$$

Складний пуассонівський розподіл

Розглянемо сумарну величину позовів S у випадку, коли їхня кількість N має розподіл Пуассона з параметром $\lambda > 0$. Тоді в.в. матиме складний розподіл Пуассона з параметрами λ і $F(x)$, де $F(x)$ – функція розподілу величини індивідуального позову X_i.

Нагадаємо, що

$$\mathrm{E}N = \mathrm{D}N = \lambda, \quad M_N(t) = \exp\{\lambda e^t - 1\}.$$

Тоді згідно з формулами (1) – (3), маємо

$$\mathrm{E}S = \lambda m_1, \quad \mathrm{D}S = \lambda m_2, \quad M_S(t) = \exp\{\lambda(M_X(t) - 1)\}, \quad (4)$$

де $m_k = \mathrm{E}X_i^k$, $k = 1, 2, \ldots$, $M_X(t)$ – генератриса моментів X_i.

Важливою особливістю складного пуассонівського розподілу є той факт, що сума незалежних в.в., які мають складний розподіл Пуассона, також розподілена за складним пуассонівським законом. Сформулюємо цю властивість у вигляді наступної теореми.

Теорема

Нехай $S_1, \dots, S_n$ – незалежні в.в., причому S_i має складний розподіл Пуассона з параметрами $\lambda_i > 0$ та $F_i(x)$. Позначимо сума $S = S_1 + \cdots + S_n$ має складний розподіл Пуассона з параметрами

$$\lambda = \sum_{i=1}^{n} \lambda_i \quad \textit{та} \quad F(x) = \frac{1}{\lambda} \sum_{i=1}^{n} \lambda_i F_i(x).$$

Доведення.

Оскільки $\sum_{i=1}^{n} \frac{\lambda_i}{\lambda} = 1$, то $F(x)$ є зваженим середнім функцій розподілу $F_i(x)$, а значить, також є функцією розподілу.
Нехай $M(t)$ – генератриса моментів в.в. із функцією розподілу $F(x)$. Тоді

$$M(t) = \int_0^{\infty} e^{tx} \mathrm{d}F(x)$$

$$= \frac{1}{\lambda} \sum_{i=1}^{n} \int_0^{\infty} e^{tx} \mathrm{d}F_i(x) = \frac{1}{\lambda} \sum_{i=1}^{n} M_i(t).$$

Розглянемо тепер генератрису моментів S. Використовуючи незалежність S_i, маємо □

$$
\begin{aligned}
M_S(t) &= \mathsf{E}e^{tS} = \mathsf{E}e^{t(S_1+\cdots+S_n)} = \mathsf{E}\prod_{i=1}^{n} e^{tS_i} \\
&= \prod_{i=1}^{n} \mathsf{E}e^{tS_i} = \prod_{i=1}^{n} \exp\{\lambda_i(M_i(t)-1)\} \\
&= \exp\left\{\sum_{i=1}^{n} \lambda_i(M_i(t)-1)\right\} \\
&= \exp\left\{\lambda \sum_{i=1}^{n} \frac{\lambda_i}{\lambda}(M_i(t)-1)\right\} \\
&= \exp\left\{\lambda \left(\sum_{i=1}^{n} \frac{\lambda_i}{\lambda} M_i(t) - 1\right)\right\},
\end{aligned}
$$

тобто

$$M_S(t) = \exp\{\lambda(M(t) - 1)\}.$$

Внаслідок взаємно однозначної відповідності між функціями розподілу і генератрисами моментів в.в. з останньої формули випливає, що S має складний розподіл Пуассона з параметром λ та $F(x)$.

Складний біноміальний розподіл

За певних умов природним вибором в якості розподілу для кількості позовів N виникає біноміальний розподіл. Наприклад, у полісі групового страхування життя для групи з n осіб розподілом кількості смертей протягом року є біноміальний, якщо припускати, що всі застраховані особи незалежні між собою відносно смерті та мають однакову інтенсивність смертності.

Нагадаємо, що для в.в. N, яка має біномільний розподіл з параметрами n та $0 < \theta < 1$,

$$\mathsf{E}N = n\theta, \quad \mathsf{D}N = n\theta(1-\theta), \quad M_N(t) = (\theta e^t + 1 - \theta)^n.$$

Коли N має біноміальний розподіл, то S – складний біноміальний розподіл. Важливою умовою для вибору біноміального розподілу для N є те, що існує верхня межа n для кількості позовів.

Згідно з формулами (1) – (3), маємо

$$\mathsf{E}S = n\theta m_1, \quad \mathsf{D}S = n\theta m_2 - \theta^2 m_1^2, \tag{5}$$

$$M_S(t) = (\theta M_X(t) + 1 - \theta)^n.$$

Складний від'ємний біноміальний розподіл

Нехай в.в. N має складний від'ємний біноміальний розподіл з параметрами k та $0 < \theta < 1$:

$$\mathsf{P}(N = n) = C_{k+n-1}^{n}\theta^{k}(1-\theta)^{n}, \quad n = 0, 1, 2, \ldots$$

Тоді

$$\mathsf{E}N = \frac{k(1-\theta)}{\theta}, \quad \mathsf{D}N = \frac{k(1-\theta)}{\theta^2}, \quad M_N(t) = \theta^k(1-(1-\theta)e^t)^{-k}.$$

Зокрема, у частковому випадку при $k = 1$ в.в. N має геометричний розподіл.

Від'ємний біноміальний розподіл виступає альтернативою розподілу Пуассона для N.

Однією з переваг від'ємного біноміального розподілу над пуассонівським є те, що його дисперсія перевищує математичне сподівання.
У розподілі Пуассона ці дві величини однакові.

Відповідно до цього від'ємний біноміальний розподіл може краще відповідати множині даних, у якій вибіркова дисперсія більша за вибіркове середнє значення.

Отже, коли кількість позовів N розподілена за від'ємним біноміальним законом, то величина S сумарних виплат має складний від'ємний біноміальний розподіл.

Знову використовуючи формули (1) – (3), отримуємо вирази для математичного сподівання, дисперсії та генератриси моментів S:

$$\mathsf{E}N = \frac{k(1-\theta)}{\theta} m_1,$$

$$\mathsf{D}N = \frac{k(1-\theta)}{\theta} m_2 + \frac{k(1-\theta)^2}{\theta^2} m_1^2,$$

$$M_N(t) = \theta^k (1-(1-\theta)M_X(t))^{-k}.$$

Вправа

Сумарну величину позовів S за портфелем моделюють як складний розподіл $S = X_1 + X_2 + \cdots + X_N$, де в.в. N – кількість позовів – має розподіл Пуассона з середнім λ, X_i – величина i-го позову, та X_i – н.о.р. в.в. та незалежні від N. Нехай $M_X(t)$ позначає генератрису моментів X_i.

i) Покажіть, використовуючи умовне математичне сподівання, що генератриса кумулянт S має вигляд $C_S(t) = \lambda(M_X(t) - 1)$.

ii) Обчисліть дисперсію S, коли $\lambda = 20$, а X має середнє 20 й дисперсію 10.

Вправа

Ймовірнісний розподіл загальної кількості позовів N, що виникають у деякій групі полісів протягом року, має такий вигляд:

$$P(N=0)=0,7;\ P(N=1)=0,15;$$

$$P(N=2)=0,1;\ P(N=3)=0,05.$$

Величина кожного позову X (в $\$1000$) має гамма-розподіл з параметрами $\alpha=2$ та $\lambda=0,1$ незалежно від величин інших позовів та кількості.

Обчисліть очікуване середнє значення та стандартний відхил сумарної величини позовів за один рік.

Вправа

Припускають, що кількість позовів N, що надходять протягом п'ятирічного періоду від вказаного полісу, має від'ємний біноміальний розподіл другого типу з середнім $\mathrm{E}N = \frac{k(1-p)}{p}$ та дисперсією $\mathrm{D}N = \frac{k(1-p)}{p^2}$. Вважають, що величина кожного позову X (в \$1000) має експоненційний розподіл з параметром λ незалежно від величин інших позовів та кількості позовів. Нехай S є сумарною величиною позовів за період 5 років для $k = 2$, $p = 0,8$ та $\lambda = 2$.

i) Обчисліть мат. сподівання та стандартний відхил S.

Припустимо тепер, що N має розподіл Пуассона з параметром $\mu = 0,5$, тобто з таким самим середнім, що й N вище, а X має гамма-розподіл із параметрами $\alpha = 2$ та $\lambda = 4$, тобто теж з тим самим середнім.

i) Обчисліть математичне сподівання та стандартний відхил S за нових припущень.
ii) Порівняйте два набори відповідей.

Вправа

Нехай в.в. N описує кількість позовів, які надходять до страховика протягом одного місяця від одного з його портфелів. Припустимо, що N має розподіл Пуассона з $\mathrm{E}N = 50$. Величина X_i кожного позову нормально розподілена з середнім $\mu = 1000$ та дисперсією $\sigma^2 = 200^2$. Сумарна величина позовів за один місяць дорівнює $S = X_1 + \cdots + X_N$. Припустимо, що в.в. $N, X_1, X_2, \ldots$ незалежні між собою.

i) Зазначте тип розподілу в.в. S.

ii) Обчисліть мат. сподівання та стандартний відхил S.

Модель індивідуального ризику

Модель індивідуального ризику відображає сумарні втрати у вигляді фіксованої суми незалежних (але не обов'язково однаково розподілених) в.в.:

$$S = X_1 + \cdots + X_n,$$

де X_j – величина позову, який виник за j-им ризиком, а n дорівнює загальній кількості ризиків.

Зауважимо, що деякі з ризиків можуть і не призвести до позовів, тобто деякі зі спостережених значень $\{X_j\}_{j=1}^{n}$ можуть дорівнювати нулю.

Таку модель зазвичай використовують під час дослідження сумарних втрат від n страхових угод, наприклад, для полісу групового страхування життя n осіб.

Отже, стосовно кожного ризику роблять такі припущення:

i) кількість позовів N_j від j-го ризику дорівнює 0 або 1;
ii) імовірність позову від j-го ризику дорівнює q_j, тобто
$$\mathrm{P}(N_j = 1) = q_j.$$

Якщо позов спричинений j-им ризиком, то величину збитків позначимо через Y_j. Нехай також $\mathrm{E}Y_j = m_{1j}$, $\mathrm{D}Y_j = \mu_{2j}$.

Припущення i) звужує сферу застосування моделі.

Зокрема, з нього випливає, що від кожного ризику може виникнути не більше одного позову.

Такий підхід охоплює договори страхування життя, у якому ймовірність смерті j-ої особи протягом року дорівнює q_j, але усуває багато типів договорів загального страхування. Наприклад, немає жодних обмежень на кількість позовів, які можуть протягом року виникнути за полісом страхування домашнього майна.

Наголосимо, які відмінності від моделі колективного ризику має модель індивідуального ризику:

- кількість ризиків у портфелі є фіксованою. У моделі колективного ризику немає необхідності точно визначати цю кількість чи вважати, що вона незмінна протягом терміну дії страхового покриття;
- кількість позовів за кожним індивідуальним полісом обмежена на відміну від моделі колективного ризику – вона таких обмежень не має;
- припускають, що індивідуальні ризики незалежні. У моделі колективного ризику незалежними є величини індивідуальних позовів.

З припущень i)-ii) видно, що в.в. N_j мають біноміальний розподіл із параметрами $n = 1$ та q_j. Тому розподіл X_j є складним біноміальним з індивідуальними позовами, розподіленими, як Y_j. З (5) випливає, що

$$\mathsf{E}X_j = q_j m_{1j}, \quad \mathsf{D}X_j = q_j \mu_{2j} + q_j(1 - q_j) m_{1j}^2.$$

Як бачимо, в.в. S є сумою n незалежних (можливо, не однаково розподілених) в.в. зі складним біноміальним розподілом. Розподіл S є згорткою n функцій розподілу в.в. X_j, який можна просто обчислити лише у випадку незалежних однаково розподілених складних біноміальних в.в. Проте легко знайти середнє та дисперсію S:

$$\mathrm{E}S = \mathrm{E}\sum_{j=1}^{n} X_j = \sum_{j=1}^{n} \mathrm{E}X_j = \sum_{j=1}^{n} q_j m_{1j},$$

$$\mathrm{D}S = \mathrm{D}\sum_{j=1}^{n} X_j = \sum_{j=1}^{n} \mathrm{D}X_j = \sum_{j=1}^{n} \left(q_j\mu_{2j} + q_j(1-q_j)m_{1j}^2\right).$$

В окремому випадку, коли $\{X_j\}$ є послідовністю н.о.р. в.в., тоді для всіх полісів значення q_j, m_{1j} та μ_{2j} однакові і рівні q, m_1 та μ_2 відповідно.

В.в. S матиме складний біноміальний розподіл з параметрами n, q та індивідуальними позовами з тією самою функцією розподілу $F(x)$. Таким чином, цей випадок зводиться до моделі колективного ризику.

Легко бачити, що

$$\mathrm{E}S = nqm_1, \quad \mathrm{D}S = nq\mu_2 + nq(1-q)m_1^2,$$

що повністю узгоджується з (5).

Мінливість чи невизначеність параметрів

Досі при вивченні моделей припускали, що розподіли кількості позовів та величини індивідуальних позовів повністю відомі.

Однак, вони можуть бути невідомими, і їх треба оцінювати за спостереженими даними. Спробуємо узагальнити введені раніше моделі, щоб охопити випадок невизначеності чи мінливості параметрів.

Для цього розглянемо кілька прикладів. Більшість із них стосується невизначеності у розподілах кількості позовів, бо саме цій проблемі приділяють більше уваги в актуарній літературі.

Мінливість у неоднорідному портфелі

Розглянемо портфель, що складається з n незалежних полісів. Сумарні позови від i-го поліса позначимо через S_i, де в.в. S_i має складний розподіл Пуассона з параметрами λ_i та $F(x)$.

Для простоти розподіл величин індивідуальних позовів $F(x)$ вважатимемо однаковим для всіх полісів. Тут розподіл $F(x)$ є відомим, але невідомими є значення пуассонівських параметрів λ_i, $i = 1, 2, \ldots, n$.

Також припускатимемо, що λ_i є вибірковими значеннями незалежних в.в. з однаковим розподілом.

Інакше кажучи, $\{\lambda_i\}$ будемо трактувати як множину н.о.р. в.в. із відомим розподілом. Це означає, що коли поліс вибрано навмання з портфелю, то значення його пуассонівського параметра невідоме, але стосовно нього можна сформулювати певні ймовірнісні твердження, наприклад, про довірчі інтервали.

Важливо розуміти, що пуассонівський параметр для поліса, вибраного з портфелю, є фіксованою величиною. Проблема полягає в тому, що ці величина невідома.

Приклад

Припустимо, що параметр Пуассона для окремих полісів вибирають із гамма-розподілу з параметрами α й δ. Треба знайти розподіл кількості позовів від поліса, вибраного навмання з портфелю.

Позначимо через N_i кількість позовів від i-го поліса із заданого портфелю та через λ_i – відповідний пуассонівський параметр. Тоді N_i має розподіл Пуассона з параметром λ_i, але проблема в тому, що значення λ_i невідоме. Відомий лише розподіл, з якого вибирають λ_i.

Знайдемо безумовний розподіл N_i.

$$f_{\lambda_i}(\nu) = \frac{\delta^\alpha}{\Gamma(\alpha)}\nu^{\alpha-1}e^{-\delta\nu}, \quad \nu > 0,$$

$$p_{N_i|\lambda_i=\nu}(x,\nu) = \mathsf{P}(N_i = x|\lambda_i = \nu) = \frac{\nu^x}{x!}e^{-\nu}, \quad x = 0, 1, 2, \ldots$$

Використаємо аналог формули повної ймовірності. Маємо

$$\mathsf{P}(N_i = x) = \int_0^\infty p_{N_i|\lambda_i=\nu}(x,\nu) f_{\lambda_i}(\nu)\mathrm{d}\nu = \int_0^\infty \frac{\nu^x}{x!}e^{-\nu}\frac{\delta^\alpha}{\Gamma(\alpha)}\nu^{\alpha-1}e^{-\delta\nu}\mathrm{d}\nu$$

$$= \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)x!}\frac{\delta^\alpha}{(\delta+1)^{\alpha+x}}.$$

Звідси видно, що маргінальний розподіл N_i є від'ємним біноміальним розподілом з параметрами α та $\delta/(\delta+1)$.

Приклад

Припустимо, що пуассонівські параметри полісів точно невідомі, але з однаковою ймовірністю можуть набувати значення 0,1 або 0,3. Треба знайти:

а) середнє та дисперсію (у термінах m_1 та m_2) сумарних позовів від поліса, вибраного навмання з портфелю;

б) середнє та дисперсію (у термінах m_1, m_2 та n) сумарних позовів від усього портфелю в цілому.

Нехай λ_i, $i = 1, 2, \dots, n$ – значення пуассонівського параметра i-го поліса, де $\{\lambda_i\}$ – н.о.р. в.в., кожна з яких має розподіл $\mathsf{P}(\lambda_i = 0,1) = \mathsf{P}(\lambda_i = 0,3) = 0,5$. Звідси

$$\mathsf{E}\lambda_i = 0,1 \cdot 0,5 + 0,3 \cdot 0,5 = 0,2;$$

$$\mathsf{D}\lambda_i = \mathsf{E}\lambda_i^2 - (\mathsf{E}\lambda_i)^2 = (0,1)^2 \cdot 0,5 + (0,3)^2 \cdot 0,5 - (0,2)^2 = 0,01.$$

Оскільки в.в. $S_i|\lambda_i$ мають складний розподіл Пуассона, то з (4) випливає, що

$$\begin{aligned}
\mathsf{E}S_i &= \mathsf{E}(\mathsf{E}(S_i|\lambda_i)) = \mathsf{E}(\lambda_i m_1) = m_1 \mathsf{E}\lambda_i = 0,2m_1;\\
\mathsf{D}S_i &= \mathsf{E}(\mathsf{D}(S_i|\lambda_i)) + \mathsf{D}(\mathsf{E}(S_i|\lambda_i)) = \mathsf{E}(\lambda_i m_2) + \mathsf{D}(\lambda_i m_1)\\
&= m_2 \mathsf{E}\lambda_i + m_1^2 \mathsf{D}\lambda_i = 0,2m_2 + 0,01m_1^2.
\end{aligned}$$

б) В.в. S_i є н.о.р. в.в., кожна з яких має розподіл, наведений у п. а). Отже,

$$\mathrm{E}\left(\sum_{i=1}^{n} S_i\right) = n\mathrm{E}S_i = 0,2m_1 n;$$

$$\mathrm{D}\left(\sum_{i=1}^{n} S_i\right) = \sum_{i=1}^{n} \mathrm{D}S_i = n\mathrm{D}S_i = (0,2m_2 + 0,01m_1^2)n.$$

Мінливість в однорідному портфелі

Розглянемо портфель із n полісів. Припустимо, що сумарні позови від одного поліса мають складний розподіл Пуассона з параметрами λ та $F(x)$.

Ці параметри є однаковими для всіх полісів у портфелі.

Якщо значення λ відоме, то сумарні позови від різних полісів будуть незалежними.

Вважатимемо, що значення λ невідоме (наприклад, воно міняється щороку), проте є певні міркування щодо ймовірності значень, які може набувати λ.

Як і у попередньому параграфі для простоти припустимо, що немає невизначеності стосовно моментів або розподілу величин індивідуальних позовів, тобто відносно $F(x)$.

Невизначеність щодо значень λ можна змоделювати, вважаючи цей параметр в.в. з відомим розподілом.

Приклад

Припустимо, що параметр λ розподілу Пуассона з однаковою ймовірністю набуває значення 0,1 або 0,3. Треба знайти:
а) середнє та дисперсію (у термінах m_1 та m_2) сумарних позовів від поліса, вибраного навмання з даного портфелю;
б) середнє та дисперсію (у термінах m_1, m_2 та n) сумарних позовів від усього портфелю.

Як і раніше, через S_i позначимо сумарні позови від i-го поліса. У нашій ситуації в.в. $\{S_i|\lambda\}$ є н.о.р. в.в., кожна з яких розподілена за складним пуассонівським законом з параметрами λ та $F(x)$. В.в. λ має розподіл $\mathsf{P}(\lambda_i = 0,1) = \mathsf{P}(\lambda_i = 0,3) = 0,5$. Звідси

$$\mathsf{E}\lambda_i = 0,1 \cdot 0,5 + 0,3 \cdot 0,5 = 0,2;$$

$$\mathsf{D}\lambda_i = \mathsf{E}\lambda_i^2 - (\mathsf{E}\lambda_i)^2 = (0,1)^2 \cdot 0,5 + (0,3)^2 \cdot 0,5 - (0,2)^2 = 0,01.$$

$$
\begin{aligned}
a)\ \mathsf{E}S_i &= \mathsf{E}(\mathsf{E}(S_i|\lambda)) = \mathsf{E}(\lambda m_1) = m_1\mathsf{E}\lambda = 0,2m_1;\\
\mathsf{D}S_i &= \mathsf{E}(\mathsf{D}(S_i|\lambda)) + \mathsf{D}(\mathsf{E}(S_i|\lambda)) = \mathsf{E}(\lambda m_2) + \mathsf{D}(\lambda m_1)\\
&= m_2\mathsf{E}\lambda + m_1^2\mathsf{D}\lambda = 0,2m_2 + 0,01m_1^2.
\end{aligned}
$$

б) Оскільки в.в. S_i є однаково розподілені, то

$$
\begin{aligned}
\mathsf{E}\left(\sum_{i=1}^{n} S_i\right) &= n\mathsf{E}S_i = 0,2m_1 n;\\
\mathsf{D}\left(\sum_{i=1}^{n} S_i\right) &= \mathsf{E}\left(\mathsf{D}\left(\sum_{i=1}^{n} S_i\Big|\lambda\right)\right) + \mathsf{D}\left(\mathsf{E}\left(\sum_{i=1}^{n} S_i\Big|\lambda\right)\right)\\
&= \mathsf{E}(n\lambda m_2) + \mathsf{D}(n\lambda m_1)\\
&= nm_2\mathsf{E}\lambda + n^2 m_1^2\mathsf{D}\lambda = 0,2m_2 n + 0,01m_1^2 n^2.
\end{aligned}
$$

Значення середніх в обох прикладах однакові, як і дисперсії для окремих полісів (п. а)). Різниця виникає, коли розглядають дисперсії більш ніж одного поліса (п. б)), де більша дисперсія виникає у прикладі 2.

Практична ситуація, до якої підходить останній приклад, стосується портфелю полісів страхування будівель у деякому районі. Кількість позовів може залежати, крім інших факторів, від погоди протягом року – незвично велика кількість штормів може призвести до великої очікуваної кількості позовів (тобто до великого значення λ) для усіх полісів одночасно, і навпаки, за гарної погоди значення λ будуть малими.

Мінливість у кількості й величині позовів та невизначеність у параметрах

Розглянемо два приклади. Перший з них містить невизначеність як у величинах позовів, так і в їх кількості.

Приклад

Страхова компанія моделює позови, викликані ураганами, при страхуванні домогосподарств за таких припущень:
i) кожного року кількість K ураганів має розподіл Пуассона з параметром λ;
ii) кількість позовів N_i, викликаних i-им ураганом, має розподіл Пуассона з параметром Θ_i, $i = 1, \ldots, K$;
iii) параметри Θ_i – це н.о.р. в.в. такі, що $\mathrm{E}\Theta_i = n$, $\mathrm{D}\Theta_i = s_i^2$, $i = 1, \ldots, K$;

iv) величина j-го позову, викликаного i-им ураганом, тобто X_{ij}, $j = 1, \ldots, N_i$, $i = 1, \ldots, K$, має логнормальний розподіл з параметрами μ_i та σ, де значення σ вважають відомими, а значення μ_i – ні. Припускають, що середні величини позовів, спричинені i-им ураганом, дорівнюють $\Lambda_i = \exp\{\mu_i + \sigma^2/2\}$, які є н.о.р. в.в. із середнім p та дисперсією s_2^2;
v) Θ_i та Λ_i незалежні.

а) Покажіть, що $\mathrm{E}X_{ij} = p$, $\mathrm{D}X_{ij} = \exp\{\sigma^2\}(p^2 + s_2^2) - p^2$.

б) Нехай S_i – сумарні виплати, спричинені i-им ураганом, тоді умовний розподіл $S_i | \{\Theta_i, \Lambda_i\}$ – це складний розподіл Пуассона. Доведіть, що $\mathrm{E}S_i = np$ та
$\mathrm{D}S_i = (p^2 + s_2^2)(n^2 + s_1^2 + n\exp\{\sigma^2\}) - n^2 p^2$.

в) Знайдіть вираз для математичного сподівання та дисперсії сумарних річних позовів від усіх ураганів.

а) Покажіть, що $\mathsf{E}X_{ij} = p$, $\mathsf{D}X_{ij} = \exp\{\sigma^2\}(p^2 + s_2^2) - p^2$.
б) Нехай S_i – сумарні виплати, спричинені i-им ураганом, тоді умовний розподіл $S_i | \{\Theta_i, \Lambda_i\}$ – це складний розподіл Пуассона. Доведіть, що $\mathsf{E}S_i = np$ та
$\mathsf{D}S_i = (p^2 + s_2^2)(n^2 + s_1^2 + n\exp\{\sigma^2\}) - n^2p^2$.
в) Знайдіть вираз для математичного сподівання та дисперсії сумарних річних позовів від усіх ураганів.

а) Використаємо формули повного математичного сподівання і дисперсії. Тоді

$$\mathsf{E}X_{ij} = \mathsf{E}(\mathsf{E}X_{ij}|\Lambda_i) = \mathsf{E}\Lambda_i = p;$$

$$\begin{aligned}
\mathsf{D}X_{ij} &= \mathsf{E}(\mathsf{D}(X_{ij}|\Lambda_i)) + \mathsf{D}(\mathsf{E}(X_{ij}|\Lambda_i)) \\
&= \mathsf{E}(\Lambda_i^2(\exp\{\sigma^2\} - 1)) + \mathsf{D}\Lambda_i \\
&= (p^2 + s_2^2)(\exp\{\sigma^2\} - 1) + s_2^2 \\
&= (p^2 + s_2^2)\exp\{\sigma^2\} - p^2.
\end{aligned}$$

б) Оскільки умовний розподіл $S_i|\{\Theta_i, \Lambda_i\}$ – це складний розподіл Пуассона, то

$$\mathsf{E}(S_i|\Theta_i, \Lambda_i) = \Theta_i \Lambda_i.$$

З формули повного математичного сподівання та незалежності Θ_i та Λ_i випливає, що

$$\mathsf{E}S_i = \mathsf{E}(\mathsf{E}(S_i|\Theta_i, \Lambda_i)) = \mathsf{E}(\Theta_i \Lambda_i) = \mathsf{E}\Theta_i \mathsf{E}\Lambda_i = np.$$

Знову застосовуючи (4), матимемо

$$\mathsf{D}(S_i|\Theta_i, \Lambda_i) = \Theta_i \mathsf{E}(X_{ij}^2|\Lambda_i) = \Theta_i(\Lambda_i^2 \exp\{\sigma^2\})$$

$$\mathsf{E}(\mathsf{D}(S_i|\Theta_i, \Lambda_i)) = n(p^2 + s_2^2)\exp\{\sigma^2\},$$

$$\mathsf{D}(\mathsf{E}(S_i|\Theta_i, \Lambda_i)) = \mathsf{D}(\Theta_i \Lambda_i)$$

$$= \mathsf{E}(\Theta_i^2 \Lambda_i^2) - n^2 p^2 = (n^2 + s_1^2)(p^2 + s_2^2) - n^2 p^2.$$

Таким чином,

$$\mathsf{D}(S_i) = (n^2 + s_1^2)(p^2 + s_2^2) - n^2 p^2 + n(p^2 + s_2^2)\exp\{\sigma^2\}.$$

в) Нехай S – в.в., яка позначає річні сумарні позови, що надійшли від усіх позовів, тобто $S = S_1 + \cdots + S_K$, де K має пуассонівський розподіл з параметром λ, а S_i – н.о.р. в.в. Тому S має складний розподіл Пуассона та

$$\mathsf{E}S = \lambda \mathsf{E}S_i = \lambda np;$$

$$\mathsf{D}S = \lambda \mathsf{E}S_i^2 = \lambda(\mathsf{D}S_i + (\mathsf{E}S_i)^2) = \lambda(p^2 + s_2^2)(n^2 + s_1^2 + n\exp\{\sigma^2\}).$$

Приклад

Щороку страхова компанія підписує ряд договорів страхування домогосподарств, для кожного з яких страхова премія становить $\$80$. Сумарні річні позови від кожного поліса мають складний розподіл Пуассона з параметром 0,4 та індивідуальними позовами з гамма-розподілу з параметрами α і λ.

Накладні витрати на обробляння позову є в.в., рівномірно розподіленою між $\$50$ та $\$b$, $(b > 50)$, і не залежать від величини цього позову. В.в. S відображає сумарні позови й витрати від усього портфелю протягом року. Вважають, що S має близький до нормального розподіл.

а) Нехай $\alpha = 1$, $\lambda = 0,01$, $b = 100$. Доведіть, що компанія повинна продати за рік не менше 884 полісів, щоб досягти принаймні 99% упевненості в тому, що надходження премій перевищить видатки на покриття позовів і накладні витрати.

б) Припустимо тепер, що значення α, λ та b достовірно невідомі, але можуть знаходитися у таких межах:
$0,95 \leq \alpha \leq 1,05$; $0,009 \leq \lambda \leq 0,011$; $90 \leq b \leq 110$.

Вибравши найгірші можливі для компанії значення α, λ, b, визначте кількість полісів, які має продати компанія, щоб на 99% бути впевненою у тому, що надходження премій перевищуватиме видатки на покриття позовів і накладні витрати на їх оброблення.

Нехай X_i – це величина i-го позову, а Y_i – величина відповідних накладних витрат. Позначимо через N загальну кількість позовів за портфелем, а через n - кількість полісів у портфелі.

Оскільки N є сумою пуассонівських н.о.р. в.в., то N має розподіл Пуассона з параметром $0,4n$, а S можна зобразити, як

$$S = \sum_{i=1}^{N} (X_i + Y_i),$$

де $\{X_i + Y_i\}_{i \geq 1}$ є послідовністю н.о.р. в.в., незалежних від N. Звідси випливає, що S має складний розподіл Пуассона, у якому $Z_i = X_i + Y_i$ відображає величину i-го індивідуального позову (i-го доданка). Тоді

$$
\begin{aligned}
\mathsf{E}S &= 0,4n\mathsf{E}(X_i + Y_i), \\
\mathsf{D}S &= 0,4n\mathsf{E}(X_i + Y_i)^2 = 0,4n(\mathsf{E}X_i^2 + 2\mathsf{E}X_iY_i + \mathsf{E}Y_i^2) \\
&= 0,4n(\mathsf{E}X_i^2 + 2\mathsf{E}X_i\mathsf{E}Y_i + \mathsf{E}Y_i^2).
\end{aligned}
$$

Обчислимо моменти X_i Y_i у термінах α, λ та b:

$$\mathsf{E}X_i = \frac{\alpha}{\lambda};$$

$$\mathsf{E}X_i^2 = \frac{\alpha(\alpha+1)}{\lambda^2};$$

$$\mathsf{E}Y_i = \frac{50+b}{2};$$

$$\mathsf{E}Y_i^2 = \frac{b^2 + 50b + 2500}{3}.$$

а) Покладемо $\alpha = 1$, $\lambda = 0,01$, $b = 100$, тоді матимемо $\mathsf{E}S = 70n$ та $\mathsf{D}S = (127,80)^2 n$. І, відповідно, S наближено має нормальний розподіл із середнім $70n$ та стандартним відхилом $127,80\sqrt{n}$.

Надходження премій склало $80n$, і нам потрібно визначити найменше значення n, яке задовольнить нерівність

$$\mathsf{P}(S < 80n) \geq 0,99.$$

Стандартизуємо в.в. S, тоді

$$\mathsf{P}\left(\frac{S-70n}{127,80\sqrt{n}} < \frac{80n-70n}{127,80\sqrt{n}}\right) \geq 0,99. \tag{6}$$

Нагадаємо, що *квантилем рівня* α або α*%-им квантилем* розподілу $F_X(x)$ називають таку точку x_α, що

$$F_X(x_\alpha) = \mathsf{P}(X < x_\alpha) = \alpha.$$

99%-ий квантиль стандартного нормального розподілу дорівнює 2,326, отже, нерівність (6) перетворюється на таку:

$$\frac{80n-70n}{127,80\sqrt{n}} \geq 2,326,$$

звідки маємо $n \geq 883,7$ або $n \geq 884$, якщо розглядати найближче ціле число.

б) Найгіршою можливою комбінацією значень α, λ та b для страхової компанії є комбінація, яка дає найбільші можливі значення для ES й DS. Для того, щоб пересвідчитися у цьому, позначимо через μ та σ математичне сподівання та стандартний відхил сумарних позовів та витрат на їх обробляння для індивідуального полісу. Параметри μ і σ будуть функціями від α, λ, b, причому $\mathrm{E}S = n\mu$ та $\mathrm{D}S = n\sigma^2$. Аналогічним чином, як у п. а), отримуємо нерівність для n:

$$\frac{(80-\mu)\sqrt{n}}{\sigma} \geq 2,326,$$

звідки маємо, що

$$n \geq \left(\frac{2,326\sigma}{80-\mu}\right)^2.$$

Отже, найбільше значення n відповідає найбільшим значенням μ і σ (за умови, що найбільше значення μ є меншим за 80).