Лекція 4. /2год./ **Задачі перевірка гіпотез. Гіпотези про параметри. Гіпотези про розподіл.**

Задача вибору статистичної гіпотези – ймовірнісний підхід та методологія. Тестування параметричних гіпотез.

ПЕРЕВІРКА ГІПОТЕЗ

Якщо можливо висунути декілька взаємовиключних "гіпотез" про розподіл елементів вибірки, то виникає задача про вибір однієї з цих гіпотез на підставі вибіркових даних. Очевидно, що за вибіркою скінченного обсягу не можна гарантувати безпомилковість висновків про розподіл, тому доводиться зважати на можливість вибрати невірну гіпотезу.

4.1. Основні поняття. Гіпотези і критерії

Нехай ξ – випадкова величина, що спостерігається в експерименті, і $x_1, x_2, ..., x_n$ – вибірка.

Гіпотезою H називається будь-яке припущення про розподіл випадкової величини ξ.

*Гіпотеза називається **простою**, якщо вона однозначно визначає розподіл випадкової величини ξ. Інакше вона називається **складною** гіпотезою.*

*Якщо гіпотез тільки дві, то одну з них H_0 прийнято називати **основною**, а іншу H_1 – альтернативою або **конкуруючою** гіпотезою*.

Якщо розподіл випадкової величини ξ відомий з точністю до параметрів і за вибіркою спостережень необхідно перевірити припущення про значення параметрів цього розподілу, то такі гіпотези називаються параметричними.

*Правило, відповідно до якого приймається або відкидається гіпотеза H_0, називається **критерієм** перевірки гіпотези H_0, який призначатимемо K .*

Через те, що рішення приймається на основі вибірки $x_1, x_2, ..., x_n$ спостережень випадкової величини ξ, необхідно вибрати придатну статистику (тобто функцію від вибірки), яка називається у цьому випадку *статистикою критерію* K. При перевірці простої параметричної гіпотези $H_0 : \vartheta = \vartheta_0$ у якості статистики критерію звичайно обирають оцінку θ^* параметра ϑ .

Перевірка статистичної гіпотези ґрунтується на принципі, відповідно до якого малоймовірні події вважаються неправдоподібними, а події, що мають велику ймовірність, вважаються практично достовірними.

Цей принцип можна реалізувати в такий спосіб. Фіксується деяка мала ймовірність α, яка називається *рівнем значущості*. Нехай V – множина значень деякої статистики Z, а $V_k \subset V$ - критична множина, $V - V_k$ – ***областю***

прийняття гіпотези. Критична множина визначається умовою $\mathbf{P}\{(Z \in V_k)|H_0\} = \alpha$, тобто ймовірність потрапляння статистики критерію в критичну область V_k дорівнює α за умови, що гіпотеза H_0 справедлива.

Позначимо через z' вибіркове значення статистики Z, обчислене за вибіркою спостережень. Критерій K формулюється в такий спосіб: гіпотеза H_0 відхиляється, якщо $z' \in V_k$, і приймається, якщо $z' \in V - V_k$.

Критерій, заснований на заздалегідь обраному рівні значущості, називають ***критерієм значущості***. *Множина* V_k *називається* ***критичною областю****, а множина* $V - V_k$ *–* ***областю прийняття гіпотези*** H_0**.**

Рівень значущості α визначає розмір критичної області. На рис. 4.1 зображено розташування критичної області при перевірці гіпотези $H_0: \vartheta = \vartheta_0$ проти альтернативної гіпотези $H_1: \vartheta \neq \vartheta_0$.

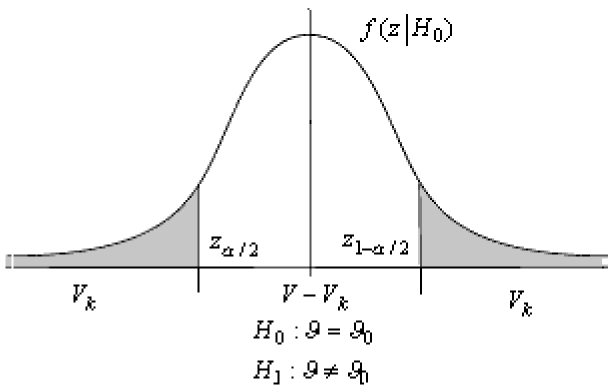

Рис. 4.1

Функція $f(z|H_0)$ – щільність розподілу статистики Z за умови, що справедлива гіпотеза H_0, $V - V_k$ – область прийняття гіпотези H_0, z_α – квантиль розподілу $f(z|H_0)$ рівня α, $\mathbf{P}\{Z \in V - V_k\} = 1 - \alpha$. Розташування критичної області на множині значень статистики Z залежить від формулювання альтернативної гіпотези H_1. Наприклад, якщо альтернативна гіпотеза H_1 формулюється як $H_1: \vartheta > \vartheta_0$ $(\vartheta < \vartheta_0)$, то критична область називається правобічною (лівобічною), тобто розміщується на правому (лівому) "хвості" розподілу статистики Z, тобто має вигляд нерівності $Z > z_{1-\alpha}$ $(Z < z_\alpha)$.

Зрозуміло, що може бути 4 варіанти:

- Гіпотеза істинна і вона приймається,
- Гіпотеза істинна і вона відхиляється
- Гіпотеза хибна і вона приймається
- Гіпотеза хибна і вона відхиляється.

Перший і четвертий варіанти не містять помилок. Другий і третій варіанти містять помилки, які мають спеціальні назви.

Помилка, зроблена при відхиленні правильної гіпотези H_0*, називається* ***помилкою першого роду***.

Ймовірність помилки першого роду дорівнює ймовірності попадання статистики критерію в критичну область за умови, що справедлива гіпотеза H_0, тобто дорівнює рівню значущості $\mathbf{P}\{(Z \in V_k)|H_0\} = \alpha$.(p-level)

*Помилкою **другого роду** є помилка, зроблена при прийнятті гіпотези H_0, у той час як справедлива гіпотеза H_1.*

Ймовірність помилки другого роду β називають потужністю критерію і обчислюють (за простою альтернативною гіпотезою) як

$$\beta = \mathbf{P}\{Z \in V - V_k | H_1\}.$$

Приклад 4.1. Є вибірка обсягу з нормального розподілу з параметрами $(a,1)$. Перевіряється проста гіпотеза $H_0 : a = 0$ проти альтернативної гіпотези $H_1 : a = 1$.

Використовується наступний критерій (для заданої константи c):

$$K(x_1) = \begin{cases} \text{ï ðèé ÿòü } H_0, & \text{åñëè } x_1 \le c; \\ \text{ï ðèé ÿòü } H_{1,} & \text{åñëè } x_1 > c. \end{cases}$$

На графіку зображено щільності, що відповідають гіпотезам, і імовірності помилок 1-го і 2-го роду критерію K :

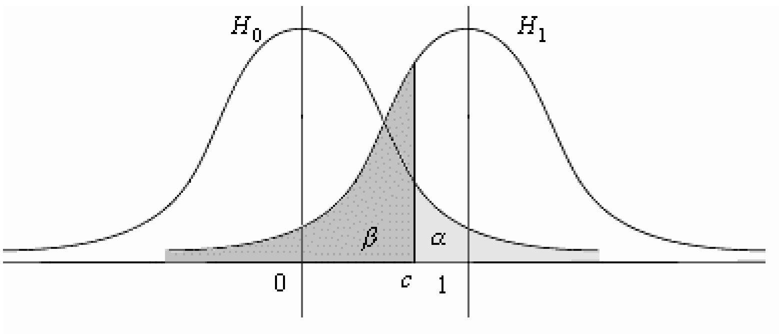

Рис. 4.2

Добре видно, якщо зменшується ймовірність помилки 1-го роду, то ймовірність помилки 2-го роду збільшується.

4.2. Категорії критеріїв. Критерії узгодженості перевірки гіпотез про розподіл

Статистичні критерії підрозділяють на такі категорії:

- **Критерій значущості.** Перевірка за значущістю припускає перевірку гіпотези про числові значення відомого закону розподілу або відомого параметру залежності:

  $H_0 : \quad a = a_0$— нульова гіпотеза
  $H_1 : \quad a > a_0 \quad (a < a_0)$або $a \neq a_0$—альтернативна, що конкурує.
- **Критерій узгодженості.** Перевірка на узгодженість має на увазі, що випадкова величина, що досліджується, узгоджується із законом, що розглядається. Наприклад, перевірка гіпотези про нормальний розподіл вибірки. Критерій узгодженості можна також сприймати, як критерій значущості.

- **Критерій однорідності.** При перевірці на однорідність випадкові величини досліджуються на факт взаємної відповідності їх законів розподілу (чи підкорюються ці величини одному і тому ж закону). Також використовуються у Факторному аналізі для визначення наявності залежностей.

Цей розділ умовний, і часто один і той же критерій може бути використаний в різних якостях.

Існує клас критеріїв, названих *критеріями узгодженості*, що використовуються для перевірки гіпотез (простих або складних) проти складних альтернатив. Усі критерії узгодженості будуються за єдиним принципом: задається деяка *функція відхилення* емпіричного розподілу від теоретичного. Вимагають, щоб ця функція збігалась до якогось власного розподілу, якщо справедлива гіпотеза, що перевіряється, і необмежено зростала, якщо гіпотеза не справедлива. Гіпотеза приймається або відкидається в залежності від величини даної функції відхилення. Ми сформулюємо ряд понять для випадку простої основної гіпотези, а надалі будемо їх коректувати у випадку необхідності.

Є вибірка $\boldsymbol{x} = (x_1, x_2, ..., x_n)$. Перевіряється проста основна гіпотеза:

$H_1 : x_i$ має закон розподілу $F_1(x) \in F$

проти складної альтернативи:

$H_2 : x_i$ має закон розподілу $\neq F_1(x) \in F$.

Відзначимо відразу, що будь-який критерій K для розрізняння цих гіпотез має цілком визначену помилку 1-го роду α. Але помилка 2-го роду може бути обчислена тільки тоді, коли відомо конкретний розподіл вибірки $F_2(x) \neq F_1(x)$ (одна з альтернатив). Тому зручно представити альтернативу у вигляді об'єднання (незліченного числа) простих альтернатив

$H_2 : x_i$ ì àº çàêî í ðî çï î ä³ëó $F_2(x) \in F$, $F_2(x) \neq F_1(x)$

за всіма можливими $F_2(x)$. Будемо розглядати помилку другого роду як функцію від $F_2(x)$: $\beta_{F_2(x)}(K) = \mathbf{P}(H_1 | F_2(x))$ – умовна ймовірність прийняти гіпотезу H_1, коли в дійсності має місце альтернативна гіпотеза H_2.

Означення 4.1. *Критерій* K *перевірки простої гіпотези* $H_1 : F_\xi(x) = F_1(x)$ *проти простої альтернативи* $H_2 : F_\xi(x) = F_2(x)$ *називається обґрунтованим якщо* $\beta_{F_2(x)}(K) = \mathbf{P}(H_1 | F_2(x)) \to 0$ при $n \to \infty$.

Критерій K *перевірки гіпотези* $H_1 : F_\xi(x) = F_1(x)$ *проти складної альтернативи* $H_2 : F_\xi(x) \in F$ *називається обґрунтованим, якщо він обґрунтований для будь-якої простої альтернативи* $H_2 : F_\xi(x) = F_2(x) \in F$.

Побудова критеріїв узгодженості. Розглянемо наступні умови.

К1. Потрібно задати функцію відхилення $\rho(\boldsymbol{x}) = \rho(\boldsymbol{x}, F_1)$ для $\boldsymbol{x} = (x_1, x_2, ..., x_n)$, яка має властивості:

а) якщо гіпотеза H_1 справедлива, то $\rho(\boldsymbol{x})$ збігається за ймовірністю до випадкової величини ξ, що має відомий розподіл $G_\xi(x)$;

б) якщо гіпотеза H_1 несправедлива, то $|\rho(\boldsymbol{x})| \to \infty$, $n \to \infty$ по ймовірності.

К2. Нехай таку функцію $\rho(\boldsymbol{x})$ задано. Для випадкової величини ξ з розподілом $G_\xi(x)$ та фіксованого $\varepsilon > 0$ визначимо постійну c з рівності $\varepsilon = \mathbf{P}\{|\xi| \geq c\}$ і побудуємо критерій:

$$K(\boldsymbol{x}) = \begin{cases} H_1, & \text{åñëè } |\rho(\boldsymbol{x})| < c; \\ H_2, & \text{åñëè } |\rho(\boldsymbol{x})| \geq c. \end{cases} \qquad (4.1)$$

Ми побудували критерій узгодженості. Він "працює" за таким принципом: якщо для даної вибірки функція відхилення велика (за абсолютним значенням), те це свідчить на користь альтернативи, і навпаки. Переконаємося у тому, що цей критерій має (асимптотичний) рівень ε і є спроможним (консистентним).

Зауваження. За визначенням, $\xi_n \to \infty$ при $n \to \infty$ еквівалентно тому, що для кожного $c > 0$ $\mathbf{P}\{\xi_n \leq c\} \to 0$ при $n \to \infty$. Для побудованого критерію K маємо

$$\alpha(K) = \mathbf{P}\left(\{|\rho(\boldsymbol{x})| \geq c\} \middle| H_1\right) \to \mathbf{P}\{|\xi| \geq c\} = \varepsilon .$$

І, крім того,

$$\beta_{F_2(x)}(K) = \mathbf{P}(H_1 | F_2(x)) = \mathbf{P}(\{|\rho(\boldsymbol{x})| < c\} | F_2(x)) = \lim_{n \to \infty} \mathbf{P}(\{|\xi| < c\} | F_2(x)) = 0$$

для будь-якої альтернативи $F_2(x) \in F$.

Критерій Колмогорова. Нехай є вибірка $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ об'єму n. Позначимо через $G(x)$ – істинну функцію розподілу, якій задовольняють реальні спостереження, через $F_n(x)$ – емпіричну (вибіркову) функцію розподілу, а через $F(x)$ – гіпотетичну функцію розподілу. Тоді гіпотеза H про те, що істинна функція розподілу є $F(x)$, записується у вигляді $H : G(x) = F(x)$. Якщо H справджується, то $F_n(x)$ та $F(x)$ мають проявляти певну схожість, і розбіжність між ними має зменшуватись зі зростанням n. Щоб показати близькість функцій можна використати віддаль між функціями. Наприклад, можна порівняти $F_n(x)$ та $F(x)$ за рівномірною метрикою, тобто розглянути величину:

$$\rho_n(\boldsymbol{x}) = \sqrt{n} \sup_y \left| F_n^*(y) - F_1(y) \right| .$$

Очевидно, що $\rho_n(\boldsymbol{x})$ – випадкова величина, оскільки її значення залежать від випадкового об'єкту $F_n(x)$. Якщо гіпотеза H справджується, то $F_n(x) \to F(x)$ при $n \to \infty$ для будь-якого $x \in \mathbb{R}$. Таким чином природно, що за цих умов $\rho_n(\boldsymbol{x}) \to 0$. Якщо ж гіпотеза H не справджується, то $F_n(x) \to G(x)$ і $G(x) \neq F(x)$, отже

$$\sup_y \left| F_n^*(y) - F(y) \right| \to \sup_y |G(y) - F(y)| .$$

Ця остання величина додатна, бо $G(x)$ не співпадає з $F(x)$. Така різниця у поведінці $\rho_n(\boldsymbol{x})$ в залежності від того, справджується H або ні, дозволяє використовувати $\rho_n(\boldsymbol{x})$ як статистику для перевірки H.

Важлива властивість $\rho_n(\boldsymbol{x})$ полягає у тому, що у разі, коли $G(x) = F(x)$ (тобто гіпотетичний розподіл визначений правильно), закон розподілу статистики виявляється одним й тим самим для всіх неперервних функцій $G(x)$. Він залежить тільки від об'єму вибірки n.

Теорема Колмогорова (див. лекція 1,2) стверджує, що у випадку, коли справджується H (та якщо $G(x)$ неперервна) величина $\mathbf{P}\{\rho_n(\boldsymbol{x})<z\}$ при $n\to\infty$ має границю, а саме:

$$\lim_{n\to\infty}\mathbf{P}\{\rho_n<z\}=1+2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2z^2}.$$

Для обчислення статистики $\rho_n(\boldsymbol{x})$ використовується формула

$$\rho_n=\max_{1\le k\le n}\left\{\frac{k}{n}-F(x_{(k)}),F(x_{(k)})-\frac{k-1}{n}\right\},$$

де через $x_{(1)},x_{(2)},...,x_{(n)}$ позначено елементи варіаційного ряду, який побудовано за вихідною вибіркою. Отриману величину ρ_n треба порівняти з вилученими з таблиць критичними значеннями. Гіпотезу H доводиться відкидати (за даним рівнем значущості), якщо отримане у результаті експерименту значення ρ_n перевищує критичне табличне значение, що відповідає цьому рівню значущості.

Критерій χ^2 (Пірсона). Критерій χ^2 базується на згрупованих даних. Область значень елементів вибірки поділяють на деяке число інтервалів. Після чого будують функцію відхилення ρ за різницями теоретичних ймовірностей потрапляння до інтервалів групування й емпіричних частот.

Нехай є вибірка $\boldsymbol{x}=(x_1,x_2,...,x_n)$. Перевіряється проста основна гіпотеза:

$H_1: x_i$ має закон розподілу $F_1(x)\in F$

проти складної альтернативи:

$H_2: x_i$ має закон розподілу $\ne F_1(x)\in F$.

Нехай $A_1,A_2,...,A_k$ – інтервали групування випадкової величини з розподілом F_1. Позначимо через v_k, $j=1,2,...,k$ число елементів вибірки, що потрапили в інтервал A_k

$$v_j=\{\text{÷èñëî }\ x_i\in A_j\}=\sum_{i=1}^{n}I(x_i\in A_j),$$

і через p_j теоретичну імовірність $\mathbf{P}(\{x_i\in A_j\}|H_1)$ потрапляння в інтервал A_j випадкової величини з розподілом F_1. Розглянуті величини задовольняють співвідношенням $\sum_{j=1}^{k}v_j=n,\ \sum_{j=1}^{k}p_j=1.$ З випадковими величинами $x_1,x_2,...,x_n$ природно пов'язана поліноміальна схема з n випробуваннями, у якій результатом i-го випробування є потрапляння x_i до якогось з інтервалів $A_1,A_2,...,A_k$. Якщо припущення про закон розподілу x_i вірно, то v_j повинно бути близько до $\mathbf{M}v_j=np_j$. Загальне відхилення усіх v_j можна вимірювати різними способами. Найчастіше як міру відхилення використовують величину

$$\rho(\boldsymbol{x})=\sum_{j=1}^{k}\frac{(v_j-np_j)^2}{np_j}=\sum_{j=1}^{k}\frac{v_j^{\,2}}{np_j}-n.$$

Теорема 4.1. (Пірсон). *Якщо справджується гіпотеза H_1, то за фіксованим k і $n\to\infty$*

$$\rho(\boldsymbol{x})=\sum_{j=1}^{k}\frac{(\nu_j-np_j)^2}{np_j}\to\chi_{k-1}^2 .$$

Доведення. Доведемо теорему для випадку $k=2$. Перетворимо вираз для $\rho(\boldsymbol{x})$

$$\rho(\boldsymbol{x})=\frac{(\nu_1-np_1)^2}{np_1}+\frac{(\nu_2-np_2)^2}{np_2}=\frac{(\nu_1-np_1)^2}{np_1}+\frac{((n-\nu_1)-n(1-p_1))}{n(1-p_1)}=$$

$$=\frac{(\nu_1-np_1)^2}{np_1}+\frac{(-\nu_1+np_1)^2}{n(1-p_1)}=\frac{(\nu_1-np_1)^2}{np_1(1-p_1)}=\left(\frac{\nu_1-np_1}{\sqrt{np_1(1-p_1)}}\right)^2 .$$

Величина ν_1 є сумою n незалежних випадкових величин з розподілом Бернуллі, і за центральною граничною теоремою $\frac{\nu_1-np_1}{\sqrt{np_1(1-p_1)}}$ збігається за ймовірністю до випадкової величини ξ, яка має стандартний нормальний розподіл. Отже,

$$\rho(\boldsymbol{x})=\frac{(\nu_1-np_1)^2}{np_1(1-p_1)}=\left(\frac{\nu_1-np_1}{\sqrt{np_1(1-p_1)}}\right)^2\to\xi^2=\chi_1^2 .$$

Наведене доведення відповідає на запитання: Чому сума квадратів k асимптотично розподілених нормальних випадкових величин збігається до розподілу χ_{k-1}^2, а не до χ_k^2? Куди зник один ступінь вільності?

Для довільного k міркування будується наступним чином: нескладно показати, що для будь-якого $j=1,2,...,k$

$$\frac{\nu_j-np_j}{\sqrt{np_j}}=\sqrt{n}\,\frac{\frac{\nu_j}{n}-p_j}{\sqrt{p_j}} .$$

Збігається за ймовірністю до випадкової величини ξ_j, яка має нормальний розподіл з параметрами $(0,1-p_j)$ і, отже, при достатньо великих k може бути замінено на стандартний нормальний розподіл. Тому за лемою Фішера

$$\sum_{j=1}^{k}\frac{(\nu_j-np_j)^2}{np_j}=\sum_{j=1}^{k}\xi_j^2\to\chi_{k-1}^2 .\square$$

Отже, довели властивість К1.(а) для $\rho(\boldsymbol{x})$.

Зауваження. Застосовуючи критерій χ^2 для розв’язання задачі про перевірку гіпотези, необхідно пам'ятати, що цей критерій не є спроможним для альтернатив з тими ж ймовірностями потрапляння в інтервали розбиття, що й у F_1. Тому беруть велику кількість інтервалів розбиття – чим більше, тим краще. Для вибірки обсягу n число інтервалів розбиття вибирають так, щоб забезпечити потрібну точність при заміні розподілу $\rho(\boldsymbol{x})$ на χ_{k-1}^2. Звичайно вимагають, щоб $np_1=np_2=...=np_k\approx 6\div 9.$ Якщо в деякому інтервалі ця умова не виконується, то його варто об'єднати із сусіднім.

За заданим ε можна знайти c таке, що $\varepsilon=\mathbf{P}\{\chi^2\geq c\}$. Залишилося побудувати критерій узгодженості χ^2:

$$K(\boldsymbol{x}) = \begin{cases} H_1, & \text{åñëè } \rho(\boldsymbol{x}) < c; \\ H_2, & \text{åñëè } \rho(\boldsymbol{x}) \geq c. \end{cases}$$

$\rho(\boldsymbol{x}) < \chi^2_{1-\alpha}(k - m - 1)$, де m – число параметрів розподілу $F(x)$, що оцінюються за вибіркою;

якщо ж $\rho(\boldsymbol{x}) \geq \chi^2_{1-\alpha}(k - m - 1)$, то гіпотеза H_1 відхиляється, теоретичні та емпіричні частоти відрізняються суттєво.

Приклад.(перевірка гіпотези про розподіл) У перших двох стовпцях таблиці 4.1 наведено дані про відмови апаратури за 10 000 годин роботи. Загальна кількість обстежених екземплярів апаратури $n = 757$, при цьому спостерігалося число відмовлень

$$0 \cdot 427 + 1 \cdot 235 + 2 \cdot 72 + 3 \cdot 21 + 4 \cdot 1 + 5 \cdot 1 = 451.$$

Перевіримо гіпотезу про те, що кількість відмов має розподіл Пуассона:

$$p_k = \mathbf{P}\{x = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots$$

при $\alpha = 0{,}01$.

Таблиця 4.1

Число відмов k	Кількість випадків, у яких спостерігалося k відмов, n_k	$p_k = \frac{0{,}6^k}{k!} e^{-0{,}6}$	Очікуване число випадків з k відмовами, np_k
0	427	0,54881	416
1	235	0,32929	249
2	72	0,09879	75
3	21	0,01976	15
4	1	0,00296	2
5	1	0,00036	0
≥ 6	0	0,00004	0
Сума	757		

Оцінка параметра λ дорівнює середньому числу відмов: $\hat{\lambda} = 451/757 \approx 0{,}6$. За таблицями розподілу Пуассона знаходимо p_k і np_k. Оскільки для $k = 4, 5, 6$ значення $np_k < 5$, поєднуємо ці рядки з рядком для $k = 3$. У результаті одержимо значення, наведені у таблиці 4.2.

Таблиця 4.2

k	n_k	np_k	$\frac{(n_k - np_k)^2}{np_k}$
0	427	416	0,291
1	235	249	0,787

2	72	75	0,120
≥ 3	23	17	2,118
			$\rho(x)$ =3,316

Оскільки за вибіркою оцінювався тільки один параметр λ, то $m=1$. Отже, число степенів вільностей дорівнює $4-1-1=2$. За таблицею квантилей розподілу χ^2 знаходимо $\chi^2_{0,99}(2)=9,21$. Гіпотеза про розподіл числа відмов за законом Пуассона приймається.