МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ

Факультет ІТМ

Кафедра прикладної математики

КУРСОВА РОБОТА

ПОЯСНЮВАЛЬНА ЗАПИСКА

Граничні теореми теорії ймовірностей (центральна гранична теорема для незалежних однаково розподілених випадкових величин). Перевірка статистичних гіпотез (критерій незалежності Спірмена)

(тема роботи)

Теорія ймовірностей і математична статистика

(дисципліна)

Керівник доц. каф. ПМ Гибкіна Н.В.

(підпис, дата, посада, прізвище, ініціали)

Студент Подшиваленко Б.О., ППМ-16-1

(група, підпис, дата, прізвище, ініціали)

Робота захищена
з оцінкою________________
«___»______________2018 р.

Харків 2018

Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту
Кафедра Прикладної математики
Дисципліна Теорія ймовірностей і математична статистика
Спеціальність 113 Прикладна математика
Курс 2 Група ППМ-16-1 Семестр 4

ЗАВДАННЯ
НА КУРСОВУ РОБОТУ

студентові Подшиваленко Борису Олександровичу
(прізвище, ім'я, по батькові)

1. Тема роботи Граничні теореми теорії ймовірностей (центральна гранична теорема для незалежних однаково розподілених випадкових величин). Перевірка статистичних гіпотез (критерій згоди Колмогорова: проста гіпотеза)

2. Термін здачі студентом закінченої роботи 28 травня 2018 р.

3. Вихідні дані до роботи: теоретичні питання та практичні завдання відповідно до варіанту

4. Зміст пояснювальної записки (перелік питань, які потрібно розробити)

Вступ
1 Теоретична частина
1.1 Граничні теореми теорії ймовірностей
1.2 Перевірка статистичних гіпотез
2 Практична частина
2.1 Розв'язання задач на тему «Основні ймовірнісні розподіли»
2.2 Розв'язання задач на тему «Граничні теореми теорії ймовірностей»
2.3 Перевірка гіпотез за допомогою критерію згоди Колмогорова: проста гіпотеза
Висновки

5. Дата видачі завдання 09 березня 2018 р.

КАЛЕНДАРНИЙ ПЛАН

Номер	Назва етапів курсової роботи	Термін виконання етапів роботи	Примітка
1	Вибір теми та отримання завдання на курсову роботу	4 тиждень	виконано
2	Підбір та вивчення літератури за темою роботи	5 – 6 тиждень	виконано
3	Виконання теоретичної частини курсової роботи	7 – 9 тижні	виконано
4	Виконання практичної частини курсової роботи	10 – 12 тижні	виконано
5	Оформлення пояснювальної записки	13 тиждень	виконано
6	Захист курсової роботи	14 тиждень	виконано

Студент ________________________________
(підпис)

Керівник роботи ________________________
(підпис)

РЕФЕРАТ

Пояснювальна записка: 20 с., 10 джерел.

Об'єкт дослідження – центральна гранична теорема для незалежних однаково розподілених випадкових величин, критерій згоди.

Мета роботи – вивчення розділу теорія ймовірностей і математична статистика, які надають можливість не тільки одержувати знання, а й допомагають розуміти закономірності навколишнього світу, та знаходити їх практичне застосування в повсякденному житті.

Робота складається з двох частин - теоретичної і практичної. У теоретичній розглянутий цикл граничних теорем, а також критерії для перевірки статистичних гіпотез. У практичній частині – розв’язання завдань за темами "Основні ймовірнісні розподіли", "Граничні теореми теорії ймовірності" і " Перевірка гіпотез за допомогою критерію незалежності Спірмена".

Ключові слова: МАТЕМАТИЧНА СТАТИСТИКА, ГРАНИЧНА ТЕОРЕМА, ВИПАДКОВА ВЕЛИЧИНА, ТЕОРЕМА ЛЯПУНОВА, ТЕОРЕМА ЛІНДЕБЕРГА, ХАРАКТЕРИСТИЧНА ФУНКЦІЯ, ГІПОТЕЗА, КРИТЕРІЙ НЕЗАЛЕЖНОСТІ.

Зміст

ВСТУП

Курс "Теорія ймовірності та математична статистика" займає особливе місце в системі математичних дисциплін, які вивчаються студентами спеціальностей «Прикладна математика», «Системний аналіз», як базовий курс. Вивчення курсу потрібне для освоєння основних понять і методів аналізу даних для вирішення конкретних завдань, а також забезпечення інших математичних дисциплін.

Метою курсової роботи є поглиблення теоретичних знань по курсу "Теорія ймовірності та математична статистика", розвиток навичок самостійної роботи; практичне застосування теорії ймовірності і математичної статистики при рішенні прикладних завдань.

Математична статистика, як галузь математичних знань, базується на теорії ймовірностей і є наукою про методи висновки щодо властивостей досліджуваної статистичної сукупності. Відповідаючи на багато запитань різних наук, математична статистика сформувалась в великий арсенал математично-статистичних прийомів обробки емпіричних даних.

Курсова робота присвячена розв’язанню окремих задач теорії ймовірності за темами «Центральна гранична теорема», «Перевірка статичних гіпотез» та більш глибокому вивченню теоретичних питань за цими темами.

1 ТЕОРЕТИЧНА ЧАСТИНА

1.1 Граничні теореми теорії ймовірностей

1.1.1 Збіжність послідовностей випадкових величин та ймовірнісних розподілів

Розглянемо ймовірнісний простір $\langle \Omega, \mathrm{F}, \mathsf{P} \rangle$ задані система випадкових величин $\xi_1,\ \xi_2,\ ...,\ \xi_n,\ ...$ та випадкова величина ξ. Так як в теорії ймовірностей існує декілька видів збіжності послідовностей випадкових величин, розглянемо їх докладніше.

Означення. Послідовність випадкових величин $\{\xi_n\}_{n=1}^{\infty}$ збігається за ймовірністю до випадкової величини ξ, якщо для будь-якого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}\left(|\xi_n - \xi| \geq \varepsilon\right) = 0 .$$

Цей вид збіжності позначають $\xi_n \xrightarrow{\mathsf{P}} \xi$ або $\mathsf{P}\lim\limits_{n\to\infty} \xi_n = \xi$.

Означення. Послідовність випадкових величин $\{\xi_n\}_{n=1}^{\infty}$ збігається за ймовірністю до випадкової величини ξ майже всюди (або з ймовірністю 1), якщо

$$\mathsf{P}(\omega : \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)) = 1 .$$

Цей вид збіжності позначають $\xi_n \xrightarrow{\text{м.в.}} \xi$.

Означення. Послідовність випадкових величин $\{\xi_n\}_{n=1}^{\infty}$ збігається за ймовірністю до випадкової величини ξ у середньому порядку p, якщо

$$\lim_{n\to\infty} M\left|\xi_n - \xi\right|^p = 0 .$$

Цей вид збіжності позначають $\xi_n \xrightarrow{(p)} \xi$.

Означення. Розподіл ймовірностей P_{ξ_n} випадкових величин ξ_n слабко збігається до розподілу P_{ξ} випадкової величини ξ, якщо

$$\forall \varphi \quad M\varphi(\xi_n) \to M\varphi(\xi) ,$$

де φ – неперервна обмежена функція.

Цей вид збіжності позначають $F_{\xi_n} \Rightarrow F_{\xi}$.

Також умову $M\varphi(\xi_n) \to M\varphi(\xi)$ можна представити у вигляді інтеграла Лебега по мірі P, наступним чином:

$$\int_{-\infty}^{\infty} \varphi(x) P_{\xi_n} dx \to \int_{-\infty}^{\infty} \varphi(x) P_{\xi} dx .$$

Теорема. Слабка збіжність має місце тоді й тільки тоді, коли в кожній точці неперервності функції $F_{\xi}(x) : F_{\xi_n}(x) \to F_{\xi}(x), n \to \infty$. Кажуть, послідовність збігається за розподілом до випадкової величини ξ.

Цей вид збіжності позначають $\xi_n \xrightarrow{d} \xi$.

Між типами збіжності є наступні імплікації: збіжність майже всюди та збіжність у середньому порядку p є найсильнішими, з кожної з них слідує збіжність за ймовірністю. Збіжність за розподілом є найслабшою (тому їй і відповідає слабка збіжність розподілів) та слідує з будь-якого з трьох попередніх типів. Ми прийшли думки , що збіжність послідовностей випадкових величин та їх розподілів є дуже важливим поняттям при формулюванні та доведенні граничних теорем теорії ймовірностей.

1.1.2 Метод характеристичних функцій

Метод характеристичних функцій, запропонований Ляпуновим, є одним з основних коштів аналітичного апарату теорії ймовірності. Також цей метод є досить ефективним у доказі найрізноманітніших граничних теорем, що і зазначає його розвиток та широке застосування. Наряду з випадковими величинами (що приймають дійсні значення) теорія характеристичних функцій вимагає залучення комплекснозначних випадкових величин.

Кожну випадкову величину можна описати не лише її функцією розподілу, а й її характеристичною функцією.

Означення. Характеристичною функцією випадкової величини ξ називається комплексна функція дійсного аргумента

$$f_{\xi}(t) = Me^{it\xi} = \int\limits_{R^n} e^{itx} dF_{\xi}(x), t \in R^n,$$

де $F_{\xi} = F_{\xi}(x_1, x_2, ..., x_n)$ – функція розподілу вектора $\xi = (\xi_1, \xi_2, ..., \xi_n)$.

Для дискретних випадкових величин вона має вигляд

$$f_{\xi}(t) = \sum_k e^{itx_k P(\xi = x_k)},$$

а для абсолютно неперервних, відповідно,

$$f_{\xi}(t) = \int\limits_{-\infty}^{+\infty} e^{itx} p_{\xi}(x) dx.$$

Характеристичні функції мають наступні властивості:

1. Будь-яка характеристична функція неперервна та обмежена:

$$\forall \xi \;\; f_\xi \in C(\mathbb{R}),\;\; \forall t \in \mathbb{R} \left| f_\xi(\mathrm{t}) \right| \le 1,\;\; f_\xi(0) = 1.$$

2. Нехай $\eta = a\xi + b$. Тоді $f_\eta(t) = f_\xi(at)e^{itb}$.

3. Якщо $\xi_1, \xi_2, ..., \xi_n$ – незалежні випадкові величини, то

$$f_{\xi_1+\xi_2+...+\xi_n}(t) = \prod_{i=1}^{n} f_{\xi_i}(t)$$

4. Якщо випадкова величина ξ має центральний момент n-го порядку m_n, то функція $f_\xi(t)$ є n раз диференційованою, та $\forall k \le n f_\xi^k(0) = i^k \mu_k$.

Характеристичні функції не лише дають ще один спосіб описувати випадкові величини, а й надають апарат для перевірки збіжності послідовності випадкових величин за розподілом. Це описується у наступній теоремі.

Теорема неперервності. Нехай $\{\xi_n\}_{n=1}^{\infty}$ – послідовність випадкових величин, кожна з яких має функцію розподілу $F_{\xi_n}(x)$ та характеристичну функцію $f_{\xi_n}(t)$. Слабка збіжність функцій розподілу до функції $F_\xi(x)$ має місце тоді й тільки тоді, коли $\forall t \;\; f_{\xi_n}(t) \to f_\xi(t)$.

1.1.3 Закони великих чисел

Закон великих чисел – сукупність теорем, визначені умови прямування середніх арифметичних значень випадкових величин до деякої константи при проведенні великого числа опитів.

Нехай $\{\xi_n\}_{n=1}^{\infty}$ – послідовність випадкових величин.

Означення. Законами великих чисел (слабкими) називають цикл теорем, в яких формуються умови збіжності за ймовірністю послідовності емпіричних середніх значень до теоретичного, тобто умови, за яких

$$\frac{1}{n}S_n - \frac{1}{n}\sum_{i=1}^{n} a_n \xrightarrow{p} 0,$$

де $a_n = M\xi_n, S_n = \sum_{i=1}^{n} \xi_i$.

Одна з найбільш важливих форм закону великих чисел – теорема Чебишова, вона встановлює зв’язок між арифметичним спостерігаємим значенням випадокової величини та її математичним сподіванням.

Теорема (закон великих чисел у формі Чебишова). Нехай $\{\xi_n\}_{n=1}^{\infty}$ – послідовність випадкових величин зі скінченними математичними сподіваннями та дисперсіями, обмеженими у сукупності ($\exists D \in {}^{\sim} : \forall n \in \quad D\xi_n \le D$), то для неї виконується закон великих чисел. Якщо $\{\xi_n\}_{n=1}^{\infty}$ – послідовність незалежних однаково розподілених випадкових величин зі скінченним математичним сподіванням a та скінченною дисперсією, то $\frac{1}{n}S_n \xrightarrow{p} a$.

Теорема (закон великих чисел у формі Хінчіна). Якщо $\{\xi_n\}_{n=1}^{\infty}$ – послідовність незалежних однаково розподілених випадкових величин зі скінченним математичним сподіванням a, то для них виконується закон великих чисел. Дана теорема дозволяє прибрати умову скінченності дисперсії.

Теорема (закон великих чисел у формі Маркова). Якщо $\{\xi_n\}_{n=1}^{\infty}$ – довільна послідовність випадкових величин, що мають моменти другого порядку, тоді

$$\lim_{n\to\infty} \frac{1}{n^2} DS_n = 0 \Rightarrow \frac{1}{n}S_n - \frac{1}{n}\sum_{i=1}^{n} a_n \xrightarrow{p} 0.$$

Теорема (закон великих чисел для схеми випробувань Бернуллі).

Покладемо у закон великих чисел у формі Хінчіна, що випадкова величина ξ_i приймає значення 1 та 0 з ймовірностями, p та $1-p$ та відповідає успіху в i-му випробуванні за схемою Бернуллі. Тоді

$$\forall\varepsilon>0\ \ P(\left|\frac{\mu_n}{n}-p\right|\geq\varepsilon)=0,$$

де $M\xi_i=p, S_n=\mu_n$ – кількість успіхів серед перших n випробувань.

Отже, чим більше випробувань за схемою Бернуллі буде проведено, тим менша ймовірність великих відхилень відносної частоти успіхів від теоретичної ймовірності успіху.

Теорема (Бернуллі). Для будь-якого $\varepsilon>0$

$$P(\left|\nu_n-p\right|\geq\varepsilon)\xrightarrow[n\to\infty]{}0,$$

де $\nu_n=\frac{1}{n}\sum_{i=1}^{n}\xi_i$ – частота появи успіху в n випробуваннях.

1.1.4 Посилені закони великих чисел

Нехай $\{\xi_n\}_{n=1}^{\infty}$ – послідовність випадкових величин.

Означення. Посиленими законами великих чисел називають цикл теорем, в яких встановлюються умови збіжності майже всюди послідовності емпіричних середніх значень до математичного сподівання, тобто умови, за яких

$$\frac{1}{n}S_n-\frac{1}{n}\sum_{i=1}^{n}a_n\xrightarrow{M.B.}0,$$

де $a_n = M\xi_n, S_n = \sum_{i=1}^{n} \xi_i$.

Посилений закон великих чисел був вперше сформульований й доведений Е. Борелем для схеми Бернуллі.

Теорема (посилений закон великих чисел для схеми випробувань Бернуллі). Нехай $\mu_n = \sum_{i=1}^{n} \xi_i$ – кількість успіхів серед n випробувань за схемою Бернуллі з ймовірністю успіху p. Тоді відносна частота успіхів збігається майже всюди до теоретичної ймовірності успіху: $\frac{\mu_n}{n} \xrightarrow{M.B.} p,\ n \to \infty$.

Теорема (перший посилений закон великих чисел Колмогорова). Якщо $\{\xi_n\}_{n=1}^{\infty}$ – послідовність випадкових величин, що мають скінченні математичні сподівання a_n та скінченні дисперсії σ_n^2, та при цьому виконується умова $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n} < \infty$, то має місце посилений закон великих чисел.

Теорема (другий посилений закон великих чисел Колмогорова). Якщо $\{\xi_n\}_{n=1}^{\infty}$ – послідовність незалежних однаково розподілених випадкових величин, що мають скінченне математичне сподівання a, то має місце посилений закон великих чисел.

1.1.5 Центральна гранична теорема для незалежних однаково розподілених випадкових величин

Різні форми центральної граничної теореми відрізняються між собою умовами, що накладаються на розподіл утворюючи суму випадкових доданків. Ми сформулюємо та доведемо одну з форм центральної граничної теореми, відносних до розподілених доданків

Закони великих чисел встановлювали лише факт збіжності (за ймовірністю чи майже) відхилення емпіричного середнього від його математичного сподівання до нуля. Центральна гранична теорема дозволяє, знаючи дисперсії випадкових величин, дещо вточнити характер збіжності.

Отже $\{\xi_n\}_{n=1}^{\infty}$ – послідовність випадкових величин, позначимо $S_n = \frac{1}{n}\sum_{i=1}^{n}\xi_i$.

Означення. Послідовність випадкових величин $\{\xi_n\}_{n=1}^{\infty}$ називається асимптотично нормальною, якщо розподіли випадкових величин $\zeta_n = \frac{S_n - MS_n}{\sqrt{DS_n}}$ слабко збігаються до стандартного нормального розподілу, тобто, якщо

$$\forall x \in {}^{\sim} \quad P(\zeta_n < x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt, n \to \infty.$$

Означення. Центральними граничними теоремами називається цикл теорем, що встановлюють умови, за яких послідовність $\{\xi_n\}_{n=1}^{\infty}$ є асимптотично нормальною.

Теорема. Нехай $\{\xi_n\}_{n=1}^{\infty}$ – послідовність незалежних однаково розподілених випадкових величин з математичним сподіванням $M\zeta_n = a$ та скінченною дисперсією $D\xi_n = \sigma^2$. Тоді вона є асимптотично нормальною.

Доведення. Спершу зауважимо, що в таких позначеннях $\zeta_n = \frac{S_n - an}{\sigma\sqrt{n}}$.

Для доведення теореми використаємо метод характеристичних функцій, а саме теорему неперервності.

Нехай $f_{\xi}(t)$ – характеристична функція кожної з випадкових величин послідовності. Знайдемо функції $f_{\zeta_n}(t)$ та покажемо, що їх послідовність збігається в усіх точках t до характеристичної функції стандартного нормального розподілу.

Згідно з властивістю 2 характеристичних функцій маємо:

$$f_{\zeta_n}(t) = f_{s_n}\left(\frac{t}{\sigma\sqrt{n}}\right)e^{-it\frac{an}{\sigma\sqrt{n}}}.$$

Маючи на увазі, що S_n є сумою незалежних випадкових величин (які до того ж однаково розподілені), за властивістю 3 характеристичних функцій отримуємо:

$$f_{\zeta_n}(t) = \left[f_{\xi}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]e^{-it\frac{an}{\sigma\sqrt{n}}}.$$

Випадкові величини ξ_n мають моменти другого порядку (дисперсію), тож їх характеристична функція є двічі диференційованою, при чому

$$f_{\xi}(0) = 1, f'_{\xi}(0) = iM\xi_n = ia, f''_{\xi}(0) = -M\xi_n^2;$$

Отже, розкладаючи за формулою Маклорена ми отримаємо:

$$\ln\phi_{\zeta_n}(t) = n\ln[1 - \frac{\sigma^2}{2}(\frac{t}{\sigma\sqrt{n}})^2 + \text{o}(\frac{t^2}{n})] = \text{n}(-\frac{2t^2}{n} + \text{o}(\frac{t^2}{n})) =$$

$$= -\frac{t^2}{2} + \text{o}(1) \to -\frac{t^2}{2},\ n \to \infty$$

звідки випливає, що

$$\forall t\ \ \phi_{\zeta_n}(t) \to e^{-\frac{t^2}{2}},\ n \to \infty.$$

Відомо, що $e^{-\frac{t^2}{2}}$ – характеристична функція стандартного нормального розподілу. За теоремою неперервності робимо висновок, що розподіли ймовірностей послідовності випадкових величин $\{z_n\}_{n=1}^{\bullet}$ слабко збігаються до стандартного нормального розподілу.

Центральна гранична теорема доведена.

1.1.6 Теореми Ляпунова та Ліндеберга

Розглянемо дві теореми, що дозволяють встановити збіжність послідовності незалежних та необов'язково однаково розподілених випадкових величин.

Нехай $\{\xi_n\}_{n=1}^{\infty}$ – послідовність незалежних випадкових величин, $S_n = \sum_{i=1}^{n} \xi_i, a_k = M\xi_k, \sigma_k^2 = D\xi_k$ та $s_n^2 = DS_n$.

Теорема Ліндеберга. Якщо для послідовності $\{\xi_n\}_{n=1}^{\infty}$ виконується умова Ліндеберга:

$$\forall \varepsilon > 0 \ \lim_{n\to\infty} \sum_{k=1}^{n} M[(\xi_k - a_k)^2 : |\xi_k - a_k| \geq \varepsilon s_n^2] = 0,$$

$$M\left[(\xi_k - a_k)^2 : |\xi_k - a_k| \geq \varepsilon s_n^2\right] = \int\limits_{\{x:|x-a_k|\geq \varepsilon s_n^2\}} (x-a_k)^2 dF_{\xi_k}(x),$$

де $F_{\xi_k}(x)$ – функція розподілу випадкової величини ξ_k, то вона збігається до випадкової величини з нормальним розподілом.

Теорема Ляпунова. Нехай для цієї ж послідовності виконується умова Ляпунова:

$$\exists \delta > 0 : \lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} M|\xi_k - a_k|^{2+\delta} = 0.$$

Тоді умови Ліндеберга також виконуються і послідовність $\{\xi_n\}_{n=1}^{\infty}$ є асимптотично нормальною.

1.2 Перевірка статистичних гіпотез

1.2.1 Основні задачі математичної статистики та їх характеристика

Теорії ймовірностей – закон розподілу випадкової величини із вже відомими параметрами. На основі цього потім розраховуються ймовірності подій, що нас можуть цікавити з тих чи інших причин, робляться висновки, що може трапитися і наскільки це ймовірно.

Математична статистика – розділ математики, який займається розробкою методів отримання науково обґрунтованих висновків про масові процеси та явища за даними спостережень або експериментів.

Основними задачами математичної статистики є:

- оцінювання невідомих параметрів;
- перевірка статистичних гіпотез;
- встановлення форми та ступеня зв’язку між випадковими величинами.

Задача перша, оцінювання невідомих параметрів виникає в тих випадках, коли функція розподілу генеральної сукупності відома з точністю до параметра θ.

У першому випадку, коли оцінкою є єдине число (функція, вектор), мова йде про точкове оцінювання. Найуживанішими методами для його виконання є метод моментів та метод максимальної вірогідності. Коли ж оцінкою є діапазон, мова йде про інтервальне оцінювання.

У другому випадку, статистичною гіпотезою називають будь-яке припущення про розподіл ймовірності спостережуваної випадкової величини -

скалярної або векторної.

В деякому розумінні завдання перевірки статистичної гіпотези є зворотним до завдання оцінювання параметра. При оцінюванні параметра ми нічого не знаємо про його істинне значення. При перевірці статистичної гіпотези ми з якихось міркувань припускаємо відомим його значення і хочемо за результатами експерименту перевірити наше припущення.

Третя задача, дозволяє підтвердити чи спростувати гіпотетичний взаємозв'язок між величинами. Щоб встановити, чи є певна закономірність між значеннями двох випадкових величин, можуть бути використані, наприклад, кореляційний чи регресійний аналіз. Можна також перевіряти й вплив невипадкових факторів. Для цього, знов-таки, використовується регресійний аналіз, а для перевірки впливу певного номінативного фактору, що має декілька ступенів, на досліджувану випадкову величину може бути застосований дисперсійний аналіз.

Також слід зауважити, що для третього випадку необхідні деякі данні з першого та другого.

1.2.2 Перевірка статистичних гіпотез: основні поняття

Означення. Статистичною гіпотезою H називається будь-яке припущення відносно розподілу випадкової вибірки, спостережуваної в стохастичному експерименті.

Означення. Гіпотеза, що перевіряється, називається основною (або нульовою) – H_0. Альтернативна гіпотеза – H_1.

Означення. Статистична гіпотеза $H_0 : \theta \in \Theta$ називається простий, якщо множина Θ_0– одноелементне, причому розподіл генеральної сукупності повністю відомий. Інакше ця гіпотеза називається складною.

Перевірка гіпотези обгрунтована на обчисленні статистики критерію за результатами спостереження.

Означення. Статистичний критерій перевірки гіпотези H_0 – це правило, відповідно до якого на підставі спостережуваного значення статистики $z_{спост} = \phi(\vec{x}_n)$ гіпотеза H_0 приймається або відкидається.

Означення. Критична область $\overline{G}$ у статистиці критерію - це область тих реалізацій $z_{спост}$ статистики ζ при яких гіпотеза H_0 відкидається.

Означення. Довірча область критерію G – це значення $z_{спост}$ статистики ζ, при яких гіпотеза H_0 приймається.

Означення. Помилкою першого роду називається відхилення нульової гіпотези за умови що вона вірна.

Означення. Помилкою другого роду називається прийняття нульової гіпотези за умови, що вірна альтернатива.

Означення. Рівень значущості (довірчий рівень, критичний рівень) статистичного критерію – це функція від параметра, завдана ймовірності помилки першого роду:

$$\mathrm{P}_\theta(\varphi(\vec{X}_n) \in \overline{G}) = \mathrm{P}_\theta(\vec{X}_n \in \omega),\ \theta \in \Theta,$$

де $\omega = \left\{\vec{x}_n \in \mathrm{X}^n : \varphi(\vec{x}_n) \in \overline{G}\right\}$ – критична область вибірки.

Означення. Потужність критерію – це ймовірність відсутності помилки другого роду, тобто ймовірність правильного альтернативного виведення, коли видно альтернатива.

$$1 - \mathrm{P}_\theta(\phi(\vec{X}_n) \in G) = \mathrm{P}_\theta(\phi(\vec{X}_n) \in \overline{G}) = \mathrm{P}_\theta(\vec{X}_n \in \omega),\ \theta \in \Theta.$$

Означення. Оперативна характеристика критерію – це ймовірність попадання випадкової вибірки в критичну область як функція від усіх значень параметра θ, $\theta \in \Theta_0 \cup \Theta_1$.

Означення. Якщо нульова гіпотеза проста, то рівень значущості задається одним числом.

Означення. Критерій називається незміщеним, якщо його потужність не менше його рівня значущості, так як:

$$\mathrm{P}_{\theta}(\varphi(\vec{X}_n) \in \overline{G}) = \mathrm{P}_{\theta}(\vec{X}_n \in \omega) \geq \alpha \ , \forall \theta \in \Theta_1 .$$

Означення. Критерій рівня значущості є слушним критерієм, якщо його потужність прагне до одиниці зі збільшенням об'єму випадкової вибірки.

$$\mathrm{P}_{\theta}(\phi(\vec{X}_n) \in \overline{G}) \to 1, \ n \to \infty \ \forall \theta \in \Theta_1 \ .$$

1.2.3 Критерій незалежності Спірмена

На практиці для перевірки гіпотези незалежності часто використовують рангові критерії. Найбільш відомим з них є критерій Спірмена, названий ім’ям видатного психолога, який ввів його в 1904 році. Він ґрунтується на коефіцієнті рангової кореляції ρ, визначеному наступним чином.

Позначимо через R_i ранг X_i серед спостережень $X_1,...,X_n$ (тобто номер місця, займаного величиною X_i в варіаційному ряду $X_{(1)} \leq ... \leq X_{(n)}$); аналогічно, нехай S_i- ранг Y_i серед спостережень $\mathrm{Y}_i,...,\mathrm{Y}_i$. Таким чином вихідні дані (X_i, Y_i), $i = 1,...,\mathrm{n}$ породжують n пар рангів $(\mathrm{R}_i, \mathrm{S}_i)$, $i = 1,...,\mathrm{n}$. Переставивши ці пари в порядку зростання першої компоненти, позначимо отриману множину пар через $(\mathrm{I}, \mathrm{T}_1),...,(\mathrm{n}, \mathrm{T}_n)$. Статистика Спірмена ρ визначається тепер як коефіцієнт кореляції двох множин рангів $(\mathrm{R}_1,...,\mathrm{R}_n)$ і $(\mathrm{S}_1,...,\mathrm{S}_n)$:

$$\rho = \frac{\sum_{i=1}^{n} (R_i - \overline{R})(S_i - \overline{S})}{[\sum_{i=1}^{n} (R_i - \overline{R})^2 \sum_{i=1}^{n} (S_i - \overline{S})^2]^{\frac{1}{2}}},$$

(тут $\overline{R}$ і $\overline{S}$ - відповідні арифметичні середні). Але $(\mathrm{R}_1,...,\mathrm{R}_n)$ і $(\mathrm{S}_1,...,\mathrm{S}_n)$ - деякі перестановки множин $(1,2,...,\mathrm{n})$, тому

$$\overline{R} = \overline{S} = \frac{1}{n}\sum_{i=1}^{n} i = \frac{n+1}{2},$$

$$\sum_{i=1}^{n} (R_i - \overline{R})^2 = \sum_{i=1}^{n} (S_i - \overline{S})^2 = \sum_{i=1}^{n} i^2 - n(\frac{n+1}{2})^2 = \frac{n(n^2-1)}{12},$$

тому що

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}.$$

Звідси

$$\rho = \frac{12}{n(n^2-1)} \sum_{i=1}^{n} (R_i - \frac{n+1}{2})(\mathrm{S}_i - \frac{n+1}{2}) = \frac{12}{n(n^2-1)} \sum_{i=1}^{n} (i - \frac{n+1}{2})(T_i - \frac{n+1}{2}).$$

Таким чином, ρ- лінійна функція рангів T_i. Часто використовують формули

$$\rho = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (R_i - \mathrm{S}_i)^2 = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (T_i - i)^2.$$

При повній відповідності рангів $(R_i = S_i, i = 1,...,n)$ статистика $\rho = 1$, а при протилежних рангах $(T_i = n - i + 1,\ i = 1,...,n)$ $\rho = -1$,

взагалі ж $-1 \le \rho \le 1$. Далі, якщо гіпотенуза H_0 вірна, то $E_\rho = 0$, $D_\rho = \dfrac{1}{(n-1)}$, тому значення ρ, близькі до крайніх, розглядають як ті що свідчать проти H_0, і, отже, критичну область критерію Спірмена задають у вигляді $\tau_\alpha = \{|\,\rho\,| > \mathrm{t}_\alpha(\mathrm{n})\}$. Для визначення числового значення критичної грані $\mathrm{t}_\alpha(\mathrm{n})$ при заданих об'єму вибірки n і рівні значимості α використовують таблиці табулірованного розподілу статистики ρ при нульовій гіпотезі, розраховані для $n = 2,...,30$. Для великих n можна скористатися асимптотичним результатом $\mathcal{L}\left(\sqrt{n}\rho \,|\, H_0\right) \to N(0,1)$. Звідси випливає, що якщо вибрати $\mathrm{t}_\alpha(\mathrm{n}) = \dfrac{t_{\frac{\alpha}{2}}}{\sqrt{n}}, \phi(-\mathrm{t}_{\frac{\alpha}{2}}) = \dfrac{\alpha}{2}$,то при великих n:

$$P\{\rho \in \tau_\alpha \,|\, H_0\} = \mathrm{P}\{\sqrt{n}\rho > \tau_{\frac{\alpha}{2}} \,|\, H_0\} \approx 2\phi(-\mathrm{t}_{\frac{\alpha}{2}}) = \alpha ,$$

тобто рівень значимості критерія приблизно дорівнює α.

ВИСНОВКИ

В результаті даної роботи було розглянуто теоретичні аспекти наступних тем: «Граничні теореми теорії ймовірностей», «Перевірка статистичних гіпотез», розв'язано задачі за темами: «Граничні теореми теорії ймовірностей», перевірено гіпотезу за допомогою критерію Спірмена.

У сучасній галузі інтелектуального аналізу даних, методи математичної статистики використовуються для пошуку прихованих нетривіальних зв'язків між величинами, які інтуїтивно помітити складно або навіть неможливо. Не менш вдало методи математичної статистики застосовують в організації виробництва, радіотехніці, теорії автоматичного керування, статистичній фізиці, розв'язанні теоретичних та практичних завдань кібернетики.

Результати курсової роботи можуть бути застосовані для подальшого вивченні розвитку теорії ймовірностей та математичної статистики, а також для вирішення прикладних задач статистичного аналізу даних.

ПЕРЕЛІК ПОСИЛАНЬ

1. Ивченко Г.И., Медведев Ю.И. Введение в математическую статистику. – М.: Изд-во ЛКИ, 2010. – 600 с.

2. Ивченко Г.И., Медведев Ю.И. Математическая статистика. – М.: Высш. шк., 1984. – 248 с.

3. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2006. – 816 с.

4. Математическая статистика / В.Б. Горяинов, И.В. Павлов, Г.М. Цветкова и др.; Под ред. В.С. Зарубина, А.П. Крищенко. – Изд. 3-е, испр. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2008. – 424 с.

5. Боровков А. А. Теория вероятностей. – М.: Наука, 1986. – 432 с.

6. Гихман И. И., Скороход А. В., Ядренко М. И. Теория вероятностей и математическая статистика. – К.: Выща школа, 1979. – 440 с.

7. Гнеденко Б. В. Курс теории вероятностей. – М.: Изд–во ЛКИ, 2007. – 448 с.

8. Розанов Ю. А. Теория вероятностей, случайные процессы и математическая статистика. – М.: Наука, 1989. – 320 с.

9. Ширяев А. Н. Вероятность. – 2–е изд., перераб. и доп. – М.: Наука, 1989. – 640 с.

10. Жлуктенко В. І., Наконечний С. І. Теорія ймовірностей і математична статистика. У 2 ч. – К.: КНЕУ, 2000. – 304 с.

11. Вадзинский Р. Н. Справочник по вероятностным распределениям. – Спб.: Наука, 2001. – 295 с.

12. Хастингс Н., Пикок Дж. Справочник по статистическим распределениям. – М.: Статистика, 1980. — 95 с.

Додаток А