ЗАКОН ВЕЛИКИХ ЧИСЕЛ ТА ЦЕНТРАЛЬНА ГРАНИЧНА ТЕОРЕМА

План

1. Відхилення відносної частоти від постійної ймовірності у незалежних дослідах.

2.Закон великих чисел.

1. **Відхилення відносної частоти від постійної ймовірності у незалежних дослідах**

Як відомо, не можна завчасно впевнено передбачити, яке із можливих значень прийме випадкова величина у досліді; це залежить від багатьох випадкових причин, врахувати які неможливо.

Здається на перший погляд, що встановити закономірності поведінки випадкових величин неможливо.

Але за деяких порівняно широких умов сумарна поведінка досить великого числа випадкових величин (за великої кількості дослідів) майже втрачає випадковий характер і набуває рис закономірності.

Важливе значення має знання умов, за яких сукупний вплив дуже великої кількості випадкових причин приводить до результату, що майже не залежить від випадку, оскільки дає можливість передбачати хід подій. Ці умови і вказуються у граничних теоремах.

Граничні теореми встановлюють граничні закони розподілу випадкових величин. Їх об'єднують під загальною назвою – закону великих чисел.

Закон великих чисел відіграє велику роль у різних процесах, пов'язаних з масовим виробництвом.

До граничних теорем теорії ймовірностей відносять локальну та інтегральну теореми Лапласа.

1. Відхилення відносної частоти від постійної ймовірності у незалежних дослідах

Нехай проводиться n незалежних дослідів в однакових умовах, причому у кожному з них з ймовірністю $p \in [0;1]$ може з'явитися подія A.

Тоді ймовірність того, що в n проведених дослідах абсолютна величина відхилення відносної частоти появи події A від ймовірності появи цієї події не перевищує деякого числа $\varepsilon \geq 0$, приблизно дорівнює подвоєному значенню інтегральної функції Лапласа у точці , тобто:

$$P_n\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{p(1-p)}}\right).$$

Приклад 8.1.

Визначити скільки необхідно провести дослідів з підкиданням монети , щоб з ймовірністю 0,92 можна було стверджувати, що частота випадання «герба» відхиляється від теоретичної ймовірності $p = 0{,}5$ на абсолютну величину, меншу, ніж 0,02.

<u>Алгоритм розв'язку.</u>

1) скористаємось записаною вище формулою:

$$2\Phi\left(\varepsilon\sqrt{\frac{n}{p(1-p)}}\right) = P_n$$

Підставимо значення $P_n = 0{,}92;\ \varepsilon = 0{,}02;\ p = 0{,}5$.

Маємо:

$$2\Phi\left(0{,}02\sqrt{\frac{n}{0{,}5(1-0{,}5)}}\right)=0{,}92.$$

Звідки $\Phi(z)=0{,}92/2=0{,}46$.

Знаходимо значення z з таблиць інтегральних функцій Лапласа (за значенням функції знаходимо аргумент) $z\approx 1{,}76$.

З іншої сторони, використовуючи знову вище записану формулу, маємо:

$$\varepsilon\sqrt{\frac{n}{p(1-p)}}=z \quad \text{і} \quad 0{,}02\sqrt{\frac{n}{0{,}5(1-0{,}5)}}=1{,}76.$$

Розв'яжемо рівняння відносно невідомої величини n:

$$1{,}76=\frac{0{,}02}{0{,}5}\sqrt{n}; \qquad \sqrt{n}=\frac{1{,}76\cdot 0{,}5}{0{,}02}=44$$

З отриманого рівняння знаходимо $n=1936$

Відповідь:

Необхідно провести щонайменше 1936 підкидань монети.

Приклад 8.2.

Перевіряється стадо теличок ВРХ 475 голів. Відомо, що ймовірність того, що теличка може бути вибракувана, дорівнює 0,05 (p). З ймовірністю 0,94 (P)визначити межі, в яких буде знаходитись число (m) вибракуваних теличок серед усіх теличок, що перевіряються.

<u>Алгоритм розв'язку.</u>

Знову скористаємось формулою:

$$P_n\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)\approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{p(1-p)}}\right).$$

Тоді:

$$2\Phi\left(\varepsilon\sqrt{\frac{475}{0{,}05(1-0{,}05)}}\right)\approx 0{,}94 \quad \text{і} \quad \Phi(z)\approx 0{,}47.$$

З таблиць інтегральних функцій Лапласа знаходимо аргумент z.

Маємо: $z=1{,}89$.

З рівняння $\varepsilon\sqrt{\frac{475}{0{,}05(1-0{,}05)}}=1{,}89$ знаходимо, що величина $\varepsilon=0{,}0189$.

Знаходимо межі числа можливо вибракуваних теличок знаходимо з нерівності:

$$p-\varepsilon \le \frac{m}{n} \le p+\varepsilon.$$

$$n(p-\varepsilon) \le m \le n(p+\varepsilon)$$

$$475\cdot(0{,}05-0{,}0189) \le m \le 475\cdot(0{,}05+0{,}0189)$$

$$14{,}77 \le m \le 32{,}73$$

Оскільки відповідь має бути в цілих числах (адже це кількість теличок), то маємо:

$$15 \le m \le 33.$$

1. Закон великих чисел

Під **"законом великих чисел"** в теорії ймовірностей мають на увазі ряд математичних теорем, які дозволяють не лише здійснити прогнози в області випадкових явищ, а й оцінити точність цих прогнозів. Розглянемо деякі з них.

Нерівність Чебишева

Лема Чебишева

Якщо випадкова величина $\boldsymbol{X}$, математичне сподівання якої $\boldsymbol{m_x}$, може приймати лише невід'ємні значення, то для довільного додатного числа $\boldsymbol{\alpha}$ має місце нерівність

$$P(X \geq \alpha) \leq \frac{m_x}{\alpha}.$$

Нерівність Чебишева

Нерівність П.Л.Чебишева справедлива, як для дискретних, так і неперервних випадкових величин.

Теорема 8.1. Ймовірність того, що відхилення випадкової величини $\boldsymbol{X}$ від її математичного сподівання за абсолютною величиною, менше додатного числа $\boldsymbol{\varepsilon}$, не менше, ніж $1-\dfrac{D(X)}{\varepsilon^2}$:

тобто:

$$P(|X-m_x|<\varepsilon)>1-\frac{D(X)}{\varepsilon^2} \quad \text{або}$$

$$P(|X-m_x|\geq\varepsilon)\leq\frac{D(X)}{\varepsilon^2}$$

Як бачимо з формули, нерівність Чебишева дозволяє оцінити ймовірність відхилення випадкової величини $\boldsymbol{X}$ від її математичного сподівання, знаючи лише її дисперсію $\boldsymbol{D(X)}$, а тому має велике практичне значення.

Приклад 8.3.

Використовуючи нерівність Чебишева, оцінити ймовірність того, що випадкова величина $\boldsymbol{X}$ відхилиться від свого математичного сподівання менше ніж на три середніх квадратичних відхилень.

Алгоритм розв'язку.

Згідно з умовою задачі $\varepsilon = 3 \cdot \sigma$, а $D(X) = \sigma^2$. Підставляючи ці параметри в нерівність, маємо:

$$P(|X - M(X)| < 3 \cdot \sigma) = 1 - \frac{\sigma^2}{9\sigma^2} = 1 - \frac{1}{9} = \frac{8}{9}.$$

Цей приклад показує, що ймовірність відхилення, що перевищує 3σ для довільного розподілу не є малою величиною (дорівнює 1/9).

Приклад 8.4.

Проводиться дослідження ваги зерен пшениці. Відомо, що середньоквадратичне відхилення ваги зерен не перевищує 0,02 г.

З усієї партії пшениці, що надійшла на елеватор від деякого господарства відібрано 100 зерен.

Яка ймовірність того, що середня вага зерен, обчислена за даною вибіркою, буде відрізнятись від середньої ваги зерна усіє партії на величину, меншу 0,004 г?

Алгоритм розв'язку

Вага зерна є випадковою величиною. Позначимо вагу зерна пшениці через X. Уся партія зерен є генеральною сукупністю.

Середню вагу зерна у партії приймемо за математичне сподівання цієї випадкової величини m_x. За умовою задачі середньоквадратичне відхилення дорівнює $\sigma_x = 0{,}02$.

Відібрані для зважування зерна є вибірковою сукупністю об'єму 100. Замітимо, що математичне сподівання m_x і середнє значення $\bar{x}$ нам невідомі.

Відносно їх різниці ставиться умова, що вона має бути меншою 0,004 г. Треба визначити ймовірність такого відхилення, тобто:

$$P(|\bar{x} - m_x| < 0{,}004).$$

За нерівністю Чебишева можна записати, що шукана ймовірність більше, ніж $1-\delta$,

де $\delta=\frac{\sigma^2}{n\cdot(\varepsilon)^2}=\frac{(0,02)^2}{100\cdot(0,004)^2}=0,25$.

Тоді $1-\delta=1-0,25=0,75$. Тобто шукана ймовірність відхилення знайдена і вона більша, ніж 0,75.

Приклад 8.5.

Умова попередньої задачі. Але вимагається знайти ймовірність того, що взята навмання одна зернина має вагу, яка відрізняється від математичного сподівання на величину, меншу, ніж 0,05 г.

Алгоритм розв'язку

Вага навмання взятої зернини є випадковою величиною X. Для Алгоритм розв'язку задачі знову скористаємось нерівністю Чебишева:

$P(|X-m_x|<0,05)>1-\delta$, де $\delta>0$.

Розрахуємо величину δ:

$$\delta=\frac{D(X)}{\varepsilon^2}=\frac{\sigma^2}{\varepsilon^2}=\frac{(0,02)^2}{(0,05)^2}=0,16.$$

Звідси $1-\delta=1-0,16=0,84$.

У такий спосіб, шукана величина ймовірності відхилення ваги, взятої навмання зернини, від математичного сподівання ваги зернини більша, ніж 0,84.

Нерівність Чебишева для практики має обмежене значення, оскільки часто дає грубу, а інколи і тривіальну оцінку.

Теоретично ж значення нерівності Чебишева досить велике. Воно використовується для виводу інших теорем, наприклад, теореми Чебишева.

Узагальнена теорема Чебишева

Теорема 8.2. Узагальнена теорема Чебишева

Якщо $X_1, X_2, \dots, X_n$ - попарно незалежні випадкові величини, причому дисперсії їх рівномірно обмежені (не перевищують постійного числа C), то яким би малим не було додатне число ε, ймовірність нерівності

$$\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n}\right| < \varepsilon$$

буде як завгодно близькою до одиниці, якщо число випадкових величин досить велике.

Іншими словами:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n}\right| < \varepsilon\right) = 1,$$

тобто відхилення середньої арифметичної випадкової величини від судньої арифметичної їх математичних сподівань буде (за абсолютною величиною) як завгодно малим.

Таким чином, хоч незалежні випадкові величини можуть набувати значень, далеких від своїх математичних сподівань, середнє арифметичне досить великого числа випадкових величин з досить великою ймовірністю набуває значень, близьких до визначеного постійного числа, а саме до числа:

$$\frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n} = a.$$

Іншими словами: окремі випадкові величини можуть мати значний розкид, то їх середнє арифметичне розсіяне мало.

Таким чином, не можна впевнено передбачити, якого можливого значення набуде кожна із випадкових величин, але можна передбачити, якого значення набуде їх середнє арифметичне.

Таким чином, середнє арифметичне досить великого числа незалежних випадкових величин, дисперсії яких рівномірно обмежені, втрачає характер випадкової величини.

Це пояснюється тим, що відхилення кожної з випадкових величин від математичного сподівання (m_x) можуть бути як додатними, так і від'ємними, але у середньому значенні вони гасяться.

$$\left|\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}-a\right|<\varepsilon\ .$$

Це правило має практичне застосування при вимірюваннях. При досить великих n середнє арифметичне значення наближається до істинної величини випадкової величини.

Але слід пам'ятати, що вимірні прилади мають точність вимірювання $\pm\alpha$; тому кожний результат, а значить і їх середнє значення матиме точність, що не перевищує точність приладу.

І, яким би не було великим число вимірювань n, ця точність лишиться незмінною.

На теоремі Чебишева базується широко застосовуваний у статистиці вибірковий метод, суть якого полягає в тому, що по порівняно невеликій вибірці роблять судження про усю генеральну сукупність і тому має велике практичне застосування.

Теорема Якоба Бернуллі

Вперше була теорема доведена Я. Бернуллі і надруковане доведення у 1713р. Доведення було складним, простіше її доведення було зроблене П.Л.Чебишевим у 1846 р.

Теорема 8.3. Якщо в кожному із $\boldsymbol{n}$ незалежних випробувань ймовірність $\boldsymbol{p}$ появи події $\boldsymbol{A}$ стала (схема Бернуллі), то як завгодно близька до одиниці ймовірність того, що відхилення відносної частоти $\frac{\boldsymbol{m}}{\boldsymbol{n}}$ від ймовірності $\boldsymbol{p}$ за абсолютною величиною буде як завгодно малим, якщо число випробувань достатньо велике.

Тобто якщо замінити $\frac{1}{n}\sum_{i=1}^{n} X_i$ на $\frac{m}{n}$, а $\frac{1}{n}\sum_{i=1}^{n} M(X_i)$ на $\boldsymbol{p}$ в попередній формулі, отримаємо:

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1.$$

Іншими словами: при збільшенні кількості дослідів ймовірність того, що відхилення частоти від ймовірності буде як завгодно малим, прагне до одиниці.

При збільшенні числа дослідів частота події збігається по ймовірності.

Теорема Бернуллі пояснює, чому відносний частоті за досить великого числа дослідів притаманна властивість стійкості, і вона оправдує статистичне визначення ймовірності.

Схема Бернуллі є математичною моделлю серії випробувань, що повторюються за однакових умов. У кожному випробуванні може настати подія $\boldsymbol{A}$, яку ми назвали успіхом.

Згідно з теоремою Бернуллі частота $\frac{m}{n}$ настання події A наближається до її ймовірності p при досить великій кількості дослідів.

Теорема Пуассона (Узагальнена Бернуллі)

Проводиться n незалежних дослідів, у кожному досліді ймовірність появи події A своя і дорівнює p_i. Тоді справедлива нерівність:

$$P\left(\left|p^* - \frac{\sum_{i=1}^{n} p_i}{N}\right| < \varepsilon\right) > 1 - \delta \quad .$$

Тобто $p^* \to \frac{1}{N}\sum_{i=1}^{n} p_i(A)$ – частота збігається по ймовірності.

Центральна гранична теорема Ляпунова

Якщо випадкова величина X є сумою дуже великого числа взаємно незалежних випадкових величин, вплив кожної з яких на всю суму надто малий, то X має розподіл ймовірностей дуже близький до нормального:

$$\lim_{n\to\infty} P\left(\frac{S_n - A_n}{B_n} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz \ ,$$

де

$$S_n = \sum_{s=1}^{n} X_s \ , \ A_n = \sum_{k=1}^{n} a_k \ , \ B_n^2 = \sum_{k=1}^{n} b_k^2 \ ,$$

тобто $A_n = M(X_k)$, а $B_n^2 = D(X_k)$.

За умови, що виконується умова рівномірної малості величин, що утворюють суму

$$\cdot \ \frac{1}{B_n^3} \cdot \sum_{i1}^{n} M\left(X_i\right)^3 \to 0$$