ПРАКТИЧНЕ ЗАНЯТТЯ № 4. (*Додаток*).
Апроксимація Муавра-Лапласа.

1. «Масові» явища зі сталою ймовірністю «УСПІХУ» (p). 2. Теорема Муавра-Лапласа. 3. Використання функції Лапласа.

1. *«Масові» явища зі сталою ймовірністю «УСПІХУ»* (*p*).

Приклад 1. Кидаємо $n = 720$ разів гральний кубик. Знайти ймовірність того, що кількість отриманих при цьому результатів **«6»** буде в межах від **110** до **135**.

Розв'язок. Отримання результату «6» при киданні грального кубика будемо називати «УСПІХОМ». Оскільки:

$$p = P\{\text{«УСПІХУ»}\} = P\{\text{Отримано результат «6»}\} = \frac{1}{6},$$

а результати чергових спроб не впливають один на одного, то випадкова величина ξ, яка показує, скільки разів отримано результатів «6» при $n = 720$ спробах має *біноміальний розподіл з параметрами* $(720, \frac{1}{6})$:

$$\xi \Leftrightarrow B(720, \frac{1}{6}).$$

Тому ймовірність випадкової події:

$$P\{\text{Кількість отриманих результатів «6» буде в межах від 110 до 135}\} =$$

$$= \mathrm{P}\{110 \le \xi \le 135\} = \sum_{k=110}^{135} C_{720}^{k} \cdot \left(\frac{1}{6}\right)^{k} \cdot \left(\frac{5}{6}\right)^{720-k}.$$

Подібно, як і в прикладі з *апроксимацією Пуассона*, отримали ситуацію, коли маємо *формулу точної відповіді*, скористатися якою на практиці неможливо, а точніше: безпосереднє обчислення *теоретично* можливе, але позбавлене *практичного сенсу*. У випадку, коли n дуже велике, навіть точні формули непридатні до практичного застосування. Між цими прикладами є однак суттєва різниця:

- Якщо в прикладі з *апроксимацією Пуассона* «УСПІХ» можна було вважати «*рідкою*» подією ($p = 0{,}01$), то в прикладі 1 (апроксимація Муавра-Лапласа) – це досить «*регулярна*» подія ($p = \frac{1}{6}$).

Варто підкреслити, що не існує чітко визначеної межі, яка встановлює, коли подія є «*рідкою*», а коли ні. Існує, однак умовний поділ вибірок на «*малі*» ($n \le 30$) та «*чисельні*» ($n > 30$). Тому, по аналогії, можна вважати подію «*рідкою*», якщо $p < 0{,}03$. В прикладі 1 (апроксимація Муавра-Лапласа) ймовірність «У» $p = 1/6$ є досить великою, щоб можна було використати апроксимацію Пуассона.

Якщо $\xi \Leftrightarrow B(n, p)$, то незалежно від величини ймовірності p, очікувана кількість «УСПІХІВ» дорівнює ($m_n = n \cdot p$). Тому можна виділити дві різні з математичної точки зору ситуації:

1) Спостерігаємо «*рідку*» подію в масовому явищі. Тоді очікувана кількість «УСПІХІВ» буде залишатись на більш-менш сталому, не дуже великому рівні ($\xi \approx m_n = n \cdot p$), навіть якщо кількість випробувань дуже велика.

2) Якщо ж ймовірності p значна, то збільшення кількості випробувань призводить до *пропорційного* збільшення кількості «УСПІХІВ».

Якщо в першому випадку ефективним інструментом є апроксимація Пуассона, то в другому – апроксимація *Муавра-Лапласа*.

2. *Теорема Муавра-Лапласа*.

Припустимо, подібно як і в теоремі Пуассона, що кількість випробувань (n) необмежено зростає: ($n \to \infty$). Ймовірність «УСПІХУ» (на відміну від теореми Пуассона) є *сталим, дійсним числом* ($0 < p < 1$), тобто не залежить від кількості випробувань (n).

Тоді для кожного n кількість «УСПІХІВ» буде випадковою величиною ξ_n, яка має біноміальний розподіл з параметрами (n, p):

$$\xi_n \Leftrightarrow B((n, p).$$

Оскільки ймовірність «УСПІХУ» p не змінюється, то одночасно з необмеженим зростанням ($n \to \infty$) кількості випробувань, зростає необмежено також кількість «УСПІХІВ» ($\xi_n \to \infty$).

Тобто для будь-якого числа $k = 0, 1, 2, \ldots$ ймовірність:

$$p_n(k) = P\{\xi_n = k\} = C_n^k \cdot (p)^k \cdot (1-p)^{n-k} \to 0,$$

якщо $n \to \infty$.

Математичне сподівання $m_n = E(\xi_n)$ (що характеризує середнє значення кількості «УСПІХІВ») дорівнює: $m_n = E(\xi_n) = n \cdot p \to \infty$, теж необмежено зростає. Можна тоді сподіватися, що при $n \to \infty$ різниця

$$\xi_n - n \cdot p$$

буде «*коливатися*» навколо *нуля*. Оскільки дисперсія $\sigma_n^2 = D(\xi_n)$ кількості «УСПІХІВ» дорівнює: $\sigma_n^2 = D(\xi_n) = n \cdot p \cdot (1-p)$, то «*в основному*» ці «*коливання*» будуть відбуватися в межах «*стандартного відхилення*»:

$$\left[-\sqrt{D(\xi_n)}; \sqrt{D(\xi_n)}\right] = \left[-\sqrt{n \cdot p \cdot (1-p)}; \sqrt{n \cdot p \cdot (1-p)}\right].$$

Ці міркування пояснюють сутність та дають фізичну інтерпретацію апроксимації *Муавра-Лапласа* для біноміального розподілу, що визначається наступною теоремою.

Теорема (*Муавра-Лапласа*). Рівномірно відносно довільних дійсних чисел $a < b$ виконується наступна рівність:

$$\lim_{n\to\infty} P\left(a < \frac{(\xi_n - n \cdot p)}{\sqrt{n \cdot p \cdot (1-p)}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx,$$

тобто:

$$\lim_{n\to\infty} P\left(n \cdot p + a \cdot \sqrt{n \cdot p \cdot (1-p)} \le \xi_n < n \cdot p + b \cdot \sqrt{n \cdot p \cdot (1-p)}\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

3. *Використання функції Лапласа.*

Визначення. Функцією Лапласа (або *інтегралом* Лапласа) називається функція, що визначається наступною формулою:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du, \ -\infty < x < \infty,$$

Враховуючи це визначення можемо записати теорему Муавра-Лапласа наступним чином:

- Якщо кількості випробувань n є «*достатньо великою*», то для довільних дійсних чисел $a < b$ виконується наступна *приблизна* рівність:

$$P\left(a \le \frac{\xi_n - np}{\sqrt{np(1-p)}} < b\right) \approx \Phi(b) - \Phi(a).$$

Або рівнозначно:

$$P\left(n \cdot p + a \cdot \sqrt{n \cdot p \cdot (1-p)} \le \xi_n < n \cdot p + b \cdot \sqrt{n \cdot p \cdot (1-p)}\right) = \Phi(b) - \Phi(a).$$

Приклад 1 (продовження). Кидаємо $\boldsymbol{n}$ **= 720** разів гральний кубік. Знайти ймовірність того, що кількість отриманих при цьому результатів **«6»** буде в межах від **110** до **135**.

Розв'язок. Ми встановили, що випадкова величина ξ, яка показує, скільки разів отримано результатів «6» при $n = 720$ спробах, має *біноміальний розподіл з параметрами* $(720, \frac{1}{6})$. Тому:

$$n = 720;\ p = 1/6,$$

$$1 - p = \frac{5}{6};\ n \cdot p = 720 \cdot \frac{1}{6} = 120;$$

$$n \cdot p \cdot (1 - p) = 120 \cdot \frac{5}{6} = 100.$$

Отже на підставі теореми Муавра-Лапласа:

$$\mathrm{P}\{110 < \xi < 135\} = P\left(\frac{110-120}{\sqrt{100}} < \frac{\xi - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} < \frac{135-120}{\sqrt{100}}\right) =$$

$$= P\left(-1 \le \frac{\xi - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} < 1{,}5\right) \approx \Phi(1{,}5) - \Phi(-1).$$

Використовуючи таблицю функції Лапласа отримуємо:

$$\Phi(1{,}5) = 0{,}9332.$$

$$\Phi(1) = 0{,}8413;\ \Phi(-1) = 1 - \Phi(1) = 1 - 0{,}8413 = 0{,}1587.$$

Тому остаточно:

$$\mathrm{P}\{110 < \xi < 135\} \approx \Phi(1{,}5) - \Phi(-1) = 0{,}9332 - 0{,}1587 = 0{,}7745.$$

(**Відповідь:** $\boldsymbol{P} \approx$ **0,7745**.)

Приклад 2. Ймовірність того, що в стохастичному експерименті відбудеться певна подія A дорівнює $\boldsymbol{p}$ **= 0,4**. Випадкова величина ξ_n визначає, скільки разів ця подія відбулася при n незалежних повторюваннях стохастичного експерименту.

Нехай $\nu_n = \frac{\xi_n}{n}$ означає частоту реалізації події A в n незалежних повторюваннях стохастичного експерименту.

Дати обґрунтовану відповідь на наступне питання:

- Скільки незалежних повторювань (n) стохастичного експерименту необхідно провести, щоб *принаймні* з ймовірністю **0,9** можна було б стверджувати, що частота ν_n події A буде відхилятися (в той чи інший бік) від ймовірності p події A не більше, ніж на **0,1**?

ЗАВДАННЯ 1)

Дати обґрунтовану відповідь на це питання, використовуючи з цією метою ***таблицю функції Лапласа***: $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$

Розв'язок. Умова:

> «*принаймні з ймовірністю* **0,9** *можна було б стверджувати, що частота* ν_n *події A буде відхилятися* (*в той чи інший бік*) *від ймовірності p події A не більше, ніж на* **0,1**»

в математичному (формальному) записі виглядає наступним чином:

- Необхідно визначити таке мінімальне значення кількості незалежних повторювань (n) стохастичного експерименту, щоб виконувалось співвідношення:

$$P(|\nu_n - p| < 0{,}1) \geq 0{,}9 .$$

Використовуючи визначення $\nu_n = \frac{\xi_n}{n}$ частоти події A можемо переписати це співвідношення наступним чином:

$$P\left(\left|\frac{\xi_n}{n} - p\right| < 0{,}1\right) \geq 0{,}9 .$$

$$P\left(-0{,}1 < \frac{\xi_n - n \cdot p}{n} < 0{,}1\right) \geq 0{,}9 .$$

$$P\left(-0{,}1 \cdot \frac{\sqrt{n}}{\sqrt{p \cdot (1-p)}} < \frac{\xi_n - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} < 0{,}1 \cdot \frac{\sqrt{n}}{\sqrt{p \cdot (1-p)}}\right) \geq 0{,}9 .$$

В прикладі 2, що розглядається, ймовірність «УСПІХУ» p = 0,4. Використовуючи теорему Муавра-Лапласа, а також міркування, наведені в попередньому прикладі 1, отримаємо:

$$P\left(-0{,}1 \cdot \frac{\sqrt{n}}{\sqrt{0{,}4 \cdot 0{,}6}} < \frac{\xi_n - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} < 0{,}1 \cdot \frac{\sqrt{n}}{\sqrt{0{,}4 \cdot 0{,}6}}\right) \approx$$

$$\approx \Phi\left(0{,}1 \cdot \frac{\sqrt{n}}{\sqrt{0{,}4 \cdot 0{,}6}}\right) - \Phi\left(-0{,}1 \cdot \frac{\sqrt{n}}{\sqrt{0{,}4 \cdot 0{,}6}}\right) =$$

$$= \Phi\left(0{,}1 \cdot \frac{\sqrt{n}}{\sqrt{0{,}4 \cdot 0{,}6}}\right) - \left(1 - \Phi\left(0{,}1 \cdot \frac{\sqrt{n}}{\sqrt{0{,}4 \cdot 0{,}6}}\right)\right) = 2 \cdot \Phi\left(0{,}1 \cdot \frac{\sqrt{n}}{\sqrt{0{,}4 \cdot 0{,}6}}\right) - 1.$$

Отже для визначення кількості незалежних повторювань n отримаємо наступну нерівність: $2\cdot\Phi\left(0{,}1\cdot\frac{\sqrt{n}}{\sqrt{0{,}4\cdot 0{,}6}}\right)-1\geq 0{,}9$. Або

$$\Phi\left(0{,}1\cdot\frac{\sqrt{n}}{\sqrt{0{,}4\cdot 0{,}6}}\right)\geq\frac{1+0{,}9}{2}=0{,}95.$$

Оскільки функції Лапласа є монотонно зростаючою, то можемо записати наступну нерівність: $0{,}1\cdot\frac{\sqrt{n}}{\sqrt{0{,}4\cdot 0{,}6}}\geq\Phi^{-1}(0{,}95)$. Використовуючи таблицю функції Лапласа отримаємо наступне співвідношення: $\Phi(1{,}64)=0{,}9495<0{,}95<0{,}9505=\Phi(1{,}65)$, або: $1{,}64<\Phi^{-1}(0{,}95)<1{,}65$. Тому:

$$0{,}1\cdot\frac{\sqrt{n}}{\sqrt{0{,}4\cdot 0{,}6}}>1{,}64;$$

або: $\sqrt{n}>\sqrt{0{,}4\cdot 0{,}6}\cdot 16{,}4$, чи $n>0{,}24\cdot(16{,}4)^2=64{,}55$.

Отже, щоб гарантувати виконання умови $P(|\nu_n-p|<0{,}1)\geq 0{,}9$ необхідно провести щонайменше $n=65$ незалежних повторювань стохастичного експерименту.

(**Відповідь: $n\geq 65$.**)

ЗАВДАННЯ 2)

Знайти розв'язок, використовуючи ***безпосереднє наближення***: $\xi_n\approx N(m_n;\ \sigma_n^2)$ біноміального розподілу нормальним розподілом $N(m;\ \sigma^2)$.

Обчислити ***наближене значення*** ймовірності:застосовуючи з цією метою статистичні функції («*НОРМ.РАСП*» та «*НОРМ. ОБР*») аркушу «***Excel***» для нормального розподілу з довільними параметрами $(m;\ \sigma^2)$.

Розв'язок. Використовуючи визначення $\nu_n=\frac{\xi_n}{n}$ частоти події A можемо переписати це співвідношення наступним чином:

$$P\left(\left|\frac{\xi_n}{n}-p\right|<0{,}1\right)\geq 0{,}9.\ \text{Або:}\ P\left(\left|\frac{\xi_n-n\cdot p}{\sqrt{n}}\right|<0{,}1\cdot\sqrt{n}\right)\geq 0{,}9.$$

Якщо випадкова величина ξ має розподіл $N(m;\ \sigma^2)$: $\xi\Leftrightarrow N(m;\ \sigma^2)$, то враховуючи його властивості, можна стверджувати:

- Випадкова величина $(\xi-m)$ буде мати розподіл $N(0;\ \sigma^2)$:

$$\xi-m\Leftrightarrow N(0;\ \sigma^2).$$

- Для випадкової величини $N(0;\ \sigma^2)$ справедлива наступна рівність:

$$P\{N(0;\ \sigma^2)<-a\}=P\{N(0;\ \sigma^2)>a\}$$

Отже: $P\{|N(0;\ \sigma^2)|<a\}=P\{-a<N(0;\ \sigma^2)<a\}=$

$$=P\{N(0;\ \sigma^2)<a\}-P\{N(0;\ \sigma^2)<-a\}=$$

$$=P\{N(0;\ \sigma^2)<a\}-1+P\{N(0;\ \sigma^2)<a\}=\mathbf{2}\cdot P\{N(0;\ \sigma^2)<a\}\mathbf{-1}.$$

Оскільки

$$m_n = E(\xi_n) = n \cdot p, \text{ а } \sigma_n^2 = D(\xi_n) = n \cdot p \cdot (1 - p),$$

то «*безпосереднім наближенням*» випадкової величини $\dfrac{\xi_n - n \cdot p}{\sqrt{n}}$ буде нормальний розподіл з параметрами $N(0; p\cdot(1 - p))$.

Нехай тепер $F_{(0,p)}(x)$ позначає функцію розподілу випадкової величини $N(0; p\cdot(1 - p))$.

$$F_{(0,p)}(x) = P\{N(0; p\cdot(1 - p)) < x\}, -\infty < x < \infty,$$

Тоді нерівність:

$$P\left(\left|\frac{\xi_n - n \cdot p}{\sqrt{n}}\right| < 0{,}1 \cdot \sqrt{n}\right) \geq 0{,}9\,,$$

означає, що

$$P\left(\left|\frac{\xi_n - n \cdot p}{\sqrt{n}}\right| < 0{,}1 \cdot \sqrt{n}\right) = 2 \cdot P\big(N(0, p \cdot (1 - p)) < 0{,}1 \cdot \sqrt{n}\big) - 1 =$$

$$= 2 \cdot F_{(0,p)}\left(0{,}1 \cdot \sqrt{n}\right) - 1 \geq 0{,}9,$$

Тобто

$$2 \cdot F_{(0,p)}\left(0{,}1 \cdot \sqrt{n}\right) \geq 1 + 0{,}9,$$

або

$$F_{(0,p)}\left(0{,}1 \cdot \sqrt{n}\right) \geq 0{,}95.$$

Звідси знаходимо:

$$0{,}1 \cdot \sqrt{n} \geq F^{-1}_{(0,p)}\left(0{,}95\right).$$

Оскільки

$$p\cdot(1 - p)) = 0{,}24,$$

то використовуючи статистичну функцію «*НОРМ.ОБР*» аркушу «*Excel*» для розподілу $N(0; 0{,}24)$, отримуємо:

$$F^{-1}_{(0,p)}\left(0{,}95\right) = 0{,}80581;$$

$$0{,}1 \cdot \sqrt{n} \geq 0{,}80581;$$

$$\sqrt{n} \geq 8{,}0581;$$

$$n \geq 64{,}55.$$

(Відповідь: $n \geq 65$.)