ЛЕКЦІЯ 7(А).
Нормальний розподіл та його властивості.

1. *Щільність нормального розподілу та її властивості*. 2. *Середнє значення нормального розподілу*. 3. *Дисперсія та стандартне відхилення нормального розподілу.*

З точки зору практичних застосувань теорії ймовірностей та математичної статистики нормальний розподіл є одним серед найважливіших, якщо не найважливішим.

Нормальний розподіл – це більше, ніж «*один з можливих*» розподілів теорії ймовірностей. Можливо він відображає якісь, ще не до кінця вивчені та зрозумілі, глибинні закономірності природи.

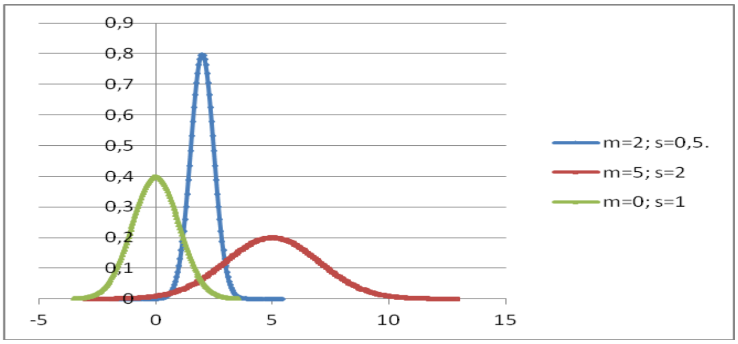
Це знаходить своє відображення в математичних властивостях цього розподілу, що й буде головною темою лекції.

1. Щільність нормального розподілу та її властивості.

Нормальний розподіл належить до неперервних випадкових величин а тому однозначно визначається своєю щільністю.

Визначення. Неперервна випадкова величина ξ має нормальний розподіл з параметрами $(m,\ \sigma^2)$: $\xi \Leftrightarrow N(m,\ \sigma^2)$, якщо її щільність $f_\xi(x)$ визначається формулою:

$$f_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}},\ -\infty < x < \infty.$$



Мал. 1. Приклади графіків нормального розподілу.

Відкриття цього розподілу пов'язують з іменами Лапласа (1749-1827) та Гауса (1777-1855). Але значно раніше Муавр довів збіжність біноміального розподілу до нормального та використовував його щільність в актуарних розрахунках.

Прийнято позначати нормальний розподіл символом $N(m,\ \sigma^2)$, підкреслюючи цим, що він залежить від двох параметрів. Таким чином маємо справу з цілою *родиною* подібних до себе розподілів. Якщо вибрати конкретні числові значення для цих параметрів, то тим самим визначимо цілком конкретну випадкову величину ξ, яка коротко позначається символом $\xi \Leftrightarrow N(m, \sigma^2)$. Малюнок 1 представляє графіки щільності нормального розподілу для різних комбінацій параметрів (m, σ^2).

- Для параметрів ($m = 2$, $\sigma^2 = 0{,}25$) графік функції $f_\xi(x)$ буде найбільш *витягнутим*. Він спочатку *стрімко* зростає до значення аргументу $x = 2$, а потім так же *стрімко* спадає.
- Для параметрів ($m = 0$, $\sigma^2 = 1$) графік, хоч і подібний до попереднього, але має дуже *плавні* форми, симетричний відносно осі OX. Ця версія нормального розподілу називається *стандартним законом*, або *розподілом* $N(0, 1)$.
- Найбільш *сплющена* крива представляє графік щільності нормального розподілу з параметрами ($m = 5$, $\sigma^2 = 4$).

Аналізуючи наведені графіки а також аналітичний вигляд щільності $f_\xi(x)$, встановимо декілька характеристичних для нормального розподілу властивості, що безпосередньо випливають з цих загальних особливостей щільності.

- Оскільки $f_\xi(x) > 0$ для будь-яких значень $-\infty < x < \infty$, то випадкова величина $\xi \Leftrightarrow N(m, \sigma^2)$, яка має нормальний розподіл з параметрами (m, σ^2) (незалежно від конкретних числових значень цих параметрів) може приймати *довільні дійсні значення* $x \in R$.
- В точці $x^* = m$ функція $f_\xi(x)$ досягає свого найбільшого значення, яке дорівнює:

$$f_\xi(x^*) = \max_{x \in R}\{f_\xi(x)\} = f_\xi(m) = \frac{1}{\sigma \cdot \sqrt{2\pi}}.$$

Отже *найбільш ймовірні* значення випадкової величини ξ, яка має нормальний розподіл з параметрами (m, σ^2), розташовані *поблизу* числа m. На практиці це означає, що спостереження випадкової величини ξ, як правило, групуються навколо значення $x = m$.

Серед кількісних характеристик статистичної ознаки є два типи *позиційних* мір середнього рівня, а саме:

- *Домінанта* (Mo) (або *мода*), тобто значення статистичної ознаки, яке *найчастіше* зустрічається в популяції;
- *Медіана* (Me), тобто *середнє* значення *варіаційного* ряду для статистичної ознаки.

2. Середнє значення нормального розподілу.

Показники, відповідні цим параметрам, існують і в теорії ймовірностей.

- Точка x^*, в якій значення щільності є найбільшим, називається *домінантою* ($Do(\xi)$) розподілу випадкової величини ξ.

Таким чином:

$$Do(N(m,\ \sigma^2)) = m.$$

- Аналізуючи зображення на малюнку 1 бачимо, що графік щільності нормального розподілу має форму «*дзвону*» та *симетричний* відносно вертикальної прямої $x = m$. Це означає, що відхилення значень випадкової величини ξ в той чи інший бік від неї однаково ймовірні.

В лекції 5 була встановлена наступна формула [**див. власт. 3**]:

$$P(c \le \xi < C) = \int_c^C f_\xi(x)dx.$$

Враховуючи ту обставину, що $\int_{-\infty}^{\infty} f_\xi(x)dx = 1$, а також симетричність щільності нормального розподілу, отримаємо:

$$P\{N(m,\ \sigma^2) < m\} = \int_{-\infty}^{m} f_\xi(x)dx = \int_{m}^{\infty} f_\xi(x)dx = \mathrm{P}\{N(m,\ \sigma^2) \ge m\} = 0{,}5.$$

- Точка $\hat{x}$, що ділить розподіл випадкової величини ξ навпіл, тобто:

$$P\{\xi < \hat{x}\} = P\{\xi \ge \hat{x}\} = 0{,}5.$$

називається *медіаною* (Me), тобто *середнім* значенням цього розподілу.

Таким чином

$$Me(N(m,\ \sigma^2)) = m.$$

Отже параметр m розподілу $N(m,\ \sigma^2)$ визначає для нього значення *позиційних мір* середнього рівня:

$$m = Me(N(m,\ \sigma^2)) = Do(N(m,\ \sigma^2)).$$

Переконаємось, що m є одночасно і *класичним середнім*, тобто математичним сподіванням розподілу $N(m,\ \sigma^2)$. Використовуючи визначення математичного сподівання неперервної випадкової величини ξ з щільністю $f_\xi(x)$ отримаємо:

$$E(\xi) = \int_{-\infty}^{+\infty} x \cdot f_\xi(x)dx = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} x \cdot e^{-\frac{(x-m)^2}{2\cdot\sigma^2}} dx =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} (x-m) \cdot e^{-\frac{(x-m)^2}{2\sigma^2}} dx + \frac{m}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

$z \cdot e^{-\frac{z^2}{2\sigma^2}}$ - парна функція, тому $\int_{-\infty}^{\infty} z \cdot e^{-\frac{z^2}{2\sigma^2}} dz = 0$, оскільки інтервал ($-\infty < z < \infty$) *симетричний* відносно точки $z = 0$.

З іншого боку:

$$\frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \int_{-\infty}^{+\infty} f_\xi(x)dx = 1.$$

Тому проводячи заміну змінних $z = x - m$ отримаємо:

$$E(\xi) = \int_{-\infty}^{+\infty} x \cdot f_\xi(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} z \cdot e^{-\frac{z^2}{2\sigma^2}} dz + m \cdot \int_{-\infty}^{+\infty} f_\xi(x)dx = m.$$

Отже параметр m визначає середнє значення розподілу $N(m, \sigma^2)$, причому:

$$m = E(N(m, \sigma^2)) = Me(N(m, \sigma^2)) = Do(N(m, \sigma^2)).$$

- В статистиці розподіл випадкової величини, який має таку властивість, називається *симетричним*.

3. Дисперсія та стандартне відхилення нормального розподілу.

Другу групу параметрів випадкової величини, подібно, як і в описовій статистиці, утворюють *міри змінності*.

Визначення. Дисперсією випадкової величини ξ називається математичне сподівання квадрату відхилення випадкової величини ξ від її математичного сподівання, тобто число $D(\xi)$, що визначається наступним чином:

$$D(\xi) = E(\xi - E(\xi))^2.$$

В результаті для дисперсії $D(\xi)$ виконуються всі необхідні умови, що дозволяють використати її в якості міри змінності, а саме:

- Дисперсії випадкової величини – це невід'ємне число $D(\xi) \geq 0$. Щоб підкреслити цю обставину, дисперсія часто позначається символом σ^2.
- Нульове значення дисперсії є свідченням *повної відсутності змінності*. Якщо дисперсія рівна 0, то випадкова величина зберігає постійне значення, яке дорівнює m.
- Більше значення дисперсії $D(\xi)$ означає «*сильніше*» відхилення значень випадкової величини від математичного сподівання m, а отже більшу її змінність.

Визначення дисперсії зберігають свою силу та фізичну інтерпретацію і у випадку неперервних випадкових величин. Єдине, що необхідно конкретизувати у випадку неперервних випадкових величин, так це *формулу для обчислення* дисперсії (σ^2) на підставі її щільності $f_\xi(x)$.

Тому, використовуючи визначення математичного сподівання, отримуємо: $\sigma^2 = D(\xi) = \int_{-\infty}^{\infty} (x-m)^2 \cdot f_\xi(x)dx$,

де

$$m = E(\xi) = \int_{-\infty}^{\infty} x \cdot f_\xi(x)dx.$$

Приклад. [***Дисперсія нормального розподілу***].

Аналізуючи максимум $f_\xi(m) = \dfrac{1}{\sigma\cdot\sqrt{2\pi}}$ щільності нормального розподілу $N(m, \sigma^2)$ бачимо, що він буде *більший*, чим *менша* величина параметр σ^2. І навпаки, як видно на малюнку 1, чим *більше* значення σ^2, тим графік щільності $f_\xi(x)$ буде більш *сплющений*. Форма кривої щільності вказує на рівень змінності відповідної випадкової величини: *витягнутий* графік відповідає розподілам з *невеликою* змінністю, а *плоский* - з *високою*. Отже приходимо до висновку:

- Чим *більше* значення σ^2, тим *вищий* рівень змінності розподілу $N(m, \sigma^2)$.

Доведемо, що ці міркування цілком обґрунтовані, а

$$\sigma^2 = D(N(m, \sigma^2)).$$

Якщо неперервна випадкова величина має щільність $f_\xi(x)$, $m = E(\xi)$, то її дисперсія обчислюється за формулою:

$$\sigma^2 = D(\xi) = \int_{-\infty}^{\infty}(x-m)^2\cdot f_\xi(x)dx .$$

Нехай $v = e^{-\frac{z^2}{2\sigma^2}}$, тоді: $D(N(m, \sigma^2)) = \dfrac{1}{\sigma\sqrt{2\pi}}\cdot\displaystyle\int_{-\infty}^{\infty}(x-m)^2\cdot e^{-\frac{(x-m)^2}{2\sigma^2}}dx =$

$$= \frac{1}{\sigma\sqrt{2\pi}}\cdot\int_{-\infty}^{\infty}z^2\cdot e^{-\frac{z^2}{2\sigma^2}}dz = -\sigma^2\cdot\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}zde^{-\frac{z^2}{2\sigma^2}} = -\sigma^2\cdot\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}zdv .$$

Так, як $\displaystyle\int_a^b zdv = z\cdot v\Big|_a^b - \int_a^b vdz$, то після простих перетворень отримаємо:

$$D(N(m, \sigma^2)) = -\sigma^2\cdot\frac{1}{\sigma\sqrt{2\pi}}\cdot\left(-\int_{-\infty}^{\infty}e^{-\frac{z^2}{2\sigma^2}}dz\right) = \sigma^2\cdot\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{z^2}{2\sigma^2}}dz = \sigma^2.$$

На практиці, коли маємо справу з величинами, яким відповідають конкретні одиниці міри, присутність «*квадрату*» у визначенні дисперсії створює деякі *технічні* незручності та ускладнює її безпосереднє використання в розрахунках. Оскільки квадратний корінь є монотонно зростаючою, невід'ємною функцією, то добуваючи корінь з дисперсії позбавляємось цих незручностей, зберігаючи при цьому всі необхідні властивості міри змінності.

Визначення. *Стандартним відхиленням* випадкової величини ξ називається число $S(\xi)$, що визначається наступним чином:

$$S(\xi) = \sqrt{D(\xi)} .$$