ЛЕКЦІЯ 7(Б).
Стандартний нормальний розподіл.

1. *Функція Лапласа*. 2. *Функція Лапласа в практичних розрахунках*. 3. *Правило «трьох сигм»*. 4. *Властивості нормального розподілу.*

1. Функція Лапласа.

Параметри $\boldsymbol{m}$ та $\boldsymbol{\sigma^2}$ однозначно визначають щільність $f_\xi(x)$ випадкової величини ξ, що має відповідний нормальний розподіл: $\xi \Leftrightarrow N(m, \sigma^2)$. А отже її функція розподілу: $F_\xi(x) = P\{\xi < x\}$ теж цілком визначається цими параметрами.

$$F_\xi(x) = P\{N(m, \sigma^2) < x\} = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{(u-m)^2}{2\sigma^2}} du, \ -\infty < x < \infty.$$

Нажаль безпосередньо скористатися з цієї формули (наприклад для підрахунку ймовірностей наступних подій:

$$P(c \le N(m, \sigma^2) < C) = F_\xi(C) - F_\xi(c))$$

практично неможливо, оскільки присутній в ній інтеграл неможливо виразити через елементарні функції.

В такому випадку, як правило, будують *спеціальні таблиці*, використовуючи з цією метою чисельні методи. Однак і такий підхід був би під знаком запитання, так як різноманітних комбінацій (m, σ^2) параметрів розподілу незлічена кількість. І фактично треба створювати *безліч* таблиць.

Однак у випадку нормального розподілу можна обійтись однією таблицею для одного «*спеціального*» набору параметрів, наприклад:

$$m = 0, \ \sigma^2 = 1.$$

Серед графіків, що представляють щільність нормального розподілу для різних комбінацій параметрів (m, σ^2) присутній і цей випадок. Підкреслювалось при цьому, що «*версія*» нормального $N(0, 1)$ розподілу називається «*стандартним законом*», графік його щільності в кшталті «*дзвону*» симетричний відносно осі OX та має дуже *плавні* форми.

Визначення. Нормальний розподіл з параметрами (0, 1) називається *стандартним*.

Випадкову величину, яка має $N(0, 1)$-розподіл, будемо позначати символом η. Її щільність $f_\eta(x)$ визначається рівністю:

$$f_\eta(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \text{ для } -\infty < x < \infty.$$

В різних розділах прикладної математики віддавна використовується спеціальна функція $\Phi(x)$, що має назву функції Лапласа:

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

Віддавна існують також впорядковані таблиці її значень для різних аргументів $x \in (-\infty, \infty)$. Приведемо для прикладу фрагмент такої таблиці.

Таблиця функції Лапласа.

u	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5861	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6518
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7637	0,7703	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8369
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9235	0,9251	0,9265	0,9278	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9648	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9777	0,9783	0,9788	0,9793	0,9798	0,9803	0,9807	0,9812	0,9816
2,1	0,9821	0,9825	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9853	0,9857
2,2	0,9861	0,9864	0,9867	0,9871	0,9874	0,9877	0,9880	0,9884	0,9887	0,9889
2,3	0,9892	0,9895	0,9898	0,9901	0,9903	0,9906	0,9908	0,9911	0,9913	0,9915
2,4	0,9918	0,9920	0,9922	0,9924	0,9926	0,9928	0,9930	0,9932	0,9934	0,9936
2,5	0,9937	0,9939	0,9941	0,9943	0,9944	0,9946	0,9947	0,9949	0,9950	0,9952
2,6	0,9953	0,9954	0,9956	0,9957	0,9958	0,9959	0,9960	0,9962	0,9963	0,9964
2,7	0,9963	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9972	0,9973
2,8	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9978	0,9979	0,9980	0,9980
2,9	0,9981	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986
3,0	0,9986	0,9986	0,9987	0,9987	0,9988	0,9988	0,9988	0,9989	0,9989	0,9990

Підставляючи значення параметрів ($m = 0$, $\sigma^2 = 1$) до функції розподілу $F_\xi(x)$ випадкової величини $\xi \Leftrightarrow N(m, \sigma^2)$, переконуємось, що функція Лапласа $\Phi(x)$ буде функцією розподілу стандартного нормального закону:

$$\Phi(x) = P\{N(0, 1) < x\}, -\infty < x < \infty.$$

2. Функція Лапласа в практичних розрахунках.

Тому, перш за все, переконаємось, що для виконання будь-яких пов'язаних з випадковою величиною $\xi \Leftrightarrow N(m, \sigma^2)$ підрахунків, цілком достатньо таблиць функції Лапласа. З цією метою доведемо такий факт:

- *Лінійне перетворення нормального розподілу залишає його нормальним, змінюючи тільки значення параметрів.*

Його доведення отримаємо, як наслідок наступної властивості нормального розподілу, що має *фундаментальне* значення як з теоретичної, так і практичної точок зору.

Лема 1. Припустимо, що між випадковими величинами ξ та η виступає наступна залежність:

$$\eta = \frac{\xi - m}{\sigma}.$$

Якщо ξ має нормальний розподіл з параметрами (m, σ^2):

$$\xi \Leftrightarrow N(m, \sigma^2),$$

то η буде мати нормальний стандартний розподіл:

$$\eta \Leftrightarrow N(0, 1).$$

Доведення. Встановимо спочатку зв'язок між функціями розподілу $F_\xi(x)$ та $F_\eta(x)$ відповідно випадкових величин ξ та η:

$$F_\eta(x) = P\{\eta < x\} = P\left\{\frac{\xi - m}{\sigma} < x\right\} = P\{\xi < m + \sigma \cdot x\} = F_\xi(m + \sigma \cdot x).$$

Використовуючи цю рівність, встановимо тепер зв'язок між щільністю $f_\xi(x)$ випадкової величини ξ та щільністю $f_\eta(x)$ випадкової величини η:

$$f_\eta(x) = [F_\eta(x)]'_x = [F_\xi(m + \sigma \cdot x)]'_x = F'_\xi(m + \sigma \cdot x) \cdot (m + \sigma \cdot x)'_x =$$

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

Тобто η є неперервною випадковою величиною з щільністю:

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

що й треба було довести.

Лема 2. Якщо випадкова величина η має нормальний *стандартний* розподіл, то для будь-яких дійсних чисел $\sigma \neq 0$ та m, розподіл випадкової величини:

$$\xi = m + \sigma \cdot \eta$$

буде нормальним з параметрами (m, σ^2): $\xi \Leftrightarrow N(m, \sigma^2)$.

Доведення випливає безпосередньо з леми 1.

Лема 3. Для довільного дійсного $a > 0$ виконується наступна рівність:

$$\Phi(-a) = 1 - \Phi(a).$$

Доведення. Нехай випадкова величина η має нормальний стандартний розподіл. Встановимо зв'язок між випадковими подіями $\{\eta < a\}$ та $\{\eta > a\}$. Оскільки щільністю $f_\eta(x)$ випадкової величини η – парна функція, тобто:

$$f_\eta(a) = f_\eta(-a),$$

то можемо записати наступні співвідношення:

$$\Phi(-a) = P\{\eta < -a\} = \int_{-\infty}^{-a} f_\eta(u)du = \int_{a}^{\infty} f_\eta(u)du =$$

$$= P\{\eta > a\} = 1 - P\{\eta < a\} = 1 - \Phi(-a).$$

Лема 4. Нехай випадкова величина η має нормальний стандартний розподіл. Тоді для довільного дійсного $A > 0$ виконується наступна рівність:

$$P\{|\eta| < A\} = 2 \cdot \Phi(A) - 1.$$

Доведення.

$$P\{|\eta| < A\} = P\{-A < \eta < A\} = P\{\eta < A\} - P\{\eta < -A\} =$$

$$= \Phi(A) - \Phi(-A) = \Phi(A) - 1 + \Phi(A) = 2 \cdot \Phi(A) - 1.$$

Лема 5. Нехай випадкова величина η має нормальний стандартний розподіл, а α – довільне число з інтервалу (0, 1). Якщо виконується рівність:

$$P\{|\eta| < A\} = 1 - \alpha,$$

то

$$\Phi(A) = 1 - \frac{\alpha}{2}.$$

Доведення. Дійсно: $1 - \alpha = P\{|\eta| < A\} = 2 \cdot \Phi(A) - 1$;

$$2 \cdot \Phi(A) = 2 - \alpha; \text{ тобто } \Phi(A) = 1 - \frac{\alpha}{2}.$$

2. Правило «*трьох сигм*».

Лема 6. Припустимо, що випадкова величина ξ має нормальний розподіл з параметрами (m, σ^2): $\xi \Leftrightarrow N(m, \sigma^2)$. Тоді ймовірність того, що випадкова величина ξ буде відхилятися від очікуваного значення m на одне, два або три *стандартні відхилення* σ відповідно дорівнює:

$$P\{m - \sigma < \xi < m + \sigma\} \approx 0{,}6826,$$

$$P\{m - 2\sigma < \xi < m + 2\sigma\} \approx 0{,}9544,$$

$$P\{m - 3\sigma < \xi < m + 3\sigma\} \approx 0{,}9972.$$

Доведення. Нехай $F_\xi(x) = P\{\xi < x\} = P\{N(m, \sigma^2) < x\}$ – функція розподілу випадкової величини ξ. Тоді:

$$P\{m - \sigma < \xi < m + \sigma\} = F_\xi(m + \sigma) - F_\xi(m - \sigma).$$

Випадкова величина:

$$\eta = \frac{\xi - m}{\sigma}$$

має нормальний стандартний розподіл:

$$\eta \Leftrightarrow N(0,\, 1).$$

Тому функція Лапласа $\varPhi(x)$ є її функцією розподілу:

$$\varPhi(x) = P\{\eta < x\}.$$

В свою чергу: $\xi = m + \sigma \cdot \eta$. Отже:

$$F_{\xi}(x) = P\{\xi < x\} = P\{m + \sigma \cdot \eta < x\} = P\left(\eta < \frac{x - m}{\sigma}\right) = \Phi\left(\frac{x - m}{\sigma}\right).$$

Тому:

$$P\{m - \sigma < \xi < m + \sigma\} = F_{\xi}(m + \sigma) - F_{\xi}(m - \sigma) = \varPhi(1) - \varPhi(-1) =$$
$$= \varPhi(1) - 1 + \varPhi(1) = 2\varPhi(1) - 1.$$

Подібним чином отримуємо:

$$P\{m - 2\sigma < \xi < m + 2\sigma\} = 2\varPhi(2) - 1.$$

$$P\{m - 3\sigma < \xi < m + 3\sigma\} = 2\varPhi(3) - 1.$$

З приведеної таблиці функція Лапласа знаходимо:

$$\varPhi(1) = 0{,}8413, \text{ тобто } 2\varPhi(1) - 1 = 0{,}6826;$$

$$\varPhi(2) = 0{,}9772, \text{ тобто } 2\varPhi(2) - 1 = 0{,}9544;$$

$$\varPhi(3) = 0{,}9986, \text{ тобто } 2\varPhi(3) - 1 = 0{,}9972.$$

що й доводить лему 6.

Рівність:

$$P\{m - 3\sigma < \xi < m + 3\sigma\} \approx 0{,}9972.$$

називається «*правилом трьох сигм*» і досить часто використовується на практиці в математичній статистиці.

Якщо статистична ознака має нормальний розподіл з параметрами m та σ^2), то при її вимірюванні *найбільш ймовірними* є значення ознаки навколо «*очікуваного*» значення m. Як випливає з доведеної леми, практично всі спостереження повинні бути зосереджені в інтервалі

$$[m - 3\sigma;\, m + 3\sigma],$$

що визначається *трьома стандартними відхиленнями* (σ) *від* «*очікуваного*» середнього значення m.

В середньому (*приблизно*) лише 28 із 10 000 отриманих спостережень можуть перетинати межі цього діапазону. Імовірність цієї події:

$$P\{|N(m,\, \sigma^2) - m| > 3\sigma\} \approx 0{,}0028$$

дуже мала, тому вона повинна *дуже рідко* реалізовуватись на практиці.

Варто однак зауважити, що для розподілів, відмінних від ненормального «*правило трьох сигм*» не обов'язково має бути вірним.

4. Властивості нормального розподілу

Вивчимо деякі корисні з практичної точки зору властивості нормального закону, які будуть широко використовуватись при побудові математичного інструментарію статистичного аналізу.

Лема 7. (*Лінійна функція нормального розподілу*.) Лінійна функція випадкової величини ξ, що має нормальний розподіл, також має нормальний розподіл, тобто, якщо:

$$\zeta = b + a\cdot\xi,$$

та

$$\xi \Leftrightarrow N(m, \sigma^2),$$

то

$$\zeta \Leftrightarrow N(b + a\cdot m, (a\cdot\sigma)^2).$$

Доведення. На підставі леми 1:

$$\eta = \frac{\xi - m}{\sigma} \Leftrightarrow N(0, 1).$$

Тобто:

$$\xi = m + \sigma\cdot\eta,$$

а

$$\zeta = b + a\cdot\xi = b + a\cdot(m + \sigma\cdot\eta) = (b + a\cdot m) + a\cdot\sigma\cdot\eta.$$

Тоді на підставі леми 2 приходимо до висновку, що

$$\zeta \Leftrightarrow N(b + a\cdot m, (a\cdot\sigma)^2).$$

🕮

Знаходження розподілу *сум випадкових величин* – це одна з основних проблем теорії ймовірностей. Розглянемо найпростіший випадок цієї задачі, що стосується *суми двох незалежних випадкових величин*.

Теорема 1. Припустимо, що ξ_1 та ξ_2 – незалежні випадкові величини, що мають неперервний розподіл, а $f_1(x)$ та $f_2(y)$ відповідні їм щільності. Тоді сума:

$$\eta = \xi_1 + \xi_2$$

теж буде неперервною випадковою величиною, щільність якої визначається наступним чином:

$$f_\eta(z) = \int_{-\infty}^{\infty} f_1(z-y)\cdot f_2(y)dx = \int_{-\infty}^{\infty} f_2(z-y)\cdot f_1(y)dx.$$

Доведення. Введемо двовимірну випадкову величину неперервного типу (ξ_1, ξ_2). Оскільки ξ_1 та ξ_2 – незалежні випадкові величини, то на підставі леми 1 (лекція 12) щільність $f(x, y)$ випадкової величини (ξ_1, ξ_2) визначається за формулою:

$$f(x, y) = f_1(x)\cdot f_2(y).$$

Тому використовуючи визначення функції розподілу, отримаємо:

$$F_\eta(z) = P\{\eta < z\} = P\{\xi_1 + \xi_2 < z\} = P\{\xi_2 < z - \xi_1\}.$$

Нехай D позначає множину точок площини, що визначаються рівністю:

$$D = \{(x, y): y < z - x\}.$$

В лекції 12 щільність приведена наступна формула:

- Якщо випадковий вектор $\xi = (\xi_1, \xi_2, \dots , \xi_n)$ має неперервний розподіл з щільністю $f_\xi(x_1, x_2, ..., x_n)$, то

$$P\{(\xi_1, \xi_2, ..., \xi_n) \in D\} = \iint\limits_D f(x_1,...x_n)dx_n...dx_1 .$$

для довільної підмножини $D \subset R_n$ n-вимірного простору. Отже:

$$F_\eta(z) = P\{\xi_2 < z - \xi_1\} = P\{(\xi_1, \xi_2) \in D\} =$$

$$= \iint\limits_D f(x,y)dydx = \int\limits_{-\infty}^{\infty} dx \int\limits_{z-x}^{\infty} f(x,y)dy = \int\limits_{-\infty}^{\infty} (1 - F_2(z-x)) \cdot f_1(x)dx .$$

Обчислюючи похідну відносно z в лівій та правій частині цієї рівності та враховуючи «*симетрію*» виразів, що розглядаються відносно випадкових величин ξ_1 та ξ_2, отримаємо доведення теореми.

🕮

Наступна властивість нормального розподілу, що має значний теоретичний і особливо практичний вимір, звучить наступним чином:

- *Сума незалежних нормально розподілених випадкових величин також має нормальний розподіл, параметри якого дорівнюють сумі відповідних параметрів окремих доданків*.

Сформулюємо її докладніше.

Лема 8. Якщо ξ_1 та ξ_2 – незалежні випадкові величини, що мають нормальний розподіл відповідно з параметрами $(m_1, \sigma^2{}_1)$ та $(m_2,\sigma^2{}_2)$: (m, σ^2):

$$\xi_1 \Leftrightarrow N(m_1, \sigma^2{}_1),\ \xi_2 \Leftrightarrow N(m_2, \sigma^2{}_2),$$

то сума:

$$\eta = \xi_1 + \xi_2$$

теж буде неперервною випадковою величиною, що має нормальний розподіл з параметрами $(m_1 + m_2, \sigma^2{}_1 + \sigma^2{}_2)$, тобто:

$$\eta \Leftrightarrow N(m_1 + m_2, \sigma^2{}_1 + \sigma^2{}_2).$$

Доведення. В справедливості твердження леми 8 можна переконатися проводячи безпосередні обчислення, використовуючи теорему 1. Пропонуємо зробити це самостійно.

🕮

Лема 9. (*Лінійна комбінація нормальних розподілів.*) Лінійна комбінація випадкових величин, що мають нормальний розподіл, також має нормальний розподіл. Тобто: якщо незалежні випадкові величини $\xi_1, \xi_2, \dots , \xi_n$ мають нормальний розподіл, $c_1, c_2, \dots , c_n$ – дійсні числа, то випадкова величина

$$\zeta = c_1 \cdot \xi_1 + c_2 \cdot \xi_2 + \dots + c_n \cdot \xi_n,$$

також буде мати нормальний розподіл.

В свою чергу, строге доведення леми 8, як і більш загальної властивості, сформульованої в лемі 9, яка поєднує результати лем 7 та 8, буде приведене в наступній лекції 13. Це доведення буде базуватися на використанні апарату *характеристичних функцій*, до вивчення яких зараз переходимо.