Лекція 4

Теорія статистичних гіпотез

Деякі розподіли функцій нормальних випадкових величин

1. Розподіл χ^2 (хі-квадрат) Пірсона

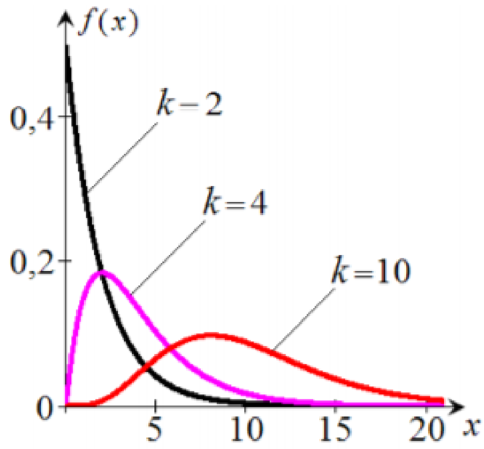
Розподілом χ^2 ***Пірсона*** з k ступенями свободи називається розподіл суми квадратів нормально розподілених незалежних випадкових величин X1, X2, ..., Xk з параметрами a = 0, σ = 1:

$$\chi^2(k) = X_1^2 + X_2^2 + \ldots + X_k^2.$$

Щільність цього розподілу визначається формулою

$$f_{\chi_k^2}(x) = \begin{cases} \dfrac{x^{\frac{k}{2}-1} \cdot e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \cdot \Gamma(k/2)}, & \text{при } x > 0, \\ 0, & \text{при } x \le 0, \end{cases}$$

$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$ – гамма-функція Ейлера (Γ(x)=(x -1)! для натуральних значень x).



Розподіл χ^2 визначається тільки одним параметром – числом ступенів свободи k. графіки функції $f_{\chi^2}(x)$ для різних значень k представлені на рис. Зі збільшенням
числа ступенів свободи k (k → ∞) розподіл χ^2 наближається до нормального закону розподілу (при k> 30 відмінностей практично немає).

Числові характеристики розподілу χ^2:

$$M[\chi^2] = k,\ D[\chi^2] = 2k,\ a_s = \sqrt{\frac{8}{k}},\ \varepsilon_k = \frac{12}{k}.$$

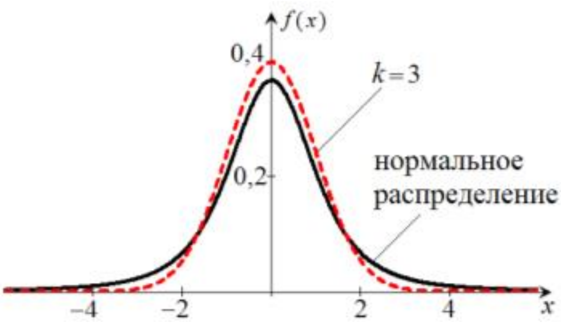
2. Розподіл Стьюдента. Нехай X, X1, X2, ..., Xk – незалежні випадкові величини, що мають стандартний нормальний розподіл з параметрами a = 0, σ = 1.

Розподілом Стьюдента (або t-розподілом) з k ступенями вободи називається розподіл відношення

$$T = \frac{X}{\sqrt{\frac{1}{k}(X_1^2 + X_2^2 + \ldots + X_k^2)}} = \frac{X}{\sqrt{\frac{\chi^2}{k}}}.$$

Щільність цього розподілу визначається за формулою

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi} \cdot \Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}, \quad -\infty < x < +\infty$$



Зі збільшенням значень k розподіл Стьюдента досить швидко наближається до нормального розподілу

Числові характеристики розподілу Стьюдента:

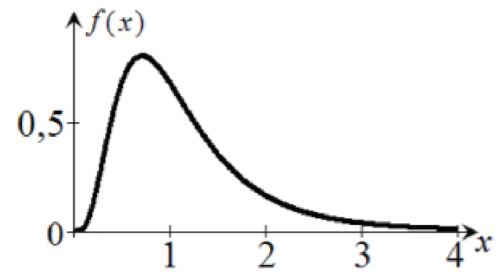
$$M[T]=0,\ D[T]=\frac{k}{k-2}\ (k>2),\ a_s=0,\ \varepsilon_k=\frac{6}{k-4}.$$

3. Розподіл Фішера-Снедекора. F-розподілом ***Фішера-Снедекора*** зі ступенями свободи k_1 і k_2 називається розподіл відношення

$$F=\frac{\chi^2(k_1)/k_1}{\chi^2(k_2)/k_2}$$

Щільність цього розподілу визначається за формулою

$$f(x)=\frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\cdot\Gamma\left(\frac{k_2}{2}\right)}\cdot\frac{k_1^{\frac{k_1}{2}}\cdot k_2^{\frac{k_2}{2}}\cdot x^{\frac{k_1}{2}-1}}{(k_1x+k_2)^{\frac{k_1+k_2}{2}}},\quad x\ge 0,$$



Графік щільності F-розподілу при k_1 = 10 і k_2 = 15 представлений на рис.

Числові характеристики розподілу Фішера-Снедекора:

$$M[F]=\frac{k_2}{k_2-2}\ (k_2>2),\ \ D[F]=\frac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}\ (k_2>4).$$

Статистична гіпотеза. Нульова і конкуруюча, проста і складна гіпотези

Статистичної називають гіпотезу про вид невідомого розподілу, або про параметри відомих розподілів.

Наприклад, статистичними є гіпотези:

1) генеральна сукупність розподілена за законом Пуассона;

2) дисперсії двох нормальних сукупностей рівні між собою.

Гіпотеза «на Марсі є життя» не є статистичною, оскільки в ній не йдеться ні про вигляді, ні про параметри розподілу.

Поряд з висунутої гіпотезою розглядають і суперечливу їй гіпотезу. Якщо висунута гіпотеза буде відкинута, то має місце суперечлива гіпотеза. З цієї причини ці гіпотези доцільно розрізняти.

Нульовою (основною) називають висунуту гіпотезу H0.

Конкуруючою (альтернативною) називають гіпотезу H1, яка суперечить нульовий.

Наприклад, якщо нульова гіпотеза полягає в припущенні, що математичне очікування a нормального розподілу дорівнює 10, то конкуруюча гіпотеза, зокрема, може полягати в припущенні, що $a \ne 10$. Коротко це записують так:

H0: $a = 10$; H1: $a \ne 10$.

Розрізняють гіпотези, які містять тільки одне і більш одного припущень.

Простою називають гіпотезу, яка містить тільки одне припущення. Наприклад, якщо λ-параметр показникового розподілу, то гіпотеза H0: $\lambda = 5$ – проста.

Складною називають гіпотезу, яка складається з кінцевого або нескінченного числа простих гіпотез. Наприклад, складна гіпотеза H: $\lambda > 5$ складається з незліченної множини простих виду Hi: $\lambda = b_i$, де b_i-будь-яке число, більше 5.

Помилки першого і другого роду

Висунута гіпотеза може бути правильною або неправильною, тому виникає необхідність її перевірки. Оскільки перевірку проводять статистичними методами, її називають статистичної. В результаті статистичної перевірки гіпотези у двох випадках може бути прийнято неправильне рішення, т. Е. Можуть бути допущені помилки двох родів.

Помилка першого роду полягає в тому, що буде відкинута правильна гіпотеза.

Помилка другого роду полягає в тому, що буде прийнята неправильна гіпотеза.

Підкреслимо, що наслідки цих помилок можуть виявитися вельми різними. Наприклад, якщо відкинуто правильне рішення, «продовжувати будівництво житлового будинку», то ця помилка першого роду спричинить матеріальну шкоду; якщо ж прийнято неправильне рішення «продовжувати будівництво», незважаючи на небезпеку обвалу будівництва, то ця помилка другого роду може спричинити загибель людей. Можна навести приклади, коли помилка першого роду тягне більш важкі наслідки, ніж помилка другого роду.

Зауваження. Імовірність припуститися помилки першого роду прийнято позначати через α; її називають рівнем значущості. Найбільш часто рівень значущості приймають рівним 0,05 або 0,01. Якщо, наприклад, прийнятий рівень значимості, рівний 0,05, то це означає, що в п'яти випадках зі ста є ризик припуститися помилки першого роду (відкинути правильну гіпотезу).

Загальна схема перевірки статистичних гіпотез

Маючи дві гіпотези H0 і H1, необхідно на основі вибіркових даних або прийняти основну гіпотезу H0, або конкуруючу H1.

Правило, за яким приймається рішення прийняти або відхилити гіпотезу H0 (або H1), називається ***статистичним критерієм*** (або просто критерієм) перевірки гіпотези H0.

Статистикою (або тестом) критерію називають випадкову величину τ, яка служить для перевірки статистичних гіпотез.

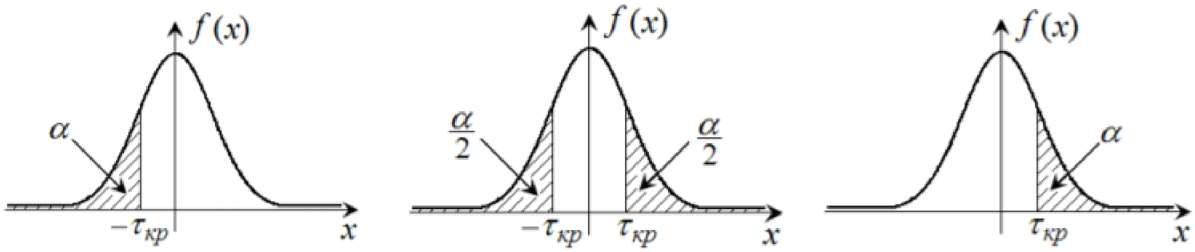
Наведемо схему перевірки статистичних гіпотез:

1. Для основної гіпотези H0 формулюється альтернативна гіпотеза H1.
2. Вибирається рівень значущості перевірки – мале число $\alpha > 0$.
3. Розглядаються теоретичні вибірки значень випадкових величин, про яких сформульована гіпотеза H0, І вибирається (формується) випадкова величина τ. Значення і розподіл τ (зазвичай з перерахованих нижче: U – нормальний розподіл, χ^2 – розподіл Пірсона, T - Стьюдента, F - Фішера-Снедекора) повністю визначаються за вибірками при припущенні про вірність гіпотези H0

4. На числової осі задають інтервал D, такий, що ймовірність попадання випадкової величини τ в цей інтервал: P D () 1 τ $\in$ = -α. Інтервал D називається областю прийняття гіпотези H0, А решта область числової осі - критичною областю (величина $\tau = \tau_{кр}$ - критичне значення тесту перевірки).

Розрізняють три типи критичних областей. Критична область визначається з урахуванням гіпотез:

H_0	H_1	*Критическая область D*	
$\theta = \tilde{\theta}$	$\theta < \tilde{\theta}$	$(-\infty; -\tau_{кр}]$	– *левосторонняя* (рис. 20)
	$\theta \neq \tilde{\theta}$	$(-\infty; -\tau_{кр}] \cup [\tau_{кр}; +\infty)$	– *двусторонняя* (рис. 21)
	$\theta > \tilde{\theta}$	$[\tau_{кр}; +\infty)$	– *правосторонняя* (рис. 22)



Відповідно інтервалах критерій перевірки називається правостороннім, двостороннім або лівостороннім.

5. За реалізацій аналізованих вибірок обчислюється конкретне (що спостерігається) значення тесту τ (позначимо його $\tau = \tau_{спост}$) і перевіряється виконання умови P D () 1 τ $\in$ = -α:

а) якщо воно виконується (наприклад, $\tau_{спост} < \tau_{кр}$ для правобічної області), то гіпотеза H0 приймається в тому сенсі, що вона не суперечить дослідними даними і немає підстав її відкинути;

б) якщо умова не виконується ($\tau_{спост} > \tau_{кр}$ для правобічної області), то гіпотеза H0 невірна і її відкидають.

Для кожного критерію є відповідні таблиці, за якими і знаходять критичне значення, яке задовольняє наведеним вище співвідношенням.

Принцип прийняття статистичної гіпотези не дає логічного докази її вірності або невірності. Ухвалення гіпотези H0 в порівнянні з альтернативної H1 не означає, що ми впевнені в абсолютній правильності H0, або, що висловлене в гіпотезі H0 твердження є найкращим, єдино придатним. Просто гіпотеза H0 який суперечить наявними у нас вибірковими даними. Таким же властивістю поряд з H0 можуть володіти і інші гіпотези. Більш того, можливо, що при збільшенні обсягу вибірки n або при випробуванні H0 проти іншої альтернативної гіпотези H2 гіпотеза H0 буде відкинута.

Таким чином, прийняття гіпотези H0 слід розцінювати не як раз і назавжди встановлений, абсолютно вірний міститься в ній факт, а лише як досить правдоподібне, що не суперечить досвіду твердження.

З представленої схеми слід, що при перевірці гіпотези H0 може бути прийнято неправильне рішення, т. е. можуть бути допущені помилки двох видів:

ошибка I *рода*	*ошибка* II *рода*
Отвергается основная (нулевая) гипотеза, хотя она верна.	Отвергается конкурирующая гипотеза, хотя она верна.
Вероятность ошибки $P(H_1 \mid H_0) = \alpha$, α – ***уровень значимости критерия*** (обычно $\alpha = 0{,}05$; 0,01; 0,005; 0,001).	Вероятность ошибки $P(H_0 \mid H_1) = \beta$ (величина β, как правило, заранее неизвестна)
Вероятность принять верную (нулевую) гипотезу $P(H_0 \mid H_0) = 1 - \alpha$.	Вероятность принять верную (конкурирующую) гипотезу $P(H_1 \mid H_1) = 1 - \beta$, $(1 - \beta)$ – ***мощность критерия***.

Наслідки помилок 1-го і 2-го роду можуть бути абсолютно різними: в одних випадках треба мінімізувати α, а в інших - β. Так, стосовно до радіолокації кажуть, що α - ймовірність пропустити сигнал, β - ймовірність помилкової тривоги. Стосовно до виробництва, до торгівлі можна сказати, що α - ризик постачальника (т. е. забракування по всій партії виробів, які відповідають стандарту), β - ризик споживача (прийом по вибірці всієї партії виробів, які задовольняють стандарту). Стосовно до судової системи, помилка 1-го роду призводить до виправдання винного, помилка 2-го - засудження невинного.

Слід зазначити, що одночасне зменшення помилок 1-го і 2-го роду можливо лише при збільшенні обсягу вибірок. Тому зазвичай при заданому рівні значущості α відшукується критерій з найбільшою потужністю.

Статистики порівняння точкових оцінок невідомих генеральних

1) Перевірка гіпотез для однієї вибірки

Нехай генеральна сукупність X розподілена за нормальним законом. Генеральна середня *a* хоча і невідома, але є підстави припускати, що вона дорівнює передбачуваному значенням a_0. Наприклад, якщо X – сукупність розмірів партії деталей, що виробляються верстатом автоматичної лінії, то можна припустити, що генеральна середня а цих розмірів дорівнює проектному розміром а0. Для того, щоб перевірити, чи правильно встановлено налаштування цього верстата, очевидно треба переконатися в тому, що середнє значення параметра у вироблених на ньому виробів буде відповідати номіналу. Таким чином, необхідно перевірити гіпотезу H0: $a = a_0$ проти альтернативної:

$H_1 : a \neq a_0$, или $H_2 : a < a_0$, или $H_3 : a > a_0$.

Якщо відмінність виявиться незначною, то верстат забезпечує в середньому проектний розмір; якщо відмінність значуща, то верстат вимагає налагодження.

При довільному налаштуванні верстата може виникнути необхідність перевірки гіпотези про те, що точність виготовлення виробів по даному параметру, що задається дисперсією σ^2, дорівнює заданій величині σ_0^2

(H0: σ2= σ_0^2), або наприклад, того, що частка бракованих виробів виготовлених верстатом, дорівнює заданій величині p_0 (H0: $p = p_0$) і т. д.

Висунуті гіпотези і відповідні критерії перевірки гіпотез про числових значеннях параметрів нормального закону наведені в табл.

H_0	Статистика критерия	H_1	Область принятия H_0
$a = a_0$ $\sigma^2 = \sigma_г^2$ известно	$U = \frac{\bar{x} - a_0}{\sigma}\sqrt{n}$	$a \neq a_0$	$\lvert U \rvert < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \frac{\alpha}{2}$
		$a < a_0$	$U > -u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$
		$a > a_0$	$U < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$
$a = a_0$ $\sigma^2 = \sigma_г^2$ неизвестно	$T = \frac{\bar{x} - a_0}{s}\sqrt{n}$	$a \neq a_0$	$\lvert T \rvert < t_{кр}$, $t_{кр} = t_{\alpha, n-1}$ для двусторонней области
		$a < a_0$	$T > -t_{кр}$, $t_{кр} = t_{\alpha, n-1}$ для односторонней области
		$a > a_0$	$T < t_{кр}$, $t_{кр} = t_{\alpha, n-1}$ для односторонней области
$\sigma^2 = \sigma_0^2$ a неизвестно	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\sigma^2 \neq \sigma_0^2$	$\chi^2_{1-\frac{\alpha}{2};n-1} < \chi^2 < \chi^2_{\frac{\alpha}{2};n-1}$
		$\sigma^2 < \sigma_0^2$	$\chi^2 > \chi^2_{1-\alpha;n-1}$
		$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{\alpha;n-1}$
$p = p_0$ достаточно большие n, $np_0 > 5$, $nq_0 > 5$, $q_0 = 1 - p_0$	$U = \frac{p^* - p_0}{\sqrt{p_0 q_0}}\sqrt{n}$, $p^* = \frac{m}{n}$	$p \neq p_0$	$\lvert U \rvert < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \frac{\alpha}{2}$
		$p < p_0$	$U > -u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$
		$p > p_0$	$U < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$

Примітка. Критичні значення статистик на рівні значущості α визначають за відповідними таблицями додатка.

2) Перевірка гіпотез для двох незалежних вибірок.

Нехай є дві незалежні нормально розподілені вибірки x_1, x_2, ..., x_{n1} і y_1, y_2, ..., y_{n2} з параметрами (a_x, σ_x^2) (a_y, σ_y^2) відповідно. Зазвичай ставиться завдання перевірки їх однорідності, або іншими словами. рівності обох параметрів, або слід перевірити рівність параметрів окремо.

Порівняння середніх двох сукупностей має важливе практичне значення. На практиці часто зустрічається випадок, коли середній результат однієї серії експериментів відрізняється від середнього результату іншої серії. При цьому виникає питання, чи можна пояснювати виявлену розбіжність середніх неминучими випадковими помилками експерименту або вона викликана деякими закономірностями. У промисловості задача порівняння середніх часто виникає при вибірковому контролі якості виробів, виготовлених на різних установках або при різних технологічних режимах, в фінансовому аналізі - при зіставленні рівня прибутковості різних активів і т. д.

Гіпотеза про рівність середніх при відомих дисперсіях перевіряється звичайно у випадку великих вибірок (об'ємом близько сотень), коли оцінки дисперсій можна прийняти за їх точні значення.

Гіпотеза про рівність середніх при невідомих дисперсіях вимагає спочатку перевірки гіпотези про рівність дисперсій двох вибірок.

Гіпотези про дисперсії виникають досить часто, так як дисперсія характеризує такі виключно важливі показники, як точність машин, приладів, технологічних процесів, ступінь однорідності сукупностей, ризик, пов'язаний з відхиленням прибутковості активів від очікуваного рівня, і т. д.

Порівняння часткою ознаки в двох сукупностях – досить часто зустрічається на практиці завдання. Наприклад, якщо вибіркова частка ознаки в одній сукупності відрізняється від такої ж частки в інший сукупності, то вказує це на те, що наявність ознаки в одній сукупності дійсно найімовірніше, або отримане розбіжність часткою є випадковим?

Сформулюємо задачу. Є дві сукупності X і Y, генеральні сукпності ознаки в яких рівні відповідно x_p і y_p. Необхідно перевірити нульову гіпотезу про рівність генеральних часткою. Для перевірки гіпотези H0 з цих сукупностей взяті дві незалежні вибірки досить великого обсягу n_1 і n_2. Вибіркові частки ознаки рівні відповідно $p_1^* = \dfrac{m_1}{n_1}$ і $p_2^* = \dfrac{m_2}{n_2}$, де m_1 і m_2 – відповідно число елементів перщj] і другj] вибірок, що володіють даними ознакою. Висунуті гіпотези і відповідні критерії перевірки гіпотез представлені в табл

H_0	Статистика критерия	H_1	Область принятия H_0
$a_x = a_y$ σ_x^2 и σ_y^2 известны	$U = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_x^2}{n_1} + \dfrac{\sigma_y^2}{n_2}}}$	$a_x \neq a_y$	$\lvert U \rvert < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \dfrac{\alpha}{2}$
		$a_x < a_y$	$U > -u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$
		$a_x > a_y$	$U < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$
$a_x = a_y$ σ_x^2 и σ_y^2 неизвестны, но равны	$T = \dfrac{\bar{x} - \bar{y}}{s \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$, $s = \sqrt{\dfrac{s_x^2 \cdot (n_1 - 1) + s_y^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}}$	$a \neq a_0$	$\lvert T \rvert < t_{кр}$, $t_{кр} = t_{\alpha, n_1 + n_2 - 1}$ для двусторонней области
		$a < a_0$	$T > -t_{кр}$, $t_{кр} = t_{\alpha, n_1 + n_2 - 1}$ для односторонней области
		$a > a_0$	$T < t_{кр}$, $t_{кр} = t_{\alpha, n_1 + n_2 - 1}$ для односторонней области
$\sigma_x^2 = \sigma_y^2$ a_x и a_y неизвестны	$F = \dfrac{s_{\max}^2}{s_{\min}^2}$	$\sigma_x^2 \neq \sigma_y^2$	$F < F_{кр}$, $F_{кр} = F_{\alpha/2,\, n_1 - 1,\, n_2 - 1}$
		$\sigma_x^2 > \sigma_y^2$	$F < F_{кр}$, $F_{кр} = F_{\alpha,\, n_1 - 1,\, n_2 - 1}$
$p_x = p_y$ n_1 и n_2 достаточно большие	$U = \dfrac{p_1^* - p_2^*}{\sqrt{p^*(1 - p^*)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$, $p_1^* = \dfrac{m_1}{n_1}$, $p_2^* = \dfrac{m_2}{n_2}$, $p^* = \dfrac{m_1 + m_2}{n_1 + n_2}$	$p_x \neq p_y$	$\lvert U \rvert < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \dfrac{\alpha}{2}$
		$p_x < p_y$	$U > -u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$
		$p_x > p_y$	$U < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$

Побудова теоретичного закону розподілу випакової величини по експериментальним даним. Перевірка гіпотез про закон розподілу

Однією з найважливіших завдань математичної статистики є встановлення теоретичного закону розподілу випадкової величини, характеризує досліджувану ознаку по експкриментальному (емпіричному) розподілу, який представляє варіаційний ряд. Для вирішення цього завдання необхідно визначити вид і параметри закону розподілу.

Гіпотеза про вид закону розподілу може бути висунута виходячи з теоретичних передумов, досвіду аналогічних попередніх досліджень і, нарешті, на підставі графічного зображення емпіричного розподілу.

Параметри розподілу, як правило, невідомі, тому їх замінюють найкращими оцінками за вибіркою ($\bar{x}$, $D_в$, $\sigma_в$ і т.д.).

Розподілом, що грає основну роль в теорії статистичного оцінювання, є нормальний розподіл. Наведемо його основні характеристики (табл.), розглянуті в курсі теорії ймовірностей.

Нормальное распределение ($a=\bar{x}_в$, $\sigma=\sigma_в$ или $\sigma=s$)		
Функция плотности	Функция распределения	Вероятность попадания в интервал
$f(x)=\frac{1}{\sigma}\cdot\varphi(t)$, где $t=\frac{x-a}{\sigma}$, $\varphi(t)$ – функция Гаусса	$F(x)=0{,}5+\Phi(t)$, где $t=\frac{x-a}{\sigma}$, $\Phi(t)$ – функция Лапласа	$p_i=P\{x_{i-1}<x<x_i\}=$ $=\frac{x_i-x_{i-1}}{\sigma}\cdot\varphi\left(\frac{\tilde{x}_i-a}{\sigma}\right)=$ $=\Phi\left(\frac{x_i-a}{\sigma}\right)-\Phi\left(\frac{x_{i-1}-a}{\sigma}\right)$, где $i=1,2,\ldots,k$.

Графики функций $\varphi(t)$ (рис. 23) и $\Phi(t)$ (рис. 24):

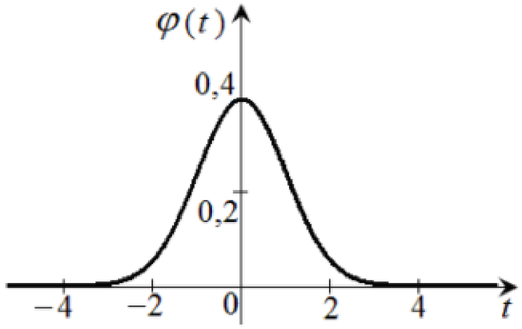


Рис. 23. Функция Гаусса

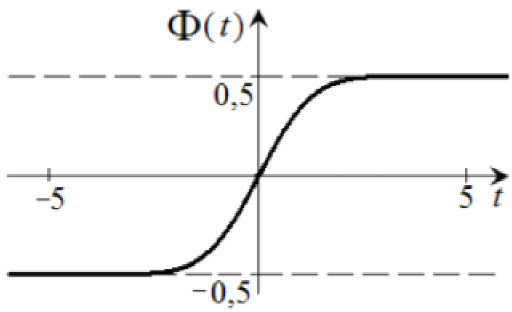


Рис. 24. Функция Лапласа

$$\varphi(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}} \text{ (приложение 1)} \qquad \Phi(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt \text{ (приложение 2)}$$

Широке використання в статистичних висновках нормального розподілу має як емпіричне, так і теоретичне обґрунтування. Численні приклади побудови гістограм і згладжування їх неперервними кривими для експериментальних даних самої різної природи показують, що в багатьох випадках нормальний розподіл є досить точним поданням таких даних. Застосовність нормального розподілу обґрунтовується і центральною граничною теоремою.

Як би добре не був обраний теоретичний закон розподілу, неминучі розбіжності між емпіричним і теоретичним розподілами. Природно, виникає питання: ці розбіжності пояснюються тільки випадковими обставинами, пов'язаними з невеликим числом спостережень, або вони є суттєвими і пов'язані з тим, що теоретичний закон розподілу підібраний невдало. Для відповіді на це питання використовуємо спеціально підібрану величину – ***критерій згоди***.

Критерієм згоди називають статистичний критерій перевірки гіпотези про передбачуваний законі невідомого розподілу. Він використовується для перевірки згоди передбачуваного виду розподілу з експериментальними даними на підставі вибірки.

Нехай необхідно перевірити гіпотезу H0 про те, що розглянута випадкова величина X підпорядковується певному законоу розподілу.

Для перевірки гіпотези H0 вибирають деяку випадкову величину τ, що характеризує ступінь розбіжності теоретичного та емпіричного розподілів, закон розподілу якої при досить великих *n* відомий і практично не залежить від закону розподілу випадкової величини X.

Знаючи закон розподілу τ, можна знайти таке критичне значення $\tau = \tau_{кр}$, що якщо гіпотеза H0 вірна, то ймовірність того, що τ прийняла значення більше ніж τ_α : $P(\tau > \tau_\alpha) = \alpha$ – мале, де α - рівень значущості критерію.

Якщо $\tau = \tau_{спост} > \tau_{кр}$,то гіпотеза H0 відхиляється.

Якщо $\tau = \tau_{спост} < \tau_{кр}$,то гіпотеза H0 приймається.

Критерій χ^2 Пірсона

Існують різні критерії згоди: Пірсона, Колмогорова, Фішера, Смирнова та ін. Критерій згоди Пірсона – найбільш часто застосовується для перевірки простої гіпотези про закон розподілу.

Алгоритм дій:

1. Вибрати закон розподілу випадкової величини.
2. За відповідною формулою обчислити точкові (або інтервальні) ймовірності p_i.
3. Обчислити вирівнюючі частоти $m'_i = n * p_i$, де *n* - обсяг вибірки.
4. Знайти статистику $$\chi^2_{набл} = \sum_{i=1}^{k} \frac{(m_i - m'_i)^2}{m'_i}.$$
5. Визначити число ступенів свободи *l = k - r -1,* де *k* – число часткових інтервалів вибірки; *r* - число параметрів диференціальної функції розподілу. Вирази для знаходження числа ступенів свободи відомих законів розподілу представлені в табл.

Закон распределения	Число степеней свободы
Биномиальный закон	$l = k - 1$, если p_A известно
	$l = k - 2$, если p_A неизвестно
Закон распределения Пуассона	$l = k - 2$
Равномерный закон	$l = k - 3$
Показательный закон	$l = k - 2$
Нормальный закон	$l = k - 3$

6. По таблиці додатке необхідно знайти критичну величину $\chi^2_{\alpha,l} = \chi^2_{кр}$, де α – заданий рівень значущоссті.

7. Якщо $\tau = \tau_{спост} > \tau_{кр}$,то гіпотеза H0 відхиляється. Якщо $\tau = \tau_{спост} < \tau_{кр}$,то гіпотеза H0 приймається.