

Софийски Университет "Св.Климент Охридски"

Физически факултет



Людмил Цанков

Вероятности и физическа статистика

(Записки на лекции)

София, 2011г.

СЪДЪРЖАНИЕ:

I. УВОД	4
II. ВЕРОЯТНОСТ	6
1. КЛАСИЧЕСКИ ВЕРОЯТНОСТЕН МОДЕЛ	6
1.1. ИНТУИТИВНО ПОНЯТИЕ ЗА ВЕРОЯТНОСТ. ВЕРОЯТНОСТТА КАТО ЧЕСТОТА	6
1.2. ГЕОМЕТРИЧНА ВЕРОЯТНОСТ	7
1.3. КОМБИНАТОРНА ВЕРОЯТНОСТ. ПРАВИЛА ЗА БРОЕНЕ	8
1.3.1. <i>Наредена извадка без повторение</i>	9
1.3.2. <i>Наредена извадка с повторение</i>	10
1.3.3. <i>Ненаредена извадка без повторение</i>	11
1.3.4. <i>Ненаредена извадка с повторение</i>	11
1.4. ИСТОРИЧЕСКИ БЕЛЕЖКИ	12
2. АКСИОМАТИЧНА ТЕОРИЯ НА ВЕРОЯТНОСТИТЕ	13
2.1. ПРОСТРАНСТВО НА ЕЛЕМЕНТАРНИТЕ СЪБИТИЯ. АЛГЕБРА НА СЪБИТИЯТА	13
2.2. АКСИОМИ НА ВЕРОЯТНОСТТА	16
2.3. УСЛОВНА ВЕРОЯТНОСТ	16
2.4. СТОХАСТИЧНА НЕЗАВИСИМОСТ.....	17
III. СЛУЧАЙНИ ВЕЛИЧИНИ	18
3. ВЕРОЯТНОСТНИ РАЗПРЕДЕЛЕНИЯ	18
3.1. ФУНКЦИИ НА СЛУЧАЙНИ ВЕЛИЧИНИ. МАТЕМАТИЧЕСКО ОЧАКВАНЕ. ДИСПЕРСИЯ. МОМЕНТИ	19
3.2. СВОЙСТВА НА МАТЕМАТИЧЕСКОТО ОЧАКВАНЕ И ДИСПЕРСИЯТА	23
4. РАЗПРЕДЕЛЕНИЯ НА НЯКОЛКО СЛУЧАЙНИ ВЕЛИЧИНИ	24
4.1. РАЗПРЕДЕЛЕНИЯ НА ДВЕ СЛУЧАЙНИ ВЕЛИЧИНИ	24
4.2. МОМЕНТИ НА СЪВМЕСТНИТЕ РАЗПРЕДЕЛЕНИЯ.....	26
4.2.1. <i>Свойства на моментите на съвместните разпределения</i>	27
4.3. СЪВМЕСТНИ РАЗПРЕДЕЛЕНИЯ НА ПОВЕЧЕ ОТ ДВЕ СЛУЧАЙНИ ВЕЛИЧИНИ.....	29
5. ТРАНСФОРМАЦИИ НА СЛУЧАЙНИ ВЕЛИЧИНИ. РАЗПРОСТРАНЕНИЕ НА ГРЕШКИТЕ	30
IV. НЯКОИ ВАЖНИ СТАТИСТИЧЕСКИ РАЗПРЕДЕЛЕНИЯ	34
6. ДИСКРЕТНИ РАЗПРЕДЕЛЕНИЯ	34
6.1. ЕКСПЕРИМЕНТ С ДВА ИЗХОДА (РАЗПРЕДЕЛЕНИЕ НА БЕРНУЛИ)	34
6.2. БИНОМНО РАЗПРЕДЕЛЕНИЕ	34
6.3. СХОДИМОСТ ПО ВЕРОЯТНОСТ. ЗАКОН ЗА ГОЛЕМИТЕ ЧИСЛА	36
6.4. ОБОБЩЕНИЯ И ГРАНИЧНИ СЛУЧАИ НА БИНОМНОТО РАЗПРЕДЕЛЕНИЕ.....	37
6.5. РАЗПРЕДЕЛЕНИЕ НА ПОАСОН.....	38
6.5.1. <i>Разпределението на Поасон като граничен случай на биномното разпределение</i>	38
6.5.2. <i>Физичен модел на разпределението на Поасон</i>	39
6.5.3. <i>Свойства на Поасоновото разпределение</i>	41
7. НЕПРЕКЪСНАТИ РАЗПРЕДЕЛЕНИЯ	42
7.1. РАВНОМЕРНО РАЗПРЕДЕЛЕНИЕ	42
7.2. НОРМАЛНО РАЗПРЕДЕЛЕНИЕ	43
7.2.1. <i>Лапласов модел на грешките</i>	43
7.2.2. <i>Свойства на нормалното разпределение</i>	45
7.2.3. <i>Нормалното разпределение и грешките при експеримента: модел на Хершел</i>	47
7.3. РАЗПРЕДЕЛЕНИЕ χ^2	49
V. ОЦЕНЯВАНЕ НА ПАРАМЕТРИ ЧРЕЗ ИЗВАДКИ	51
8. ИЗВАДКИ	51

8.1.	СЛУЧАЕН ИЗБОР. РАЗПРЕДЕЛЕНИЕ НА ИЗВАДКИТЕ	51
8.2.	ОЦЕНКИ НА ПАРАМЕТРИ. СВОЙСТВА НА ОЦЕНКИТЕ.....	53
8.3.	ОЦЕНКИ ЗА СРЕДНАТА СТОЙНОСТ И ЗА ДИСПЕРСИЯТА НА ВЕРОЯТНОСТНОТО РАЗПРЕДЕЛЕНИЕ	54
VI. МЕТОД НА МАКСИМАЛНОТО ПРАВДОПОДОБИЕ.....		57
9.	ФУНКЦИЯ НА ПРАВДОПОДОБИЕ	57
10.	МЕТОД НА МАКСИМАЛНОТО ПРАВДОПОДОБИЕ.....	59
VII. ПРОВЕРКА НА СТАТИСТИЧЕСКИ ХИПОТЕЗИ		62
11.	ОБЩА ТЕОРИЯ.....	62
12.	КРИТЕРИЙ ЗА ПРИНАДЛЕЖНОСТ КЪМ СТАНДАРТНО НОРМАЛНО РАЗПРЕДЕЛЕНИЕ	64
13.	F-КРИТЕРИЙ ЗА РАВЕНСТВО НА ДИСПЕРСИИТЕ.....	66
14.	T-КРИТЕРИЙ ЗА СРАВНЯВАНЕ НА СРЕДНИ.....	69
14.1.	T-КРИТЕРИЙ ЗА СРАВНЯВАНЕ НА СРЕДНА СТОЙНОСТ С КОНСТАНТА.....	69
14.2.	T-КРИТЕРИЙ ЗА СРАВНЯВАНЕ НА СРЕДНИТЕ НА ДВЕ ИЗВАДКИ	71
15.	КРИТЕРИЙ ЗА СЪГЛАСИЕ χ^2	72
15.1.	СРАВНЯВАНЕ НА ЕМПИРИЧНАТА ХИСТОГРАМА С ДАДЕНА ВЕРОЯТНОСТНА ПЛЪТНОСТ	72
15.2.	СРАВНЯВАНЕ НА ВЕРОЯТНОСТНИТЕ ПЛЪТНОСТИ НА ДВЕ ИЗВАДКИ	74
15.3.	ПРОВЕРКА НА СЛОЖНИ ХИПОТЕЗИ.....	74
VIII. МЕТОД НА НАЙ-МАЛКИТЕ КВАДРАТИ		76
16.	ОБЩИ БЕЛЕЖКИ.....	76
17.	МЕТОД НА НАЙ-МАЛКИТЕ КВАДРАТИ ПРИ ПРЕКИ НАБЛЮДЕНИЯ.....	77
18.	МЕТОД НА НАЙ-МАЛКИТЕ КВАДРАТИ ПРИ КОСВЕНИ НАБЛЮДЕНИЯ	80
18.1.	ЛИНЕЕН СЛУЧАЙ.....	80
18.2.	НЕЛИНЕЕН СЛУЧАЙ	83
18.3.	АЛГОРИТЪМ НА МНК ЗА ИЗМЕРВАНИЯ БЕЗ ОГРАНИЧЕНИЯ	85
19.	СВОЙСТВА НА РЕШЕНИЯТА, ПОЛУЧЕНИ ЧРЕЗ МЕТОДА НА НАЙ-МАЛКИТЕ КВАДРАТИ...86	86

I. Увод

Физиката се занимава с изучаването на процесите и явленията в природата. Един от главните методи на научно изследване (и то не само във физиката) е извършването на експерименти върху изучаваните обекти.

Експериментът се дефинира като съвкупност от строго организирани последователни действия, насочени към получаването на определени резултати, които характеризират изследваните обекти. Тези резултати като правило са свързани с получаването на числени стойности на измерваните величини. Основно качество на научните експерименти е тяхната повторемост. Това означава осигуряване постоянството и възможността за възпроизводимост на всички условия, при които се провежда експериментът. Може да се каже, че науката започва тогава, когато е налице повторемостта на експериментите.

Въпреки полагането на усилия за осигуряване на повторемост обаче, често при различните реализации на един експеримент се получават различни стойности на изследваните величини. Нещо повече - оказва се, че не е възможно да се предвиди отнапред каква точно стойност на дадената величина ще се получи при следващото измерване.

Има два главни възгледа за обяснение на този факт. Според единия от тях, всички физични величини имат някакви точни, напълно определени "истински" стойности, но в един реален експеримент не е възможно пълното познаване и възпроизвеждане на условията, при които той протича. Различията между отделните реализации на експеримента се проявяват като грешки, които влияят върху стойностите на търсените величини, поради което всеки път се получават, изобщо казано, различни резултати. Според другото основно схващане, физичните величини имат вътрешно присъща неопределеност, т.е. те съдържат потенциално в себе си набор от различни възможни стойности, и при всеки опит да бъдат измерени се получава само една от тях. Тези схващания имат и философски аспекти, които се свързват с понятията детерминизъм и индетерминизъм, дискутирани особено много във връзка с квантовата теория.

Независимо от същностните причини за получаването на различни резултати при видимо еднакви условия на експеримента, за външния наблюдател нещата изглеждат по еднакъв начин – така или иначе съществуват някои величини, които при различни реализации на експеримента заемат различни стойности, които при това не могат отнапред да бъдат точно предсказани. Такива величини се наричат случайни.

Науката, която се занимава с изучаването на свойствата на случайните величини, се нарича стохастика. Тя обединява теорията на вероятностите, математическата статистика и техните приложения (приложната статистика, обработката на данни, теорията на вземане на решения и

др.). Стохастиката е един от математичните фундаменти на съвременното естествознание и преди всичко на физиката. Стохастичните методи обаче се използват изключително широко и в биологията, медицината, икономиката, социалните науки.

Цел на този курс е изучаването на основните понятия и методи на стохастиката, връзката им с някои основни физични схващания и прилагането им за решаването на проблеми, свързани с правилната интерпретация на резултатите от измерванията на физични величини.

II. Вероятност

1. Класически вероятностен модел

1.1. Интуитивно понятие за вероятност. Вероятността като честота

Както казахме, експериментът представлява съвкупност от условия (т.е. предписания, свързани с условията на наблюдение), чието изпълнение гарантира получаването на някакъв резултат/резултати. Те се наричат още изходи от експеримента или събития. Ако изходите от даден експеримент при различни негови реализации са различни, ние говорим за случайни събития.

Оказва се, че не е възможно да се предскаже дали дадено събитие ще се сбъдне или не при следващия опит. При многократно провеждане на един и същ експеримент обаче се наблюдават определени закономерности, които като правило стават все по-отчетливи с увеличаването на броя на опитите. Поради това в стохастиката се разглеждат само случайни събития, които могат да се възпроизвеждат многократно, т.е. имат масов характер.

Класическите примери за случайни събития са добре известни. Ето кратка таблица:

експеримент	случайно събитие
хвърляне на монета	попадане на лице/герб
хвърляне на зарове	получаване на определено число
изтегляне на числа от Тото2	получаване на печеливша комбинация

Ако при n експеримента събитието A се е сбъднало m пъти, числото $\nu = m/n$ се нарича относителна честота (понякога само честота) на сбъдане на събитието A .

Ясно е, че винаги е в сила:

$$\nu = \frac{m}{n} \in [0, 1] \quad (1)$$

За да се определи вероятността за настъпване на дадено събитие A могат да се извършат голям брой опити E_1, E_2, \dots при еднакви условия. За всяка редица от опити $\{E_1, \dots, E_k\}$ може да се определи относителната честота $\nu_k = \frac{m_k}{n}$ за сбъдане на A . При голям брой опити стойностите на тази честота се стабилизират около някакво число, т.е. редицата от числа $\{\nu_k\}$ изглежда да е сходяща. Ние не можем да установим сходимостта на редицата от относителни честоти опитно, тъй като не сме в състояние да извършим безкраен брой експерименти.

Естествено е обаче да направим такова предположение за сходимост като обобщение на натрупания опит и да дефинираме вероятността като граница:

$$p(A) = \lim_{n \rightarrow \infty} \frac{m}{n} \quad (2)$$

Така се достига до статистическото (наричано още честотно или класическо) определение за вероятността. Най-общо казано, вероятността за дадено събитие е количествена мярка за възможността събитието да настъпи.

В някои случаи е възможно вероятността за дадено събитие да бъде определена чрез пресмятане, без да се извършват голям брой експерименти. Необходимо условие за това е наличието на (или предположението за) равна възможност за събъждане на различните изходи от експеримента. Такива експерименти се наричат още експерименти с равновероятни изходи.

Примери за експерименти с краен брой равновероятни изходи бяха вече посочени (хвърляне на монета, зарове, изтегляне на топки от урна или на игрални карти от колода). В такива случаи общият брой изходи и благоприятните изходи просто могат да се изброят. Има и случайни величини, при които изходите са също равновероятни, но не са краен брой. Така е например при така наречените геометрични вероятности.

1.2. Геометрична вероятност

Най-простият пример за геометрична вероятност е следният мислен експеримент: Нека върху отсечката $[a, b]$ напосоки се избира една точка. Интересуваме се каква е вероятността за това, тя да попадне в интервала $[\alpha, \beta] \in [a, b]$?

В този прост случай отговорът изглежда очевиден. Доколкото вероятността да попаднем в интервала $[\alpha, \beta]$ не зависи от това, къде се намира той спрямо основния интервал $[a, b]$, то търсената вероятност ще бъде: $p(A) = \frac{\beta - \alpha}{b - a}$, т.е. ще зависи само от отношението на дължините на интервалите. Ясно е, че този извод изцяло се обуславя от интерпретацията на условията на експеримента, съгласно които точката се избира “напосоки” върху интервала.

В този случай броят на елементарните изходи е безкраен, дори е континуум. В такива случаи вероятността е удобно да се пресмята чрез така наречената вероятностна плътност dp , която се определя като вероятност за попадане върху някакъв елементарен интервал с дължина dx : $dp = \frac{dx}{b - a}$. Тогава вероятността за попадане в интервал с крайна дължина е:

$p(A) = \int_{\alpha}^{\beta} dp = \frac{\beta - \alpha}{b - a}$, т.е. получава се резултатът, който вече знаем. Този подход (с интегриране

на вероятностната плътност) обаче е много по-мощен и позволява да се решават много по-сложни задачи.

Ето още един пример за геометрична вероятност: Във вътрешността на квадрат със страна a се избира по случаен начин точка. Каква е вероятността разстоянието от точката до центъра на квадрата да не е по-голямо от $\frac{a}{2}$?

Нека означим с A търсеното събитие (точката да е на разстояние до центъра на кръга, не по-голямо от $\frac{a}{2}$). Очевидно събитието A настъпва тогава и само тогава, когато точката лежи в кръг с радиус $\frac{a}{2}$ около центъра на квадрата. Тогава вероятността да настъпи A ще е отношението на площта на кръга (т.е. “благоприятните изходи”) към площта на квадрата (“всичките изходи”).

$$\text{Или: } p(A) = \frac{\pi \left(\frac{a}{2}\right)^2}{a^2} = \frac{\pi}{4}.$$

1.3. Комбинаторна вероятност. Правила за броене

Тук ще разгледаме по-подробно комбинаторната вероятност предвид голямата ѝ практическа важност. Припомняме, че условията за приложимост на комбинаторната вероятност са:

- броят на елементарните изходи от експеримента да е краен;
- всички елементарни изходи да са “равновъзможни” в смисъл, че няма основания да очакваме, че някой от тях ще настъпи по-често от другите. (Често понятията “равновъзможни” и равновероятни се използват като еквивалентни).

Типични примери за приложимост на комбинаторната вероятност са хазартните игри, където се вземат специални мерки за осигуряване на “равновъзможност” на изходите (балансиране на заровете и рулетките, еднаквост на гърбовете на картите за игра и пр.). В тези случаи вероятността за дадено събитие може да се дефинира просто като относителна честота, разделяйки броя на “равновъзможните” начини за настъпване на търсеното събитие към броя на всички “равновъзможни” изходи на експеримента.

Правилата за пресмятане на комбинаторната вероятност са известни още като правила за броене. Те указват начина, по който да се определи броят на изходите от експеримента и броят на благоприятните изходи.

В общия случай предполагаме, че броят на всички възможни изходи е $n \geq 1$, а броят на реализациите на експеримента е $k \geq 1$. Аналогията с изваждане на топки от урна тук е много полезна. Множеството от всички възможни изходи се нарича популация, а това от резултатите от експеримента – извадка с обем k .

Обемът на популацията е свойство на условията на експеримента и за всеки конкретен експеримент е фиксирано число. Напротив, броят на реализациите на експеримента зависи от (търпението на) експериментатора и може да е произволен.

В зависимост от условията на експеримента извадката може да бъде:

- с повторение – когато резултатът от експеримента може да бъде получаван неограничен брой пъти (изтеглената топка се връща в урната);
- без повторение – когато даден изход може да бъде реализиран само веднъж (т.е. изтеглена топка не се връща повече в урната).

Пример за извадка с повторение е хвърлянето на зарове, а за извадка без повторение – изтеглянето на числата на Тото2.

Очевидно при извадката с повторение нейният обем е неограничен, т.е. k може да е и по-голямо от n , докато при извадка без повторение винаги $k \leq n$.

Друга характерна особеност на извадките е тяхната наредба. Една извадка се нарича:

- наредена, ако за експеримента е от значение редът на последователно получаване на резултатите;
- ненаредена – в противния случай.

Пример за наредена извадка е съобщаването на числата от Тото2 в последователността, в която са изтеглени, тъй като всяка от топките в урната има свой номер. Напротив, ако всички топки са еднакви (бели), всяка извадка би била ненаредена.

Понятията, изброени по-горе, имат важни аналогии в статистическата физика, където се говори за ансамбли (т.е. популации) от различими или от неразличими частици (това е аналогията на наредени и ненаредени извадки съответно), или пък за системи, подчиняващи се на принципа на Паули, където броят на частиците, които могат едновременно да бъдат в едно и също състояние е ограничен (това е аналогия на извадка без връщане).

Премаваме към правилата за броене в зависимост от типовете извадки.

1.3.1. Наредена извадка без повторение

Първият елемент от извадката може да бъде получен без ограничения измежду всичките n елемента на популацията. Вторият елемент може да бъде “равновъзможно” избран между оставащите $n-1$ елемента. Това означава, че два определени елемента могат да бъдат получени

по $n \cdot (n-1)$ начина. Този процес може да се обобщи по очевиден начин. Така излиза, че k елемента могат да се получат измежду популация с обем n по $V_n^k = n \cdot (n-1) \cdot (n-2) \dots (n-k+1)$ начина. Числата V_n^k се наричат вариации от n елемента k -ти клас. Очевидно:

$$V_n^k = n \cdot (n-1) \cdot (n-2) \dots (n-k+1) = \frac{n!}{(n-k)!} \quad (3)$$

Тук, разбира се, $k \leq n$ и си спомняме, че по дефиниция $0! = 1$.

При $k = n$ имаме $V_n^n = n! \stackrel{def}{=} P_n$. В този случай различните изходи представляват само размествания на всички елементи на популацията на различни места. Числата P_n се наричат пермутации на n елемента.

Пример. От колода от 52 карти изтегляме последователно три. Каква е вероятността последната изтеглена карта да е дама?

Решение. Всичките начини, по които се получават три карти от 52 са $52 \cdot 51 \cdot 50 = 132600$. Това са именно всички изходи. За да преброим благоприятните изходи, предполагаме, че резултатът е налице, т.е. последната изтеглена карта е дама. Нека за конкретност тя е дама спатия. В такъв случай дамата спатия не е била изтеглена нито първия, нито втория път. Следователно, популацията за първото теглене се редуцира до 51 карти (всички без дамата спатия). От тази популация броят на възможните изходи за първите две тегления е $51 \cdot 50 = 2550$. По-нататък съобразяваме, че благоприятен изход се получава също ако третата изтеглена карта е друга дама (каро, купа или пика). Очевидно в тези случаи броят на изходите за първите две тегления е същия, т.е. търсената вероятност е:

$$p = \frac{V_{51}^2 + V_{51}^2 + V_{51}^2 + V_{51}^2}{V_{52}^3} = \frac{4 \cdot 51 \cdot 50}{52 \cdot 51 \cdot 50} = \frac{4}{52} = \frac{1}{13}. \quad \text{Сигурно трябва да се}$$

досетим, че този резултат може да се получи и по-просто. Как?

1.3.2. Наредена извадка с повторение

Първият елемент от извадката може да бъде равновъзможно получен измежду всичките n елемента. След това, за разлика от предишното разглеждане, го връщаме. При това вторият елемент се изтегля измежду популация със същия размер. Оттук следва, че първият и вторият елементи могат да бъдат изтеглени общо по n^2 начина. Тази аналогия може да се продължи неограничено, т.е. k може да бъде и по-голямо от n . Общият брой изходи е:

$$\bar{V}_n^k = n^k \quad (4)$$

Пример. Колко са различните четирибуквени думи, които могат да бъдат образувани от буквите на българската азбука?

Решение. Всяка от буквите може без ограничение да стои на всяка от четирите позиции на думата (ако се абстрахираме за момент от правилата на българската граматика!). Нашите букви са 30, така че броят на различните думи е:

$$\bar{V}_{30}^4 = 30^4 = 810000.$$

1.3.3. Ненаредена извадка без повторение

Означаваме търсения брой на ненаредените извадки без повторение с C_n^k . В раздел 1.3.1. ние получихме броя на наредените извадки без повторение (вариациите V_n^k). Явно броят на ненаредените извадки ще е по-малък от броя на наредените (защото поредиците от вида ABC, BAC, BCA, \dots ще са неразличими в нашия случай (подредбата на елементите не играе роля). По-точно, C_n^k ще е толкова пъти по-малко от V_n^k , по колкото начина може да се нареди една ненаредена извадка с обем k . Този брой обаче е известен – това е броят на пермутациите от k елемента или $P_k = k!$. Окончателно:

$$C_n^k = \frac{V_n^k}{P_k} = \frac{n!}{(n-k)!k!} \stackrel{\text{def}}{=} \binom{n}{k} \quad (5)$$

Последният символ е известният биномен коефициент. Числата C_n^k се наричат комбинации от n елемента k – ти клас.

Класическият пример в този случай е за пресмятане на броя на комбинациите, които трябва да попълни играч на Тото2, за да улучи със сигурност максималната печалба. Този брой съответствува на ненаредена извадка от 6 числа без повторение. Популацията е от всичко 49 числа. Така че резултатът е:

$$C_{49}^6 = \frac{49!}{6!(49-6)!} = 13983816.$$

Разбира се, заедно с максималната печалба от 6 познати числа при това ще се получат и всички по-малки печалби, при това точно по един път.

1.3.4. Ненаредена извадка с повторение

Броят на различните начини, по които може да се получи ненаредена извадка с обем k измежду n елемента с повторение, е:

$$\bar{C}_n^k = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!} \quad (6)$$

Този резултат привеждаме без доказателство.

1.4. Исторически бележки

Развитието на стохастиката и оформянето ѝ като математична дисциплина е тясно свързано с хазартните игри и по-специално с осъзнаването на факта, че резултатите от тях не могат да се опишат в рамките на известните дотогава класически раздели на математиката (алгебра, анализ, геометрия). Класическата теория на вероятностите е развита през XVII-XIX век по пътя на формализацията на интуитивните представи за вероятност, за които вече стана дума. За основатели на математичната теория на вероятностите се смятат Пиер Ферма (1601-1665) и Блез Паскал (1623-1662), които през 1654г. установяват някои от основните ѝ положения. Въз основа на тях Христиан Хюйгенс (1629-1695) публикува през 1657г. първия трактат по теория на вероятностите, озаглавен “За пресмятанята при хазартните игри”. За научната общност още тогава е било ясно, че резултатите на Ферма, Паскал и Хюйгенс далеч не се отнасят само до въпросите на хазартните игри, а полагат основите на нова дълбока математична теория. Следващата важна крачка принадлежи на Якоб Бернули (1654-1705) в неговия труд “Изкуството на предположението” (“*Ars conjectandi*”). Фундаменталните трудове на Авраам де Моавр (1667-1754) “Учение за случая” и особено на Пиер Симон Лаплас (1749-1827) “Аналитична теория на вероятностите” придават до голяма степен завършен вид на класическата теория на вероятностите.

След епохата на Лаплас интересът към теорията на вероятностите е отслабнал значително до степен да бъде поставена под съмнение принадлежността ѝ към математическите дисциплини. Главната причина за това е несигурната в логическо отношение основа на теорията. Такова е самото честотно определение за вероятност, което се основава на свой ред на понятието “равновероятен изход”, т.е. получава се логически порочна схема. Освен това широк клас случайни явления остават необясними в рамките на честотното определение. Такива са например геометричните вероятности, както и всички други експерименти с изходи, които образуват непрекъснато множество (интервал, област и др.).

По-нататъшното развитие на теорията на вероятностите е свързано с имената на Пафнутий Львович Чебышев (1821-1894), Андрей Андреевич Марков (1856-1922) и др. Съвременен вид този дял от математиката е получил през XX век, когато е създадена аксиоматичната теория на вероятностите на Андрей Николаевич Колмогоров (1903-1987) (“Основни понятия на теорията на вероятностите”, 1933).

2. Аксиоматична теория на вероятностите

Едно от големите ограничения на разгледаната дотук класическа вероятностна схема е изискването за равновъзможност на отделните изходи от експеримента (другият недостатък е от теоретичен порядък и е свързан с неудовлетворителния начин за дефиниране на вероятността като граница на честотата при очевидната невъзможност тази граница да бъде определена). В съвременната теория на вероятностите тези трудности се преодоляват и тя се изгражда дедуктивно, т.е. тръгвайки от определен малък брой начални предположения, които се приемат без доказателство (т.е. аксиоми). В тази глава ние ще дадем представа за основите на тази теория.

2.1. Пространство на елементарните събития. Алгебра на събитията

В общата теоретико-вероятностна схема се предполага, че за всеки експеримент със случаен изход трябва да бъдат указани всички възможни елементарни изходи, които се характеризират със следното основно свойство: при всяка реализация на експеримента е възможен един и само един от тези изходи. Всеки такъв изход се нарича още елементарно събитие; то е елементарно в смисъл, че не може да бъде разложено на “още по-елементарни”.

Множеството от всички елементарни събития се нарича пространство на елементарните събития Ω за дадения експеримент. Отделните елементарни събития ω се наричат точки в пространството на елементарните събития.

Всеки резултат от експеримента пък се нарича “събитие” (A). За всяко елементарно събитие ω може да се определи дали неговото настъпване води непременно до настъпването и на събитието A (бележи се $\omega \subset A$ и се казва “ ω влече A ”). По този начин, всяко конкретно събитие напълно се характеризира с множеството от всички елементарни събития, при които то настъпва. (Например, при хвърлянето на зар събитието “получаване на нечетно число” е сложно; то е съвкупност от събитията “1”, “3” и “5”, които са вече елементарни). Обратно, всяко множество A от точки $\omega \in \Omega$ може да се разглежда като събитие A , което или настъпва, или не в зависимост от това дали конкретният изход от експеримента ω_0 ще бъде точка от A или не. По този начин на всяко събитие A се съпоставя едно подмножество $A \in \Omega$ и те могат да се разглеждат като еквивалентни.

Сега пристъпваме към формулировката на естествените операции над събитията и на техните аналози от теорията на множествата. Съвкупността от тези операции характеризира алгебричната структура на пространството на елементарните събития.

1. Ако събитието A настъпва винаги, когато настъпва събитието B , казваме, че A е следствие от B и пишем: $B \subset A$ или $A \supset B$. В термините на теорията на множествата това означава, че всяка точка от B се съдържа в A или че B е подмножество на A .
2. Ако $A \subset B$ и $B \subset A$, то събитията A и B настъпват винаги едновременно. В такъв случай пишем $A = B$ и казваме, че събитията A и B са еквивалентни; при това множествата A и B съвпадат.
3. Събитието, което настъпва тогава и само тогава, когато не настъпва A , се нарича противоположно на A и се означава с \bar{A} . Множеството \bar{A} се състои от всички точки на Ω , които не принадлежат на A и се нарича допълнение на A .
4. Ако събитието A не съдържа нито едно елементарно събитие, то се нарича невъзможно събитие и се бележи с \emptyset . Противоположно на \emptyset е очевидно събитието Ω , което се случва винаги. То се нарича достоверно събитие. Може да се каже още, че \emptyset е празното подмножество на Ω .
5. Събитието C , настъпващо тогава и само тогава, когато настъпват събитията A и B , се нарича тяхно произведение или сечение на A и B и се бележи с AB или $A \cap B$. Множеството C се състои от всички точки, които принадлежат както на A , така и на B , и също се нарича сечение на множествата A и B .
6. Събитията A и B се наричат несъвместими, ако тяхното едновременно настъпване е невъзможно, т.е. ако $A \cap B = \emptyset$. На несъвместимите събития съответствуват непресичащи се (т.е. нямащи общи точки) множества.
7. Събитието C , състоящо се в настъпването на поне едно от събитията A и B , се нарича тяхна сума или обединение на A и B . За обединението се използва символичния запис $C = A \cup B$. Ако $A \cap B = \emptyset$, може да се напише още $C = A + B$. От гледна точка на теорията на множествата, множеството C се състои от всички точки, които принадлежат поне на едно от множествата A и B .

Допълнение: Обединението и сечението се дефинират за произволен брой събития. Например, събитието $C = A_1 \cup A_2 \cup \dots \cup A_n$ настъпва тогава, когато настъпи поне едно от събитията A_1, \dots, A_n . Аналогично се дефинира и сечение на повече от две събития.

Операциите обединение и сечение са комутативни:

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

и асоциативни:

$$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$$

$$(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$$

8. Събитието C , състоящо се в това, че настъпва събитие A и едновременно с това не настъпва B , се нарича разлика на събитията A и B и се бележи с $C = A \setminus B$. Изказано чрез термините на теорията на множествата, множеството $A \setminus B = A \cap \bar{B}$ се състои от всички точки, които принадлежат на множеството A и не принадлежат на B . То се нарича също разлика на множествата A и B .

Изброените 8 основни свойства на операциите над събитията дават възможност да бъдат извършвани и други операции. Ще отбележим специално взаимната дистрибутивност на операциите обединение (\cup) и сечение (\cap):

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

В сила са и следните равенства, известни като закони на де Морган:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

Всяко множество (клас) F от събития в Ω , за което:

- 1) $\emptyset \in F, \Omega \in F$;
- 2) за всяко събитие $A \in F$, е в сила и $\bar{A} \in F$;
- 3) за всеки две събития $A \in F, B \in F$ следва, че и $A \cup B \in F, A \cap B \in F$,

се нарича алгебра на събитията в пространството Ω .

Пример. Да се опитаме да определим алгебрата на събитията при хвърляне на зар. За простота ще номерираме събитията “попадане на числото 1” с 1 и т.н.

Пространството на елементарните събития се състои от 6 елементарни събития: $\Omega = \{1, 2, 3, 4, 5, 6\}$. Към алгебрата се отнася преди всичко (винаги!) празното множество \emptyset . След това трябва да добавим единичните събития 1, ..., 6 (6 на брой), обединенията по двойки от типа $1 \cup 2, \dots$ (броят на различните двойки е 15), обединенията по тройки от типа $(1 \cup 2) \cup 3$ (на брой 20), после обединенията по четворки (15), по петици (6) и накрая обединението на всички 6 елементарни събития $1 \cup 2 \cup 3 \cup 4 \cup 5 \cup 6$ (то е само едно). Общият брой на елементите от алгебрата (класа F) е $2^6 = 64$. По-общо, за всяко пространство Ω , състоящо се от n елементарни събития, F се състои от 2^n събития. Често се казва, че алгебрата е множеството от всички подмножества на основното пространство. Това е вярно твърдение.

Понятията пространство на елементарните събития и алгебрата на събитията дават възможност да бъдат въведени аксиомите на вероятността.

2.2. Аксиоми на вероятността

Вероятността може да се въведе и аксиоматично (Колмогоров) чрез един минимален брой аксиоми. Тук ние ще ги скицираме:

1. (неотрицателност): На всяко събитие A съответствува неотрицателно число, което се нарича негова вероятност: $P(A) \geq 0$;
2. (нормировка): Достоверното събитие Ω има вероятност единица: $P(\Omega) = 1$;
3. (адитивност): Ако събитията A и B са несъвместими (не могат да се случат едновременно), то вероятността за настъпване на A или B е: $P(A \cup B) = P(A) + P(B)$;
4. (непрекъснатост): За всяка редица от вложени събития, клоняща към празното множество, редицата от техните вероятности е сходяща с граница 0:

$$A_1 \supset A_2 \supset \dots \supset A_n \xrightarrow{n \rightarrow \infty} \emptyset \Rightarrow \lim_{n \rightarrow \infty} P(A_n) = 0.$$

От тези аксиоми лесно се получават някои важни следствия:

- $P(A) + P(\bar{A}) = 1$.
- $P(\emptyset) = 0$
- За всеки две събития A и B е в сила: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cup B) \leq P(A) + P(B)$ (неравенство на Бул)
- Ако $A \subset B$, то $P(A) \leq P(B)$.

2.3. Условна вероятност

До понятието условна вероятност се достига винаги, когато се интересуваме от вероятността за настъпване на някакво събитие B и имаме някаква допълнителна информация от експеримента, която се изразява в това, че някакво друго събитие A вече е настъпило. Тогава, ако $P(A) > 0$, то условната вероятност да се случи събитието B при условие, че вече е настъпило събитието A , се определя така:

$$P(B | A) = \frac{P(A \cap B)}{P(A)}$$

Оттук:

$$P(A \cap B) = P(A)P(B | A). \quad (1)$$

В тези изрази $A \cap B$ е сечението на събитието A със събитието B .

Установява се, че така дефинираната величина удовлетворява всички свойства на вероятността (както е определена чрез аксиомите на Колмогоров), т.е. условната вероятност е вероятност.

Изразът (1) се обобщава за повече от две събития по следния начин: Ако имаме $n+1$ събития A_0, A_1, \dots, A_n , за които $P(A_0 \cap A_1 \cap \dots \cap A_n) > 0$, то е в сила:

$$P(A_0 \cap A_1 \cap \dots \cap A_n) = P(A_0)P(A_1 | A_0)P(A_2 | A_0 \cap A_1) \dots P(A_n | A_0 \cap A_1 \cap \dots \cap A_{n-1}).$$

Този израз е известен като правило за умножение на условната вероятност.

Друго следствие е формулата за пълната вероятност за дадено събитие:

Нека даден експеримент води към някои от следните несъвместими събития:

$$\Omega = A_1 + A_2 + \dots + A_n$$

Тогава вероятността да се осъществи кое да е събитие B е:

$$P(B) = \sum_{i=1}^n P(A_i)P(B | A_i).$$

2.4. Стохастична независимост

Разгледаното в предния раздел понятие условна вероятност е мярка за влиянието на едно събитие A върху друго B . В практиката има много случаи, когато е налице такова влияние, макар че то може да бъде понякога много сложно (например влияе ли слънчевата активност върху количеството на валежите?). Има обаче случаи, когато настъпването (или ненастъпването) на едно събитие не зависи и не влияе на настъпването (ненастъпването) на друго събитие. В такива случаи ние казваме, че двете събития са независими едно от друго. Пощната дефиниция на независимите събития е свързана с едно основно тяхно свойство: две събития A и B се наричат независими помежду си, ако $P(A \cap B) = P(A)P(B)$. Това определение лесно се обобщава за повече независими събития.

Връзката между понятията независимост и условна вероятност се установява чрез следната теорема: Ако $P(A) > 0$, то необходимото и достатъчно условие събитията A и B да са независими е $P(B | A) = P(B)$, т.е. условната вероятност за настъпване на B при условие, че е настъпило A да е равна на самата (безусловна) вероятност $P(B)$, или с други думи $P(B | A)$ да не зависи от $P(A)$. Този резултат се установява лесно:

$$P(B | A) \stackrel{\text{def}}{=} \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B) \quad (\text{достатъчност}).$$

Докажете, че условието на теоремата е и необходимо.

III. Случайни величини

3. Вероятностни разпределения

Всяко случайно събитие може да се характеризира чрез някаква величина, която получава в резултат на настъпването му една от възможните си стойности. Тъй като при всяка реализация на един експеримент със случаен изход тази величина приема различни стойности (които са неизвестни отнапред), тя се нарича случайна. Случайните величини се делят на дискретни или непрекъснати според типа на множествата от стойности, които могат да заемат.

В абстрактната теория на вероятностите случайните величини се дефинират като функции X , дефинирани в пространството на елементарните събития Ω , чиито стойности са реални числа, определени по такъв начин, че за всяко реално число x множеството от елементарни събития ω , за които $X(\omega) \leq x$ е събитие, т.е. елемент от алгебрата на събитията в Ω .

Така дефинираните функции обаче са сложни за практическо използване, понеже нямат аналитични свойства. Поради тази причина вместо самите случайни величини се разглеждат други функции, свързани с тях - т.нар. функции на разпределение. Това е едно от главните различия между стохастиката и традиционния математичен анализ - в анализа функциите се разглеждат сами по себе си, а тук свойствата на случайните величини (които са функции), се разглеждат с помощта на други функции - функциите на разпределение. Оказва се, че познаването на вида на функцията на разпределение на една случайна величина позволява да бъдат изучени всички нейни свойства, и в този смисъл функциите на разпределение са предмет на централен интерес в стохастиката.

Функцията на разпределение се определя така:

ако \bar{x} е случайна величина и $x \in (-\infty, \infty)$ е реално число, то вероятността за събитието $(\bar{x} < x)$ е функция на x и се нарича функция на вероятностното разпределение на случайната величина \bar{x} :

$$F(x) = P(\bar{x} < x). \quad (1)$$

Чрез последователното увеличаване на аргумента x се обхващат последователно все повече елементарни събития от пространството Ω . Поради това, ако \bar{x} е дискретна случайна величина, $F(x)$ е стъпаловидна функция, която се изменя със скок всеки път когато се включи ново събитие от Ω . В общия случай $F(x)$ е монотонно нарастваща функция, т.е. за всяко $x_1 < x_2$ е в сила: $F(x_1) \leq F(x_2)$ (спомнете си следствието в раздел 2.2: Ако $A \subset B$, то $P(A) \leq P(B)$!).

От определението следва:

$$P(\bar{x} \geq x) = 1 - P(\bar{x} < x) = 1 - F(x) \quad (2)$$

Специален интерес представлява не само (и не толкова) функцията $F(x)$, а нейната производна по аргумента x (разбира се, ако такава съществува):

$$f(x) = \frac{dF(x)}{dx}, \quad (3)$$

която се нарича плътност на вероятностното разпределение (вероятностна плътност) на случайната величина \bar{x} . Тя е пропорционална на вероятността за събитието $(x \leq \bar{x} < x + dx)$ - т.е. след експеримента случайната величина \bar{x} да се окаже в един безкрайно малък интервал около числото x . Очевидно:

$$P(\bar{x} \leq a) = \int_{-\infty}^a f(x) dx = F(a) \quad (4)$$

$$P(a \leq \bar{x} \leq b) = \int_a^b f(x) dx = F(b) - F(a) \quad (5)$$

В частност:

$$\int_{-\infty}^{\infty} f(x) dx = 1 = P(\Omega) \quad (6)$$

3.1. Функции на случайни величини. Математическо очакване. Дисперсия. Моменти

Често ние се интересуваме освен от дадена случайна величина още и от някакви други величини, които са нейни функции. Всяка функция на една случайна величина е също случайна величина, доколкото също заема различни стойности при различни реализации на един и същ експеримент (в абстрактната теория на вероятностите това твърдение се доказва последователно за сума от случайни величини, умножение на сл. величина с число и т.н.). Следователно тя има също функция на вероятностно разпределение, както и вероятностна плътност.

Средна стойност (математическо очакване) на една дискретна случайна величина \bar{x} се нарича сумата от всевъзможните стойности на тази величина, умножени по съответните им вероятности:

$$E(\bar{x}) = \sum_{i=1}^n x_i P(\bar{x} = x_i) \quad (1)$$

Средната стойност на случайната величина обаче не е случайна величина. Тя всъщност не зависи от \bar{x} , тъй като в (1) се сумира по всички възможни стойности на случайната величина.

Средната стойност не е функция, а функционал на \bar{x} , т.е. правило, по което на една функция (случайната величина \bar{x}), се съпоставя едно число – нейното математическо очакване.

Ако $H(\bar{x})$ е някаква зададена функция на случайната величина \bar{x} , средната стойност на функцията $H(\bar{x})$ се определя така:

$$E\{H(\bar{x})\} \stackrel{\text{def}}{=} \sum_{i=1}^n H(x_i)P(\bar{x} = x_i) \quad (2)$$

При непрекъснати случайни величини тези формули се обобщават непосредствено:

$$E(\bar{x}) = \int_{-\infty}^{\infty} xf(x)dx \quad (3)$$

$$E\{H(\bar{x})\} = \int_{-\infty}^{\infty} H(x)f(x)dx \quad (4)$$

Сега ще разгледаме една функция от специален вид:

$$H(\bar{x}) = (\bar{x} - c)^k \quad (5)$$

Нейното математическо очакване:

$$\alpha_k = E\{(\bar{x} - c)^k\} \quad (6)$$

се нарича момент от ранг k на случайната величина \bar{x} спрямо точката c .

Особен интерес представляват моментите спрямо средната стойност на случайната величина:

$$\mu_k = E\{[\bar{x} - E(\bar{x})]^k\} = \int_{-\infty}^{\infty} [x - E(\bar{x})]^k f(x)dx \quad (7)$$

Те се наричат централни моменти на случайната величина \bar{x} . Причината за специалното предпочитание към централните моменти се основава на обстоятелството, че функцията $\alpha_k(c)$ има минимум при $c = E(\bar{x})$ поне при $k = 2$.

Очевидно, винаги е в сила: $\mu_0 = 1, \mu_1 = 0$. Следователно, μ_2 е най-низшият по порядък централен момент, който има нетривиална стойност. Той съдържа информация за отклонението на случайната величина от средната ѝ стойност. Централният момент от втори порядък μ_2 се нарича дисперсия (variance) на случайната величина \bar{x} . Ето няколко синонимни означения, използвани в различни книги:

$$\mu_2(\bar{x}) = \sigma^2(\bar{x}) = \text{var}(\bar{x}) = E\{[\bar{x} - E(\bar{x})]^2\}$$

Моментите на случайните величини имат и физична интерпретация:

При множество измервания на една и съща случайна величина резултатите се групират предимно около една “истинска” стойност на тази величина. Или, казано по-точно, по-вероятно

е да се получат (и по-често се получават) стойности близо до някакво число (напр. x_0), отколкото такива, които са далеч от него. Вероятностната плътност на случайната величина тогава ще има типичната и често срещана камбановидна форма. Математическото очакване може да се интерпретира като оценка за действителната стойност на измеряемата величина. Изразът за него прилича на този за координатите на центъра на масите на едно (едномерно) тяло. Дисперсията пък в такъв случай може да се интерпретира като израз за инерционния момент на такова тяло (относно центъра на масите му), и тя е мярка за *разсейването* на резултатите от измерванията около средната им стойност. Ако дисперсията е малка, измерванията могат да се разглеждат като по-точни (тъй като по-голяма част от тях са близо до средната стойност).

Дисперсията е квадратична мярка (тя има размерността на квадрата на изследваната случайна величина). Затова се използва величината $\sigma(\bar{x}) = +\sqrt{\mu_2(\bar{x})}$, която се нарича средноквадратично (стандартно) отклонение на случайната величина \bar{x} . Обикновено именно тази величина се посочва като мярка за *грешката* (правилният термин е статистическата неопределеност) на \bar{x} и се пише: $x = E(\bar{x}) \pm \sigma(\bar{x})$.

В много случаи математическото очакване и дисперсията не дават достатъчна информация за изследваните случайни величини; затова за по-пълното им характеризиране се използват и моменти от по-висок ред.

Третият централен момент μ_3 характеризира несиметричността на вероятностното разпределение около математическото очакване. Прието е безразмерната величина

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \quad (8)$$

да се използва за количествена характеристика на различията между относителния дял на положителните и отрицателните отклонения от средната стойност. Тази величина се нарича асиметрия (skew) на разпределението на случайната величина. За симетрични разпределения (които се представят с четни функции на вероятностната плътност) $\gamma_1 = 0$.

Четвъртият централен момент $\mu_4 = E\{[x - E(\bar{x})]^4\}$ определя характера на максимума в точката на математическото очакване при симетрични разпределения. За количествена характеристика се приема безразмерната величина

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 \quad (9)$$

която се нарича ексцес (kurtosis).

Величините асиметрия и ексцес са въведени чрез равенствата (8) и (9) така, че за най-често срещаното разпределение - нормалното - те да са нули: $\gamma_1 = 0$, $\gamma_2 = 0$. При $\gamma_1 > 0$ по-голямата

част от вероятностното разпределение се намира вдясно от математическото очакване (респ. обратното). При $\gamma_2 > 0$ пък максимумът на вероятностната плътност е по-остър от този при нормалното разпределение с параметри като в равенството (9), съответно при $\gamma_2 < 0$ максимумът на разпределението е по-тъп от този на нормалното разпределение.

По-висшите моменти на разпределенията рядко се използват. Трябва да се отбележи, че колкото по-голям е рангът k на даден момент, толкова по-голям е приносът на отдалечените от средната стойност участъци от функцията на вероятностната плътност (тъй наречените "опашки" на разпределението) при определянето на k -ия момент.

Освен важността на отделните моменти, съществено теоретично значение има и пълният набор от *всички* моменти (централни или нецентрални) на даденото разпределение.

Нецентралните моменти се дефинират така:

$$\mu'_k = \int_{-\infty}^{\infty} x^k f(x) dx \quad (10)$$

Доказва се (това ще видим малко по-нататък), че при много общи условия (които са изпълнени за всички разпределения, представляващи практически интерес) съвкупността от всички моменти напълно определя вероятностното разпределение. Има случаи, когато е по-лесно да се построят моментите на едно неизвестно разпределение, отколкото самото разпределение (т.е. вероятностната му плътност). Това позволява тогава да се построи и самото разпределение чрез неговите моменти.

В сила е по-конкретно следното твърдение (теорема):

Ако $f_1(x)$ и $f_2(x)$ са две вероятностни плътности, всичките съответни нецентрални моменти на които са равни, и ако разликата $f_1(x) - f_2(x)$ е разложима в степенен ред, то $f_1(x) \equiv f_2(x)$.

Ние ще установим верността на това твърдение. Нека означим:

$f_1(x) - f_2(x) = c_0 + c_1x + c_2x^2 + \dots$ и нека моментите на двете функции са съответно $\mu'_1 = \mu''_1, \mu'_2 = \mu''_2, \dots$

Тогава:

$$\begin{aligned} \int_{-\infty}^{\infty} [f_1(x) - f_2(x)]^2 dx &= \int_{-\infty}^{\infty} [f_1(x) - f_2(x)](c_0 + c_1x + c_2x^2 + \dots) dx \\ &= c_0 \int_{-\infty}^{\infty} [f_1(x) - f_2(x)] dx + c_1 \int_{-\infty}^{\infty} x[f_1(x) - f_2(x)] dx + c_2 \int_{-\infty}^{\infty} x^2[f_1(x) - f_2(x)] dx + \dots \\ &= c_0(1-1) + c_1(\mu'_1 - \mu''_1) + c_2(\mu'_2 - \mu''_2) + \dots = 0, \end{aligned}$$

откъдето следва, че $f_1(x) \equiv f_2(x)$ (тъй като подинтегралната функция е неотрицателна и тогава горният интеграл е 0 тогава и само тогава, когато самата подинтегрална функция е тъждествено равна на нула).

При зададена вероятностна плътност $f(x)$ пресмятането на моментите $\{\mu(\bar{x})\}$ не представлява проблем. Проблем представлява обаче често самото намиране на вероятностната плътност, която да описва адекватно процеса на измерването на изучаваната случайна величина.

3.2. Свойства на математическото очакване и дисперсията

1. Ако $c = const$, то:

$$E(c) = c$$

$$E(c\bar{x}) = cE(\bar{x}) \quad (1)$$

$$\sigma^2(c) = 0$$

$$\sigma^2(c\bar{x}) = c^2\sigma^2(\bar{x}) \quad (2)$$

От последното равенство следва:

$$\sigma^2(\bar{x}) = E\{[\bar{x} - E(\bar{x})]^2\} = E\{\bar{x}^2 - 2\bar{x}E(\bar{x}) + E^2(\bar{x})\} = E(\bar{x}^2) - E^2(\bar{x}) \quad (3)$$

2. Разглеждаме функцията

$$\bar{u} = \frac{\bar{x} - E(\bar{x})}{\sigma(\bar{x})} \quad (4)$$

Тази функция е също случайна величина. Нейното математическо очакване е:

$$E(\bar{u}) = \frac{1}{\sigma(\bar{x})} \{E[\bar{x} - E(\bar{x})]\} = \frac{1}{\sigma(\bar{x})} [E(\bar{x}) - E(\bar{x})] = 0 \quad (5)$$

Дисперсията на случайната величина (4) е:

$$\sigma^2(\bar{u}) = \frac{1}{\sigma^2(\bar{x})} E\{[\bar{x} - E(\bar{x}) - 0]^2\} = \frac{\sigma^2(\bar{x})}{\sigma^2(\bar{x})} = 1 \quad (6)$$

Така определената функция \bar{u} се нарича приведена (стандартизирана, нормирана) случайна величина. Такива величини са удобни с това, че математическото им очакване е винаги нула, а дисперсията - винаги единица. Таблиците и процедурите за компютърни пресмятания на статистическите разпределения като правило се дават именно за нормираните случайни величини. Самите случайни величини (ненормираните) се получават от нормираните чрез прости линейни трансформации (решавайки ур.(4) спрямо \bar{x}).

Ето някои други мерки за характеристика на местоположението в статистическите разпределения:

мод, мода (mode) се нарича стойността x_m на аргумента, която съответствува на максималната вероятност: $P(\bar{x} = x_m) = \max$. Тя може да се определи чрез диференциране на вероятностната плътност по аргумента. Разпределенията с един максимум се наричат унимодални. Такива са почти всички разпределения, които представляват интерес във физиката;

медиана (median) се нарича стойността на аргумента $x(0.5)$, която съответствува на $F(x(0.5)) = P(\bar{x} < x(0.5)) = 0.5$. Медианата дели дефиниционната област на вероятностната плътност на две части, във всяка от които случайната величина попада с равна вероятност (т.е. това са две равноплощни части на разпределението). За симетричните разпределения математическото очакване, модата и медианата съвпадат;

по-общо квантил (quantile) на ниво q , $0 < q < 1$ се нарича стойността на аргумента x_q , за която: $F(x_q) = P(\bar{x} < x_q) = q$. Така че медианата е квантилът на ниво $1/2$. Квантилите на нивата $1/4$, $2/4$, $3/4$ се наричат квартили, на нивата 0.1 , $0.2, \dots, 0.9$ се наричат децили, а на нивата 0.01 , $0.02, \dots, 0.99$ - процентили. За често използваните разпределения най-важните квантили са дадени в таблици и има готови изчислителни процедури за тяхното пресмятане.

4. Разпределения на няколко случайни величини

4.1. Разпределения на две случайни величини

При статистическия анализ на експериментални данни рядко се оказва възможно да се ограничим с разглеждането само на една случайна величина. Дори и да се опитваме да измерваме само една физична величина (параметър), ние се натъкваме на наличието и на други фактори, които влияят върху резултатите от измерването и които също се описват със случайни величини (типичен пример са измерванията на някакъв сигнал при наличието на фон). Адекватното описание на такива експерименти предполага едновременното изучаване на статистическите свойства на всички величини, които реално участвуват във формирането на резултатите. Затова е нужно да се изучават *съвместните им вероятностни разпределения*. При това възникват и някои нови свойства, които не съществуват при едномерните разпределения.

Най-простият вариант е на две случайни величини, \bar{x} , \bar{y} . Ще ни интересува вероятността едновременно $\bar{x} < x$, $\bar{y} < y$. Както и при една случайна величина, предполагахме, че съществува функция на разпределение:

$$F(x, y) = P(\bar{x} < x, \bar{y} < y) \quad (1)$$

Във всички интересни случаи $F(x, y)$ съществува. Ако тя е диференцируема по двата си аргумента, то функцията:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \quad (2)$$

се нарича съвместна вероятностна плътност на случайните величини \bar{x}, \bar{y} . Тогава, очевидно:

$$P(a \leq x < b, c \leq y < d) = \int_a^b dx \int_c^d f(x, y) dy \quad (3)$$

Понякога възниква следният проблем: в резултат на измервания се получават множество стойности на случайните величини \bar{x}, \bar{y} . По тях може приблизително да се построи съвместната функция на разпределение $F(x, y)$, респективно съвместната вероятностна плътност $f(x, y)$. Ние обаче можем да разглеждаме проблем, при който да се интересуваме само от поведението на величината \bar{x} , независимо от \bar{y} . Плътността на вероятностното разпределение по \bar{x} може да се получи чрез интегриране на $f(x, y)$ по y в цялото пространство на изменение на y :

$$P(a \leq x < b, -\infty < y < \infty) = \int_a^b dx \left[\int_{-\infty}^{\infty} f(x, y) dy \right] = \int_a^b g(x) dx \quad (4)$$

Тук:

$$g(x) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(x, y) dy \quad (5)$$

е вероятностната плътност на случайната величина \bar{x} . $g(x)$ се нарича безусловна (маргинална) плътност на разпределението на \bar{x} . По подобен начин за \bar{y} се дефинира функцията:

$$h(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad (6)$$

Аналогично на понятието независими събития ние можем да определим независимостта на случайните величини \bar{x} и \bar{y} . Именно, двете случайни величини \bar{x} и \bar{y} се наричат независими, ако е изпълнено условието:

$$f(x, y) = g(x)h(y) \quad (7)$$

Условната вероятностна плътност също се определя с помощта на маргиналните разпределения:

$$f(y|x) = \frac{f(x, y)}{g(x)} \quad (8)$$

се нарича условна плътност на вероятността за случайната величина \bar{y} при зададено x .

За пълната вероятностна плътност е в сила съотношението:

$$h(y) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(x, y) dx = \int_{-\infty}^{\infty} f(y | x) g(x) dx \quad (9)$$

За независими случайни величини:

$$f(y | x) = \frac{g(x)h(y)}{g(x)} = h(y) \quad (10)$$

т.е. знанията ни за статистическото разпределение на случайната величина \bar{x} не увеличават информацията за разпределението на \bar{y} (и обратно).

4.2. Моменти на съвместните разпределения

Аналогично на едномерния случай, определяме математическото очакване на една дадена функция $H(\bar{x}, \bar{y})$ на случайните величини \bar{x} и \bar{y} като:

$$E\{H(\bar{x}, \bar{y})\} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} H(x, y) f(x, y) dy \quad (1)$$

а дисперсията:

$$\sigma^2 \{H(\bar{x}, \bar{y})\} = E\{[H(\bar{x}, \bar{y}) - E\{H(\bar{x}, \bar{y})\}]^2\} \quad (2)$$

Нека разгледаме функцията:

$$H(\bar{x}, \bar{y}) = \bar{x}^l \bar{y}^m \quad (3)$$

Математическото очакване на тази функция се нарича *(lm)-ти момент на вероятностното разпределение на случайните величини \bar{x} и \bar{y} спрямо точката (0,0)*:

$$\lambda_{lm} \stackrel{\text{def}}{=} E\{\bar{x}^l \bar{y}^m\} \quad (4)$$

Това е нецентрален момент. По-общо се разглежда функцията:

$$H(\bar{x}, \bar{y}) = (\bar{x} - a)^l (\bar{y} - b)^m \quad (5)$$

Нейното математическо очакване определя моментите на разпределението на \bar{x} и \bar{y} относно точката-център (a, b) :

$$\alpha_{lm} \stackrel{\text{def}}{=} E\{(\bar{x} - a)^l (\bar{y} - b)^m\} \quad (6)$$

Особен интерес представляват моментите μ_{lm} спрямо точката $(\lambda_{10}, \lambda_{01})$, която е точно математическото очакване на $f(x, y)$. Младшите по ранг моменти са:

$$\begin{aligned} \lambda_{00} &= 1 \\ \lambda_{10} &= E\{\bar{x}\} \stackrel{\text{def}}{=} \hat{x} \\ \lambda_{01} &= E\{\bar{y}\} \stackrel{\text{def}}{=} \hat{y} \end{aligned} \quad (7)$$

$$\begin{aligned} \mu_{00} &= 1 \\ \mu_{10} &= 0 \\ \mu_{01} &= 0 \\ \mu_{11} &= E\{(\bar{x} - \hat{x})(\bar{y} - \hat{y})\} \stackrel{def}{=} \text{cov}\{\bar{x}, \bar{y}\} \\ \mu_{20} &= E\{(\bar{x} - \hat{x})^2\} = \sigma^2\{\bar{x}\} \\ \mu_{02} &= E\{(\bar{y} - \hat{y})^2\} = \sigma^2\{\bar{y}\} \end{aligned} \tag{8}$$

4.2.1. Свойства на моментите на съвместните разпределения

При повече от една случайна величина може да се говори и за по-сложни свойства на моментите, свързани с операциите с няколко променливи. Ето някои от тях:

$$E\{a\bar{x} + b\bar{y}\} = aE\{\bar{x}\} + bE\{\bar{y}\} = a\hat{x} + b\hat{y} \tag{9}$$

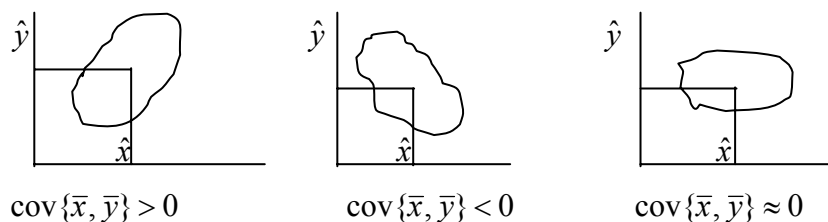
(следва непосредствено от дефиницията);

$$\begin{aligned} \sigma^2\{a\bar{x} + b\bar{y}\} &= E\{[(a\bar{x} + b\bar{y}) - E\{a\bar{x} + b\bar{y}\}]^2\} \\ &= E\{[a\bar{x} + b\bar{y} - a\hat{x} - b\hat{y}]^2\} \\ &= E\{[a^2(\bar{x} - \hat{x})^2 + b^2(\bar{y} - \hat{y})^2 + 2ab(\bar{x} - \hat{x})(\bar{y} - \hat{y})]\} \end{aligned} \tag{10}$$

Или:

$$\sigma^2\{a\bar{x} + b\bar{y}\} = a^2\sigma^2(\bar{x}) + b^2\sigma^2(\bar{y}) + 2ab \text{cov}\{\bar{x}, \bar{y}\} \tag{11}$$

Величините $E\{\bar{x}\}, E\{\bar{y}\}, \sigma^2\{\bar{x}\}, \sigma^2\{\bar{y}\}$, които участвуват тук, са много сходни с математическото очакване и дисперсията за едномерния случай. Що се отнася до $\text{cov}\{\bar{x}, \bar{y}\}$, която се нарича ковариация на \bar{x} и \bar{y} , тя изисква специално внимание предвид голямата ѝ важност. От дефиницията ѝ следва, че ковариацията е положителна, ако събитията $\bar{x} > \hat{x} \cap \bar{y} > \hat{y}$ (респективно $\bar{x} < \hat{x} \cap \bar{y} < \hat{y}$) се появяват по-често едновременно, отколкото комбинациите на малки \bar{x} с големи \bar{y} . Ковариацията е отрицателна в обратния случай, т.е. когато по-често малки (спрямо средното) стойности на \bar{x} се наблюдават едновременно с големи стойности на \bar{y} . Накрая, $\text{cov}\{\bar{x}, \bar{y}\} \approx 0$, ако знанието на \bar{x} не ни дава информация за стойностите на \bar{y} (и обратно). Това се пояснява на следните картинки:



В много случаи е удобно вместо ковариацията да се използва величината:

$$\rho(\bar{x}, \bar{y}) \stackrel{def}{=} \frac{\text{cov}(\bar{x}, \bar{y})}{\sigma(\bar{x})\sigma(\bar{y})} \quad (12)$$

която се нарича коэффициент на корелация.

Ковариацията и коэффициентът на корелация дават обобщена оценка за взаимната зависимост между случайните величини \bar{x} и \bar{y} .

За да изясним смисъла на коэффициента на корелация, разглеждаме нормираните случайни величини:

$$\bar{u} = \frac{\bar{x} - \hat{x}}{\sigma(\bar{x})}; \quad \bar{v} = \frac{\bar{y} - \hat{y}}{\sigma(\bar{y})} \quad (13)$$

Дисперсията на сумата $\bar{u} + \bar{v}$ е:

$$\sigma^2(\bar{u} + \bar{v}) = \sigma^2(\bar{u}) + \sigma^2(\bar{v}) + 2\rho(\bar{u}, \bar{v})\sigma(\bar{u})\sigma(\bar{v}) \quad (14)$$

Тъй като за нормираните случайни величини дисперсията е винаги 1, то:

$$\sigma^2(\bar{u} + \bar{v}) = 2[1 + \rho(\bar{u}, \bar{v})] \quad (15)$$

По аналогичен начин за разликата $\bar{u} - \bar{v}$ се получава:

$$\sigma^2(\bar{u} - \bar{v}) = 2[1 - \rho(\bar{u}, \bar{v})] \quad (16)$$

Но дисперсията е винаги неотрицателна, $\sigma^2 \geq 0$. Тогава от (15) и (16) следва:

$$-1 \leq \rho(\bar{u}, \bar{v}) \leq 1 \quad (17)$$

В сила са следните равенства:

$$\begin{aligned} \rho(\bar{u}, \bar{v}) &\stackrel{def}{=} \frac{\text{cov}(\bar{u}, \bar{v})}{1 \cdot 1} \stackrel{def}{=} E\{(\bar{u} - 0)(\bar{v} - 0)\} = E\left\{\frac{(\bar{x} - \hat{x})}{\sigma(\bar{x})} \cdot \frac{(\bar{y} - \hat{y})}{\sigma(\bar{y})}\right\} \\ &= \frac{\text{cov}(\bar{x}, \bar{y})}{\sigma(\bar{x})\sigma(\bar{y})} = \rho(\bar{x}, \bar{y}) \end{aligned} \quad (18)$$

Следователно, неравенството (17) е в сила за всеки две случайни величини (не непременно нормирани):

$$-1 \leq \rho(\bar{x}, \bar{y}) \leq 1 \quad (19)$$

Оттук разбираме още, че коэффициентът на корелация ρ е и нормиран; той не зависи от мащаба по x и по y (за разлика от ковариацията).

Граничните случаи са:

$$\text{а) } \rho = +1, \Rightarrow \sigma^2(\bar{u} - \bar{v}) = 0, \Rightarrow \bar{u} - \bar{v} = \frac{\bar{x} - \hat{x}}{\sigma(\bar{x})} - \frac{\bar{y} - \hat{y}}{\sigma(\bar{y})} = \text{const} \quad (20)$$

Това е *функционално* равенство. То трябва да бъде изпълнено за всяко \bar{x} и всяко \bar{y} , които могат да се изменят независимо едно от друго. Това е възможно само ако е в сила:

$$\bar{y} = a + b\bar{x} \quad (21)$$

т.е. случайните величини \bar{x} и \bar{y} са линейно свързани и $b > 0$. Равенството (21) съответства на случая на пълна линейна корелация между случайните величини \bar{x}, \bar{y} ;

б) напълно аналогично при $\rho = -1, \Rightarrow \sigma^2(\bar{u} + \bar{v}) = 0$, откъдето следва:

$$\bar{y} = a - b\bar{x} \quad (22)$$

който случай се нарича пълна линейна антикорелация.

Накрая, ако случайните величини \bar{x} и \bar{y} са независими, то $\text{cov}(\bar{x}, \bar{y}) = \rho(\bar{x}, \bar{y}) = 0$.

Наистина, ако $f(x, y) = g(x)h(y)$, тогава:

$$\text{cov}(\bar{x}, \bar{y}) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} (x - \hat{x})(y - \hat{y})g(x)h(y)dy = \left\{ \int_{-\infty}^{\infty} (x - \hat{x})g(x)dx \right\} \left\{ \int_{-\infty}^{\infty} (y - \hat{y})h(y)dy \right\} = 0.$$

Обратното обаче не е в сила, т.е. ако ковариацията на две случайни величини е нула, те *не са* непременно независими (!).

4.3. Съвместни разпределения на повече от две случайни величини

Съвместната функция на разпределение на n случайни величини $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$ се определя така:

$$F(x_1, x_2, \dots, x_n) = P(\bar{x}_1 < x_1, \bar{x}_2 < x_2, \dots, \bar{x}_n < x_n) \quad (1)$$

Съвместната вероятностна плътност се дефинира по следния начин:

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n F}{\partial x_1 \partial x_2 \dots \partial x_n} \quad (2)$$

По аналогичен начин се въвеждат всички останали характеристики: маргинално разпределение, математическо очакване, дисперсия, моменти и т.н. Този случай не ни дава нищо ново, но позволява да се направи обобщено разглеждане. Именно, случайните величини $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$ могат да се разглеждат като компоненти на вектор в n -мерното пространство:

$x = \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix}$. Това е вектор-стълб. Транспонираният му вектор е: $x^T = (x_1, x_2, \dots, x_n)$. Тогава

функцията на разпределение се записва така: $F = F(x)$, съответната вероятностна плътност:

$f(x) = \frac{\partial F(x)}{\partial x}$. Математическото очакване на една функция $H(x)$ добива вида:

$$E\{H(x)\} = \int H(x)f(x)dx. \quad (3)$$

Тук навсякъде x е вектор с n компоненти, интегрирането е също в n -мерното пространство.

Вторите централни моменти - дисперсиите и ковариациите - могат да се обединят в една матрица с размерност $n \times n$, която се нарича ковариационна матрица:

$$c = \begin{pmatrix} c_{11} & \dots & c_{1n} \\ \dots & \dots & \dots \\ c_{n1} & \dots & c_{nn} \end{pmatrix} \quad (4)$$

Нейните елементи са:

$$c_{ij} = \text{cov}(\bar{x}_i, \bar{x}_j) = E\{(\bar{x}_i - \hat{x}_i)(\bar{x}_j - \hat{x}_j)\} \quad (5)$$

Диагоналните елементи $c_{ii} = \sigma^2(x_i)$ са дисперсиите (които се наричат още вариации).

Матрицата c е симетрична по определение (5).

Като си дадем сметка, че множителите в (5) са компоненти на n -мерен вектор, от правилото за умножение на матрици се получава:

$$c = E\{(x - \hat{x})(x - \hat{x})^T\} \quad (6)$$

Ковариационната матрица (наричана понякога вариационно-ковариационна матрица, а също и матрица на грешките) съдържа основната информация за разсейването на компонентите на един n -мерен случаен вектор. По тази причина нейното пресмятане е една от главните задачи при обработката на експерименталните данни.

5. Трансформации на случайни величини. Разпространение на грешките

Стана ясно, че функцията на една случайна величина е също случайна величина: $\bar{y} = y(\bar{x})$. Въпросът, който възниква често, е: каква е вероятностната плътност $g(y)$, ако е известна плътността $f(x)$?

Общият отговор на този въпрос е свързан с изискването за взаимна еднозначност и непрекъснатост на изображенията $y(x)$ и $x(y)$ и непрекъснатост на производната $\frac{dx}{dy}$. Тогава търсената връзка е от вида:

$$g(y) = \left| \frac{dx}{dy} \right| f(x) \quad (1)$$

Съотношението (1) има и многомерно обобщение. Ние няма да се занимаваме с общия случай, а ще разглеждаме само най-важния специален случай на линейни трансформации.

Линейните трансформации са най-често използваните на практика. Причината е в простотата на действията с тях. Те например водят до линейни уравнения и системи, които

много по-лесно се решават. Затова при възможност ние се стараем да преинемем от нелинейни към линейни трансформации. Обикновено това се постига чрез ограничаване на Тейлоровото (Taylor) разложение на функциите до линейни членове.

Нека x и y са n -мерни вектори от случайните величини $(\bar{x}_1, \dots, \bar{x}_n)$ и $(\bar{y}_1, \dots, \bar{y}_n)$. Нека y е линейна функция на x . Това се записва в матричен вид така:

$$y = a + Tx \quad (2)$$

където a е n -мерен вектор, а T – $n \times n$ матрица.

Математическото очакване на случайната величина y е:

$$E(y) \stackrel{def}{=} \hat{y} = a + T\hat{x} \quad (3)$$

Нека сега да видим как ще изглежда ковариационната матрица на линейната трансформация $y(x)$:

$$\begin{aligned} C_y &= E\{(y - \hat{y})(y - \hat{y})^T\} \\ &= E\{(a + Tx - a - T\hat{x})(a + Tx - a - T\hat{x})^T\} \\ &= E\{T(x - \hat{x})(x - \hat{x})^T T^T\} = TE\{(x - \hat{x})(x - \hat{x})^T\}T^T = TC_x T^T \end{aligned} \quad (4)$$

където C_x е ковариационната матрица на случайната величина x

(Тук помним, че матричното умножение не е комутативно!).

И така:

$$\boxed{C_y = TC_x T^T} \quad (5)$$

С помощта на съотношението (5) ще формулираме едно от най-известните правила в статистиката, наречено закон за разпространение на грешките (англ. error propagation).

Нека $x = (\bar{x}_1, \dots, \bar{x}_n)$ е n -вектор от известни измервания. Ние предполагаме, че:

- средните стойности на измерванията са известни; нека те образуват вектора \hat{x} ;
- средноквадратичните отклонения ("грешките") на x са също известни; те образуват ковариационната матрица C_x .

Ние искаме да разберем каква е "грешката" на някаква функция на измерването $y = y(x)$.

Ако грешките на x са сравнително малки, то вероятностната му плътност $f(x)$ ще е съществено различна от нула само в неголяма околност на \hat{x} (тази околност е от порядъка на $\sigma(x)$). Тогава ще е в сила разложението:

$$y_i = y_i(\hat{x}) + \sum_{j=1}^n \left. \frac{\partial y_i}{\partial x_j} \right|_{x=\hat{x}} (x_j - \hat{x}_j) + O(x - \hat{x})^2, \quad i = 1, \dots, n \quad (6)$$

Или, в матрична форма:

$$y = y(\hat{x}) + T(x - \hat{x}) + O(x - \hat{x})^2, \quad T_{ij} = \left. \frac{\partial y_i}{\partial x_j} \right|_{x=\hat{x}} \quad (7)$$

Тогава, използвайки (5), получаваме:

$$\boxed{C_y = TC_x T^T} \quad (8)$$

където матрицата на трансформацията T се определя от (7b). Равенството (8) изразява закона за разпространение на грешките при измерванията.

Казаното дотук може да се обобщи така:

- законът за разпространение на грешките (5) е точен за линейни трансформации на случайните величини;
- този закон може да се използва приближено и за нелинейни трансформации на случайни величини във вида (8). Това приближение е добро, когато дисперсиите на базисните случайни величини $\{x\}$ са малки.

Тук трябва да се направи следната важна забележка:

Дисперсиите са, както знаем, диагоналните елементи на съответната ковариационна матрица. Тъй като обикновено ние се интересуваме главно от дисперсиите като мярка за разсейването на резултатите от измерванията, ние сме склонни да оценяваме само $\sigma^2(x)$. Но, забележете, че дисперсиите на трансформациите y зависят не само от дисперсиите на x , но и от ковариациите $\text{cov}(x_i, x_j)$. Без отчитането на ковариациите между измерванията (т.е. техните взаимни зависимости), ние можем да получим съвършено нереалистични резултати за $\sigma^2(y)$. Само ако наблюденията $\{x\}$ са взаимно независими помежду си, тогава $\text{cov}(x_i, x_j) = 0$ ($i \neq j$) и тогава те могат да се пренебрегнат. Тогава матрицата C_x става диагонална и е в сила:

$$(C_y)_{ii} = \sigma^2(y_i) = \sum_{j=1}^n \left(\frac{\partial y_i}{\partial x_j} \right)^2 \sigma^2(x_j) \quad (9)$$

Последното съотношение е известно също като закон за разпространение на грешките и фигурира в много учебни пособия именно в този вид. Напомняме още веднъж, че (9) е валидно само при независими измервания $\{x\}$.

Ето формули за дисперсията на някои прости функции от случайни величини, пресметнати според закона за разпространение на грешките:

Нека x_1 и x_2 са две независими случайни величини. Тогава:

$y = x_1 \pm x_2$	$\sigma^2(y) = \sigma^2(x_1) + \sigma^2(x_2)$
$y = x_1 \cdot x_2, \quad y = \frac{x_1}{x_2}$	$\frac{\sigma^2(y)}{y^2} = \frac{\sigma^2(x_1)}{x_1^2} + \frac{\sigma^2(x_2)}{x_2^2}$
$y = x^n$	$\frac{\sigma^2(y)}{y^2} = \frac{n^2 \sigma^2(x)}{x^2}$
$y = e^x$	$\frac{\sigma^2(y)}{y^2} = \sigma^2(x)$

Накратко тези правила могат да се изкажат така:

- при събиране и изваждане се сумират дисперсиите на събираемите;
- при умножение и деление се сумират относителните дисперсии на множителите/делителите.

Ако x_1 и x_2 **не са независими**, то съответните формули за дисперсията на y в случая на произведение и частно са:

$y = x_1 \cdot x_2$	$\frac{\sigma^2(y)}{y^2} = \frac{\sigma^2(x_1)}{x_1^2} + \frac{\sigma^2(x_2)}{x_2^2} + 2\rho(x_1, x_2) \frac{\sigma(x_1)}{x_1} \frac{\sigma(x_2)}{x_2}$
$y = \frac{x_1}{x_2}$	$\frac{\sigma^2(y)}{y^2} = \frac{\sigma^2(x_1)}{x_1^2} + \frac{\sigma^2(x_2)}{x_2^2} - 2\rho(x_1, x_2) \frac{\sigma(x_1)}{x_1} \frac{\sigma(x_2)}{x_2}$

С други думи, дисперсията на произведението се увеличава при положителна корелация между множителите, а дисперсията на частното намалява при положителна корелация между делителите (в сравнение с дисперсиите за независими множители/делители). При антикорелация между множителите/делителите е обратното.

IV. Някои важни статистически разпределения

Тук ще разгледаме свойствата на някои конкретни статистически разпределения, които се срещат най-често в практиката на обработката на данни.

6. Дискретни разпределения

6.1. Експеримент с два изхода (разпределение на Бернули)

Най-простият възможен експеримент е този, чиято реализация води до едно от двете несъвместими събития A и \bar{A} (например хвърляне на монета). Той е известен и като схема на Бернули (Bernouli). В такъв случай е в сила разложението:

$$E = A + \bar{A} \quad (1)$$

Означаваме вероятността за "благоприятен изход" (нека за такъв приемем настъпването на събитието A) с:

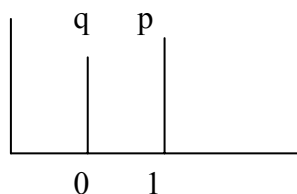
$$P(A) = p, \quad P(\bar{A}) = 1 - p = q \quad (2)$$

Резултатът от един такъв експеримент може да се изрази чрез една случайна величина \bar{x} , която приема стойности 1 (0) в зависимост от това, дали е настъпило събитието A (\bar{A}). Статистическите характеристики на случайната величина \bar{x} са:

$$E\{\bar{x}\} = \sum xp(x) = 1 \cdot p + 0 \cdot q = p \quad (3)$$

$$\sigma^2\{\bar{x}\} = \sum [x - p]^2 p(x) = (1 - p)^2 \cdot p + (0 - p)^2 \cdot q = pq^2 + p^2q = pq \quad (4)$$

Разпределението на тази случайна величина изглежда така:



Това е почти всичко, което може да се каже за тази най-проста случайна величина. Тя е интересна не толкова сама по себе си, колкото с това, че тя участва в изграждането на всички важни дискретни разпределения.

6.2. Биномно разпределение

Това е вероятностното разпределение на случайна величина, описваща сумарния резултат от n -кратното повторение на експеримента с два изхода. Тя се представя като следната сума:

$$\bar{x} = \sum_{i=1}^n \bar{x}_i \quad (1)$$

където \bar{x}_i е случайна величина, описваща единичен експеримент с два изхода. Нейното разпределение ни е вече известно.

Важно допълнение тук е, че отделните единични експерименти се предполагат *независими* помежду си. Тогава вероятността за k на брой благоприятни изходи (т.е. събитието A в предишната схема) при n -кратно повторение ще бъде:

$$W_k^n = \binom{n}{k} p^k q^{n-k} \quad (2)$$

Отделните членове имат такъв смисъл: $p^k q^{n-k}$ е вероятността точно k на брой реализации на единичния експеримент да дадат благоприятен изход, а $\binom{n}{k} = \frac{n!}{k!(n-k)!} = C_n^k$ е броят на всички възможни начини, по които може да се получат k благоприятни изхода при n на брой експерименти от този тип.

W_k^n е вероятностното разпределение на случайната величина (1). То се нарича биномно разпределение. При него k е *аргумент* (това е броят на благоприятните изходи, който е независимата променлива), а n и p са *параметри* (това са съответно броят реализации на единичния експеримент с два изхода и вероятността за благоприятен изход при единична реализация). Ясно е, че: $0 \leq k \leq n$, $0 \leq p \leq 1$, а $0 \leq n$ е цяло число.

Очевидно:

$$\sum_{k=0}^n W_k^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1^n = 1 \quad (3)$$

така че вероятността (6) е правилно нормирана.

Математическото очакване на биномното разпределение е:

$$\begin{aligned} E\{\bar{x}\} &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k} \\ &= np \sum_{k=0}^n \frac{(n-1)!}{(k-1)![n-1-(k-1)]!} p^{k-1} q^{n-1-(k-1)} = np \sum_{l=0}^{n-1} \binom{n-1}{l} p^l q^{n-1-l} = np \end{aligned} \quad (4)$$

Този резултат може да се пресметне и по-лесно - като използваме вече известния ни израз за математическото очакване на единичния експеримент с два изхода (6.1.3) и правилото за математическото очакване на сума от независими случайни величини:

$$E\{\bar{x}\} = \sum_{i=1}^n p = np \quad (5)$$

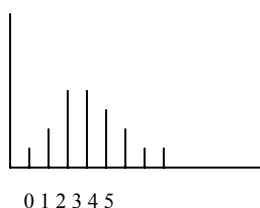
Съответно, за дисперсията на биномното разпределение имаме:

$$\sigma^2\{\bar{x}\} = \sum_{k=0}^n (k - np)^2 W_k^n \quad (6)$$

Пресмятанията по (6) са подобни на тези по (4), но са по-сложни. Ние можем обаче да използваме отново формулата за дисперсията на сума от независими случайни величини (ковариациите са нули). Така получаваме:

$$\sigma^2\{\bar{x}\} = \sum_{k=0}^n \sigma^2(x_k) = npq \quad (7)$$

Ето как изглежда биномното разпределение:



То е дефинирано само за цели неотрицателни стойности на аргумента k , има максимум при $k \approx np$ и е несиметрично (с коефициент на асиметрия $\gamma_1 > 0$).

6.3. Сходимост по вероятност. Закон за големите числа

От свойствата на биномното разпределение може да се направи едно важно наблюдение за връзката между честота и вероятност. Именно, да дефинираме честотата на събитието A (благоприятния изход) в n Бернулиеви опита като:

$$\bar{v} = \frac{\bar{x}}{n}. \quad (1)$$

Величината \bar{v} е случайна величина с биномно разпределение. Нейното математическо очакване е:

$$E\{\bar{v}\} = \frac{np}{n} = p, \quad (2)$$

а дисперсията ѝ е:

$$\sigma^2(\bar{v}) = \sigma^2\left(\frac{\bar{x}}{n}\right) = \frac{1}{n^2} \sigma^2(\bar{x}) = \frac{npq}{n^2} = \frac{p(1-p)}{n}. \quad (3)$$

Какво се вижда оттук?

- Първо, че математическото очакване на честотата съвпада с вероятността (която можем да предполагаме, че е неизвестна в един реален експеримент);
- Второ - с нарастването на n дисперсията на честотата намалява: $\lim_{n \rightarrow \infty} \sigma^2(\bar{v}) = 0$. Тъй като

в (3) $p(1-p) \leq \frac{1}{4}$, то при фиксирано n следва, че $\sigma(\bar{v}) \leq \frac{1}{2\sqrt{n}}$. Следователно, с

нарастването на n честотата \bar{v} на дадено събитие все повече се доближава до вероятността за това събитие.

Тук се сблъскваме с едно понятие, специфично за стохастиката. Става дума за това, че доближаването на честотата \bar{v} на благоприятния изход до вероятността p с нарастването на броя Бернулиеви опити не е в смисъла на обикновената сходимост, известна от математичния анализ, при която (припомняме си дефиницията за обичайната сходимост) - една редица от числа $\{a_n\}$ се нарича сходяща с граница a , ако за всяко $\varepsilon > 0$ съществува n такава, че за всяко $k > n$ е в сила: $|a_k - a| < \varepsilon$.

В разглеждания по-горе случай става дума не за сходимостта на числова редица към число, а за сходимостта на една редица от случайни величини към число. Това е така наречената сходимост по вероятност. Ето нейната дефиниция: една редица от случайни величини $\{\bar{a}_n\}$ се нарича сходяща по вероятност към числото a , ако за всяко число $\varepsilon > 0$ и всяко $\delta > 0$ съществува такава n , че за всяко $k > n$ е в сила: $P(|a_k - a| > \delta) < \varepsilon$. Това означава, че не самата разлика между членовете на редицата и нейната граница клони към нула, а вероятността тази разлика да не намалява неограничено, клони към нула.

Ясно е, че наличието на сходимост в обичайния смисъл е достатъчно условие за сходимостта по вероятност. Обратното обаче не е вярно. Поради това обикновената сходимост се нарича още силна сходимост, а сходимостта по вероятност се нарича слаба сходимост.

Слабата сходимост на честотата към вероятността популярно се нарича закон за големите числа. Всъщност обаче става дума само за един важен частен случай на закона за големите числа, който (частен случай) е известен като теорема на Бернули. По-общата формулировка на закона за големите числа е следната: Ако $\{X_n\}$ е редица от независими случайни величини с едно и също вероятностно разпределение, което има математическо очакване M и крайна дисперсия σ^2 , тогава редицата от средните им аритметични $\left\{ \tilde{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} \right\}$ клони по вероятност към M при $n \rightarrow \infty$.

Можем да заключим, че законът за големите числа дава сериозна подкрепа за обосновката на честотното определение на вероятността. С негова помощ става ясно, че граница на редицата от честоти съществува винаги и че тази граница съвпада с вероятността.

6.4. Обобщения и гранични случаи на биномното разпределение

Ще споменем някои обобщения на биномното разпределение:

Полиномното разпределение е непосредствено обобщение на биномното за експерименти с краен брой изходи (повече от два). Предполагайки, че е в сила разложението:

$$\Omega = A_1 + A_2 + \dots + A_l \quad (1)$$

където събитията A_j са несъвместими, и ако

$$P(A_j) = p_j, \quad \sum_{j=1}^l p_j = 1 \quad (2)$$

то вероятността при n опита всяко от събитията A_j да се случи k_j пъти ($j = 1, \dots, l$) се дава с:

$$W_{(k_1, k_2, \dots, k_l)}^n = \frac{n!}{\prod_{j=1}^l k_j!} \prod_{j=1}^l p_j^{k_j} \quad (3)$$

(На това разпределение се подчиняват например игрите със зар).

По-сложно е т.нар. хипергеометрично разпределение, което описва вероятността от n извадени топки (бели и черни) k да са бели и $l = n - k$ - черни, ако K и L са всичките бели и черни топки и $N = K + L$ е общият брой топки в урната. Това е изваждане без повторение. В този случай последователните опити не са независими, тъй като те зависят от предисторията (така например цветът на последната топка е еднозначно известен от предишните резултати).

Тази зависимост може да се усили, ако всеки път, когато извадим топка от определен цвят, добавим в урната фиксиран брой топки от същия вид. Това е разпределението на Пойа, което има отношение към възникването на епидемии - тогава появяването на един случай увеличава вероятността за подобни заболявания.

Нека да се върнем към биномното разпределение и да разгледаме граничните случаи. Те са два:

- при фиксирана вероятност p за благоприятен изход при единичния експеримент, ако броят на неговите повторения n нараства неограничено, биномното разпределение клони към нормално разпределение (разпределение на Гаус (Gauss));
- ако с нарастването на n вероятността за благоприятен изход p намалява така, че $np = const$, то биномното разпределение при $n \rightarrow \infty$ клони към разпределението на Поасон (Poisson).

6.5. Разпределение на Поасон

6.5.1. Разпределението на Поасон като граничен случай на биномното разпределение

Нека означим:

$$\lambda = np \quad (1)$$

Предполагаме, че λ е фиксирано. Тогава:

$$W_k^n = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \quad (2)$$

$$= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n(n-1)\dots(n-k+1)}{n^k} \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k}$$

Първият множител в (2) е $const$, при $n \rightarrow \infty$ третият и четвъртият клонят всеки към единица, а вторият клони към експоненциална функция. Следователно:

$$W_k^n \xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda} \quad (3)$$

6.5.2. Физичен модел на разпределението на Поасон

Подходът, който описваме тук, се основава на използването на един конкретен физичен модел и по тази причина изяснява по-добре свойствата на Поасоновото разпределение от гледна точка на физическия му смисъл.

Нека имаме радиоактивен източник с множество радиоактивни ядра, които се разпадат спонтанно. При това смятаме, че са в сила следните условия:

1. Отделните събития са независими във времето (т.е. наличието на разпадане в момента t не зависи от предисторията на процеса);
2. Вероятността за отделно събитие (разпадане) за *малък интервал* от време δt е пропорционална на дължината на интервала:

$$p(t, t + \delta t) = \mu \delta t + O(\delta t)^2 \quad (4)$$

3. Вероятността за повече от едно събитие (разпадане) в *малък интервал* от време δt е 0.

Ние търсим вероятността $P_k(t)$ в интервала време $(0, t)$ да се случат k на брой разпадания. За целта ще образуваме $P(t)$ и $P(t + \delta t)$ и ще получим диференциално уравнение за $P(t)$ при $\delta t \rightarrow 0$.

Първо търсим P_0 . За да има нула разпадания в интервала $(0, t + \delta t)$ трябва да няма разпадания както в интервала $(0, t)$, така и в интервала $(t, t + \delta t)$. Съгласно първото условие тези събития са независими, следователно:

$$P_0(t + \delta t) = P_0(t)(1 - \mu \delta t) \quad (5)$$

(напомняме, че $1 - \mu \delta t$ е вероятността за неразпадане в интервал с дължина δt).

Следователно:

$$\frac{P_0(t + \delta t) - P_0(t)}{\delta t} = -\mu P_0(t) \quad (6)$$

При $\delta t \rightarrow 0$:

$$\frac{dP_0}{dt} = -\mu P_0(t) \quad (7)$$

Общото решение на (7) е:

$$P_0(t) = Ae^{-\mu t} \quad (8)$$

където A е произволна константа.

За да се определи стойността на A , е необходимо едно начално условие. Логично е да предположим:

$$P_0(0) = 1 \quad (9)$$

т.е. вероятността за неразпадане в интервал с дължина нула е равна на тази на достоверното събитие. Тогава, окончателно:

$$P_0(t) = e^{-\mu t} \quad (10)$$

Това всъщност е известният закон за радиоактивното разпадане.

Нека сега $k = 1$, т.е. търсим вероятността за *точно едно* разпадане в интервала $(0, t + \delta t)$. Има две възможности: едно разпадане в $(0, t)$ и нула разпадания в $(t, t + \delta t)$ или обратното. Тези сложни събития са несъвместими, така че общата вероятност е сума от техните вероятности:

$$P_1(t + \delta t) = P_1(t)(1 - \mu\delta t) + P_0(t)\mu\delta t \quad (11)$$

Следователно:

$$\frac{P_1(t + \delta t) - P_1(t)}{\delta t} = -\mu P_1(t) + \mu e^{-\mu t} \quad (12)$$

При $\delta t \rightarrow 0$ се получава съответното диференциално уравнение, чието решение е:

$$P_1(t) = \mu t e^{-\mu t} \quad (13)$$

(проверете чрез заместване на (13) в (12)). Това е вероятността за точно едно разпадане в интервала $(0, t)$.

Горното разглеждане е напълно приложимо и за общия случай ($k \geq 1$):

$$P_k(t + \delta t) = P_k(t)(1 - \mu\delta t) + P_{k-1}(t)\mu\delta t \quad (14)$$

тъй като според третото предположение на модела повече от едно разпадане в интервала $(t, t + \delta t)$ не е възможно. Следователно:

$$\frac{P_k(t + \delta t) - P_k(t)}{\delta t} = -\mu P_k(t) + \mu P_{k-1}(t) \quad (15)$$

При $\delta t \rightarrow 0$, това диференциално уравнение има решение:

$$P_k(t) = \frac{(\mu t)^k}{k!} e^{-\mu t} \quad (16)$$

което е всъщност Поасоновото разпределение.

Разглеждайки тази картина статично, (т.е. като анализираме броя на наблюдаваните разпадания за даден фиксиран интервал от време), ние виждаме, че $\lambda = \mu t$ има смисъл на среден брой разпадания за дадения интервал, а P_k е вероятността да получим k на брой разпадания, ако средният им брой е λ . Оттук се вижда и физическият смисъл на параметъра μ ; това е *средната скорост на разпадане* на радиоактивните ядра.

6.5.3. Свойства на Поасоновото разпределение

Разпределението на Поасон (3) е дефинирано за всички цели неотрицателни стойности на своя аргумент k . То зависи от един параметър λ , който е реално неотрицателно число.

Нормировка:

$$\sum_{k=0}^{\infty} P_k(\lambda) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} \equiv 1 \quad (17)$$

откъдето се вижда, че разпределението, дефинирано чрез (3), е нормирано коректно.

Средна стойност:

$$E\{P(\lambda)\} = e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \quad (18)$$

Дисперсия:

$$\sigma^2\{P\} = e^{-\lambda} \sum_{k=0}^{\infty} (k - \lambda)^2 \frac{\lambda^k}{k!} \quad (19)$$

Развиваме израза за квадрата в скобите. Вторият и третият членове са от познат вид.

Първият е:

$$\begin{aligned} e^{-\lambda} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} (j+1) \frac{\lambda^j}{j!} = \lambda(\lambda+1) \end{aligned} \quad (20)$$

Следователно:

$$\sigma^2\{P(\lambda)\} = \lambda^2 + \lambda - 2\lambda^2 + \lambda^2 = \lambda \quad (21)$$

Забележка: Дисперсията може да се пресметне и като се използва резултата (6.2.7), за дисперсията на биномното разпределение, тъй като Поасоновото разпределение е всъщност един негов частен случай. Именно:

$$\sigma^2\{P(\lambda)\} = npq = \lambda(1 - \lambda/n) \xrightarrow{n \rightarrow \infty} \lambda \quad (22)$$

Асиметрия:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\sum_{k=0}^{\infty} (k - \lambda)^3 \frac{\lambda^k}{k!} e^{-\lambda}}{\lambda^{3/2}} = \frac{1}{\sqrt{\lambda}} \quad (23)$$

От написаното се вижда, че:

- параметърът λ действително има смисъл на средна стойност, което оправдава названието му при втория подход (среден брой разпадания);
- дисперсията на Поасоновото разпределение е равна на средната му стойност;
- асиметрията на Поасоновото разпределение е винаги положителна; с увеличаването на λ тя намалява, т.е. разпределението става по-симетрично.

Поасоновото разпределение се получава от биномното при голям брой повторения на експеримент с малка вероятност за благоприятен изход; поради това то е известно още като *разпределение на редките събития*.

Радиоактивното разпадане е вероятно най-известният пример за явление, подчиняващо се на Поасоновото разпределение. Има обаче още много други такива явления:

- броят на научните открития, направени няколкократно при независими изследвания;
- броят на звездите в определен ъглов диапазон на небесния свод,
- и др.

7. Непрекъснати разпределения

7.1. Равномерно разпределение

Това е най-простият пример за непрекъснато разпределение. Едно статистическо разпределение се нарича равномерно разпределение, ако вероятностната му плътност е константа в някакъв краен интервал и нула извън него:

$$f(x) = \begin{cases} c & a \leq x \leq b \\ 0 & x < a, \quad x > b \end{cases} \quad (1)$$

Нормировка:

$$\int_{-\infty}^{\infty} dx f(x) = 1 \Rightarrow c(b - a) = 1 \Rightarrow c = \frac{1}{b - a} \quad (2)$$

Функция на разпределение:

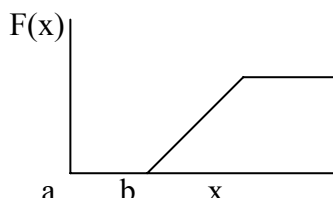
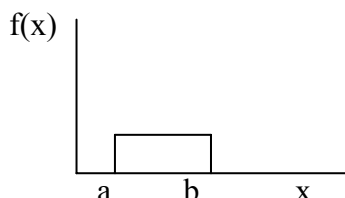
$$F(x) = \int_{-\infty}^x dx f(x) = \begin{cases} 0 & x < a \\ \frac{x - a}{b - a} & a \leq x \leq b \\ 1 & x > b \end{cases} \quad (3)$$

Математическо очакване:

$$E(\bar{x}) = \hat{x} = \int_{-\infty}^{\infty} dx \, x f(x) = \frac{a+b}{2} \quad (4)$$

Дисперсия:

$$\sigma^2(\bar{x}) = \frac{1}{12}(b-a)^2 \quad (5)$$



Равномерното разпределение само по себе си е от неголямо значение за практически приложения. То обаче е много удобно за пресмятания и играе важна роля при трансформациите на статистическите разпределения (например при моделирането със случайни числа).

7.2. Нормално разпределение

7.2.1. Лапласов модел на грешките

В произведението си "Философско есе върху теорията на вероятностите" през 1783г. Лаплас (P.S.Laplace) е разгледал въпроса за произхода на грешките в наблюденията. Той е предполагал, че:

- истинската стойност на измеряемата величина съществува обективно и има някаква точна и постоянна големина, равна на някаква константа m ;
- m не може да се измери точно, тъй като резултатите от измерването се изкривяват от голям брой (n) независими помежду си смущения, всяко от които е с големина ε ;
- всяко едно от смущенията с равна вероятност води до отклоняване на стойността на наблюдаемата величина на $\pm \varepsilon$;
- грешката на измерването се образува от сумирането на влиянието на отделните смущения.

Ясно е, че в този модел вероятностното разпределение на резултатите от измерванията се описва с биомно разпределение. Развитието на влиянието на грешките върху стойността на резултатите от наблюденията изглежда по следния начин:

бр.смущения	-3ε	-2ε	-1ε	m	+1ε	+2ε	+3ε
-------------	-----	-----	-----	---	-----	-----	-----

0				1			
1			1/2		1/2		
2		1/4		1/4+1/4		1/4	
3	1/8		1/8+1/4		1/4+1/8		1/8

В таблицата е дадена вероятността да се получи даден резултат от измерването след въздействието на определен брой смущения. В началната точка отсъствуват смущения и затова с вероятност 1 ние получаваме истинската стойност на измеряемата величина m . С появата на едно смущение с равна вероятност ($1/2$) се получава резултат $m - \varepsilon$ или $m + \varepsilon$. Същото става и при всяко следващо смущение, т.е. всеки ред се получава от предишния чрез сумиране на ефекта от предишните смущения и последното по схемата на Лаплас. Това е впрочем вероятността W_k^n за $p = q = 1/2$.

Сега, по-интересно е да видим какво става при неограничено нарастване на броя на смущенията: $n \rightarrow \infty$. Макар че в оригиналния модел на Лаплас $p = q = 1/2$, то резултатът, който следва, е валиден и за $p \neq q$ (т.е. за асиметрия в посоките на отклонение на резултата под действието на смущението).

Този резултат (известен като локална теорема на Моавр-Лаплас (Moivre-Laplace), която няма да доказваме) гласи, че за всички нормирани бернулиеви случайни величини \bar{x} от вида:

$$\bar{x} = \frac{k - np}{\sqrt{npq}} \quad (1)$$

където k е някое цяло число, при $n \rightarrow \infty$ е в сила:

$$P(x) \equiv \Phi_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (2)$$

По-общо, разпределението (2) има вероятностна плътност от вида:

$$\Phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_0)^2}{2\sigma^2}} \quad (3)$$

Математическото очакване на случайната величина с вероятностна плътност (3) е:

$$E(\bar{x}) = x_0 \quad (4)$$

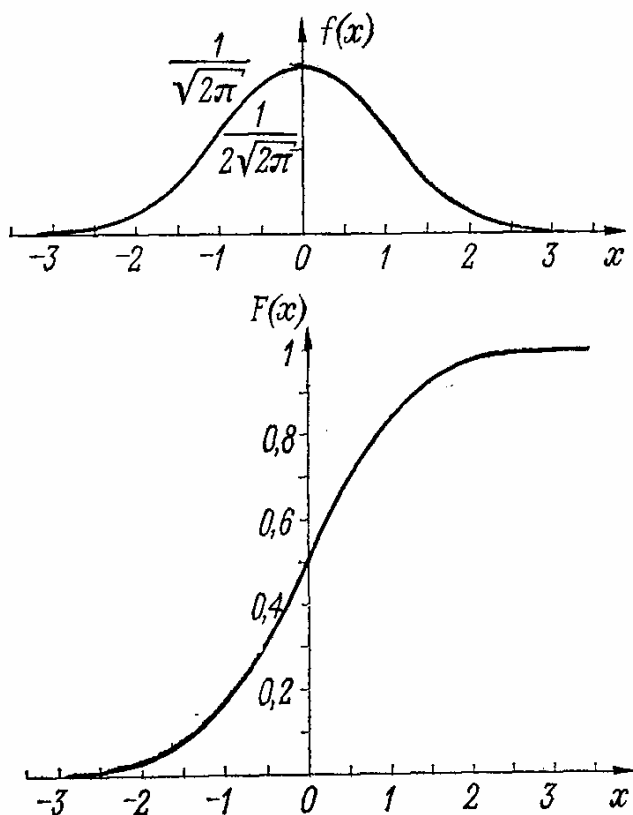
а дисперсията ѝ е:

$$\sigma^2(\bar{x}) = \sigma^2 \quad (5)$$

Това разпределение се нарича нормално, или Гаусово. То заема централно място в математическата статистика и по-специално в теорията на грешките. Разпределението,

описвано с вероятностната плътност (2), се нарича стандартно нормално разпределение. То има средна стойност нула и дисперсия 1.

7.2.2. Свойства на нормалното разпределение



Вероятностната плътност на нормалното разпределение (3) представлява камбановидна крива, симетрична спрямо точката x_0 . Тя има инфлексни точки при стойности на аргумента $x = \pm\sigma$, където втората производна става нула и характерът на наклона на кривата се сменя (от вдлъбнат в изпъкнал). Това разпределение зависи от два параметъра - средната стойност x_0 и дисперсията σ^2 . Използва се и означението:

$$N(x_0, \sigma^2) \equiv \Phi(x) \quad (6)$$

Функцията на разпределение на стандартното нормално разпределение е:

$$\Psi_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad (7)$$

Тази функция е свързана с известната специална функция $erf(x)$ (функция на грешките, англ. error function):

$$erf(x) \stackrel{def}{=} \frac{2}{\sqrt{\pi}} \int_0^x dt e^{-t^2} \quad (8)$$

Именно:

$$\Psi_0(x) = \frac{1}{2} + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)$$

Функцията (7) се нарича *интеграл на вероятностите (probabilities integral)*.

Вероятността една нормално разпределена случайна величина \bar{x} да се отличава от средната си стойност с по-малко от $n\sigma$ е:

$$\begin{aligned} P(|\bar{x} - x_0| < 1\sigma) &= 0.682 \\ P(|\bar{x} - x_0| < 2\sigma) &= 0.954 \\ P(|\bar{x} - x_0| < 3\sigma) &= 0.998 \end{aligned} \tag{9}$$

От тези съотношения се вижда, че нормално разпределената случайна величина е групирана в близка околност на средната си стойност (с радиус от порядъка на σ). По-нататък ще видим, че в много случаи в добро приближение може да се счита, че *неопределеностите (грешките) при измерванията са разпределени нормално около средната стойност x_0* . В такъв случай стандартното отклонение σ може да бъде разглеждано като мярка за големината на случайните грешки при измерванията.

Едно друго важно свойство на нормалното разпределение е, че сумата на две независими случайни величини с Гаусови разпределения е също нормално разпределена със средна стойност $\hat{x} = x_1 + x_2$ и с дисперсия $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Това свойство обаче е само частен случай от един много по-общ резултат, известен като *централна гранична теорема*. Самата централна гранична теорема (ЦГТ) има различни формулировки с различна степен на общност. Може да се установи например, че ако $\{\bar{x}_i, i=1..n\}$ е множество от независими случайни величини, еднакво разпределени със средни стойности x_0 и дисперсии σ^2 , то сумата $\frac{1}{n} \sum_{i=1}^n \bar{x}_i$ при $n \rightarrow \infty$ се подчинява на Гаусово разпределение с параметри $N(x_0, \frac{\sigma^2}{n})$. Забележете, че в тази формулировка никъде не става дума за конкретния вид на вероятностното разпределение на всяка от величините $\{\bar{x}_i\}$ (т.е. тяхното разпределение може да е произволно; достатъчно е само средните стойности и дисперсиите на всички $\{\bar{x}_i\}$ да са съответно равни, и теоремата пак ще е вярна!). Може да се покаже, че сумите от случайни величини се подчиняват на нормално разпределение дори и при по-слаби предположения от тези, в частност когато не всичките $\{\bar{x}_i\}$ имат еднакви разпределения. Ние няма да се занимаваме с най-общата формулировка на ЦГТ, но е важно да знаем, че нормалното разпределение е добро приближение за статистическото разпределение на измерваните случайни величини в много широк клас случаи.

7.2.3. Нормалното разпределение и грешките при експеримента: модел на Хершел

С помощта на ЦГТ Лапласовият модел за грешките може да бъде изтълкуван като модел за сумарното въздействие на малки индивидуални ефекти върху резултата от измерването. Първоначалното предположение за еднакви амплитуди на всичките смущения сега може да се отхвърли и въпреки това се получава нормално разпределение на грешките в резултата (вж. предишния раздел). В стремеж да се изяснят границите на приложимост на нормалното разпределение като модел за описание на грешките при физичните измервания е интересно да разгледаме модела на астронома Хершел (W.Herschel), който е извел нормалното разпределение при много общи предположения.

Хершел разглежда падането от точка O върху една равнина на малки топчета. Поради случайни смущения не всички топчета падат върху една и съща точка от равнината. Ние се интересуваме при това от вероятностната плътност на разпределение на координатите на топчетата, т.е. от вероятността за попадане на едно топче върху елементарната площ dA . Този елемент може да се разглежда както в декартови, така и в полярни координати. Предполагаме, че са взети мерки да няма систематични отклонения от центъра; тогава всички ъгли на отклонение са равновероятни (т.е. имаме ъглова симетрия). В такъв случай вероятностната плътност в полярни координати ще зависи само от радиус-вектора. Ние я означаваме с $h(r)$.

Предполагаме още, че попадането на топчето в интервала $(x, x+dx)$ е независимо от попадането му в интервала $(y, y+dy)$. Поради вече предположената љглова симетрия, вероятностната плътност по x и по y ще има еднакъв вид, т.е. ще е в сила:

$$g(x, y) = f(x)f(y) \quad (1)$$

Или, вероятността за попадане в елементарната площ dA ще бъде:

$$h(r)dA = f(x)f(y)dA \quad (2)$$

Логаритмуваме двете страни:

$$\ln h(r) = \ln f(x) + \ln f(y) \quad (3)$$

Преминавайки и в дясната страна на (3) към полярни координати, имаме:

$$\ln h(r) = \ln f(r \cos \theta) + \ln f(r \sin \theta) \quad (4)$$

Сега диференцираме двете страни спрямо полярния љгъл θ :

$$0 = -r \sin \theta \frac{f'(r \cos \theta)}{f(r \cos \theta)} + r \cos \theta \frac{f'(r \sin \theta)}{f(r \sin \theta)} \quad (5)$$

Връщайки се отново в декартови координати, получаваме:

$$0 = -y \frac{f'(x)}{f(x)} + x \frac{f'(y)}{f(y)} \quad (6)$$

Прегрупираме и стигаме до израза:

$$\frac{f'(x)}{xf(x)} = \frac{f'(y)}{yf(y)} \quad (7)$$

Поради независимостта на x и y това функционално равенство трябва да бъде изпълнено за всяко x и за всяко y . Това е възможно само ако всяка от двете му страни поотделно е равна на константа:

$$\frac{f'(x)}{xf(x)} = \text{const} = c \quad (8)$$

Следователно (интегрирайки (8) по x):

$$\begin{aligned} \ln[f(x)] &= \frac{1}{2} cx^2 + \text{const} \\ f(x) &= k \exp\left(\frac{1}{2} cx^2\right) \end{aligned} \quad (9)$$

Тъй като вероятността трябва да е ограничена при $x \rightarrow \infty$, то очевидно трябва да е изпълнено: $c < 0$. Означаваме: $c = -1/\sigma^2$. Тогава получаваме последователно:

$$\begin{aligned} f(x) &= k \exp\left(-\frac{1}{2} \frac{x^2}{\sigma^2}\right) \\ h(r) &= k^2 \exp\left(-\frac{1}{2} \frac{(x^2 + y^2)}{\sigma^2}\right) \\ h(r) &= k^2 \exp\left(-\frac{1}{2} \frac{r^2}{\sigma^2}\right) \end{aligned} \quad (10)$$

Какво се оказа? При много общи и много естествени предположения за независимост и симетрия ние получихме нормално разпределение за отклоненията. Това е демонстрация на правдоподобността при използване на Гаусовото разпределение като модел на грешките при измерванията. Заключение за наличието на Гаусово разпределение на грешките е от решаваща важност в много случаи, особено в теорията на метода на най-малките квадрати.

Въпреки това ние трябва да помним, че *нормалното разпределение не е универсален природен закон*. Причините за появата на неопределености при измерванията са твърде многообразни, за да могат да се обхванат с една формула всевъзможните им проявления. Симетрията и независимостта, макар и да са много общи качества, не винаги са изпълнени в конкретните експерименти. Затова трябва да се проверява нормалността на разпределението на отклоненията винаги, когато това е възможно, още *преди* да са извършени (понякога твърде обширните) пресмятания, които са валидни обаче само за случая на Гаусово разпределение.

7.3. Разпределение χ^2

Тук ще дадем кратки сведения (без доказателства) за още едно важно непрекъснато разпределение.

Нека $\{\bar{x}_i, i = 1, \dots, n\}$ са независими случайни величини, които имат стандартно нормално разпределение $N(0,1)$. Тогава сумата:

$$\bar{u} = \sum_{i=1}^n \bar{x}_i^2 \quad (1)$$

е случайна величина, чиято вероятностна плътност има вида:

$$f(u) = \frac{1}{\Gamma(\lambda)2^\lambda} u^{\lambda-1} e^{-\frac{u}{2}} \quad (2)$$

където:

$$\lambda = \frac{n}{2} \quad (3)$$

$$\Gamma(x+1) = \int_0^\infty dt t^x e^{-t}$$

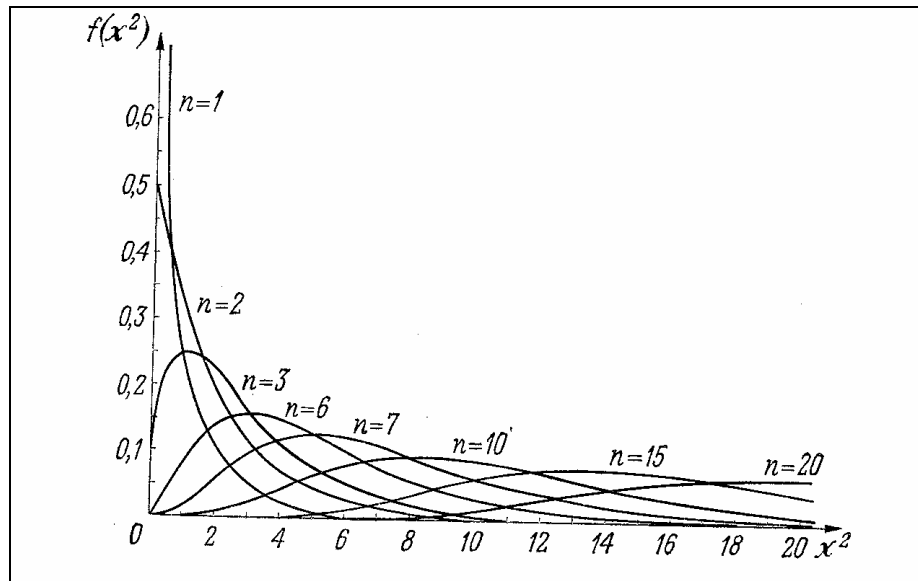
$\Gamma(x)$ е гама-функцията на Ойлер (L.Euler). Числото n , което е единственият параметър на разпределението (2), се нарича брой на степените на свобода. Разпределението (2) се нарича разпределение хи-квадрат (χ^2).

Едно важно свойство на разпределението χ^2 е следното:

сумата на две независими случайни величини с разпределения χ^2 съответно с n_1 и n_2 степени на свобода се подчинява на разпределение χ^2 с n_1+n_2 степени на свобода.

Математическото очакване и дисперсията на χ^2 разпределението с плътност (2) са съответно:

$$\begin{aligned} E\{\bar{\chi}^2\} &= 2\lambda = n \\ \sigma^2\{\bar{\chi}^2\} &= 4\lambda = 2n \end{aligned} \quad (4)$$



На фигурата е показана вероятностната плътност на разпределението χ^2 за различен брой степени на свобода. Вижда се, че с увеличаването на n разпределението става все по-симетрично (и все по-широко).

Разпределението χ^2 намира приложение поне в два много важни случая:

- при проверката на хипотезата за вероятностната плътност на случайна величина, измерена в някакъв експеримент (χ^2 критерий за съгласие);
- при проверката на резултатите от оценките по метода на най-малките квадрати.

Ако величините $\{\bar{x}_i, i = 1, \dots, n\}$ имат разпределение не $N(0,1)$, а $N(x_0, \sigma^2)$, то величината:

$$\bar{\chi}^2 = \sum_{i=1}^n \frac{(\bar{x}_i - x_0)^2}{\sigma^2} \quad (5)$$

се подчинява на χ^2 разпределение с n степени на свобода.

V. Оценяване на параметри чрез извадки

Досега ние разглеждахме общия вид на статистическите разпределения, а също така и редица техни специфични свойства. Ние обаче не се интересувахме от това, как тези свойства се проявяват при отделните експерименти. Фактически, вероятностната плътност в реалните ситуации зависи от някои параметри (напр. средната стойност при Поасоновото разпределение, средната стойност и дисперсията при Гаусовото разпределение и пр.). Числените стойности на тези параметри като правило са неизвестни в един конкретен експеримент. Следователно, ние не знаем *точно* вероятностно разпределение и сме принудени да го апроксимираме с някакво честотно разпределение, каквото може да се получи експериментално (трябва да си спомним, че конкретните експерименти са свързани винаги с краен брой измервания). Съвкупността от експерименти, провеждани с тази цел, се нарича извадка.

8. Извадки

8.1. Случаен избор. Разпределение на извадките

Извадката се прави измежду множество елементи. Размерът на това множество е обикновено безкраен. То всъщност се състои от всевъзможните изходи от експериментите от даден вид. Такова множество се нарича популация (често се използва и терминът генерална съвкупност).

Ако е осъществена извадка от n елемента, казва се, че извадката е с обем n . Важно е да се знае, че обемът на извадката е винаги краен.

Нека разпределението на случайната величина \bar{x} в популацията се описва с вероятностната плътност $f(x)$. Ние се интересуваме от стойностите на \bar{x} , които се съдържат в елементите на извадката. Предполагаме, че сме извършили последователно l извадки с обем за всяка от тях n и при това сме получили следните резултати:

$$\begin{aligned} & x_1^1, x_2^1, \dots, x_n^1 \\ & \dots \\ & x_1^l, x_2^l, \dots, x_n^l \end{aligned} \tag{1}$$

Ние можем да разглеждаме всеки пореден елемент от извадката като случайна величина $\bar{x}_i, i = 1, \dots, n$. Тогава цялата извадка може да бъде разглеждана като една n -мерна случайна величина:

$$x^j = (x_1^j, x_2^j, \dots, x_n^j), \tag{2}$$

която е вектор в пространството на n -мерните извадки (напомниме, че случайната величина \bar{x} , която всъщност изучаваме и чиито всевъзможни реализации образуват популацията, е само една). Тази случайна величина има вероятностна плътност:

$$g(x) = g(x_1, x_2, \dots, x_n) \quad (3)$$

Ние казваме, че плътността $g(x)$ описва процеса на случаен избор тогава и само тогава, когато са изпълнени следните условия:

- случайните величини \bar{x}_i (компонентите на извадката) са независими, т.е.:

$$g(x) = g_1(x_1)g_2(x_2)\dots g_n(x_n); \quad (4)$$

- всяка вероятностна плътност в (4) е равна на вероятностната плътност на случайната величина \bar{x} в популацията:

$$g_1(x) = g_2(x) = \dots = g_n(x) = f(x) \quad (5)$$

Ако не е уговорено друго, по-нататък ще считаме, че извадките, които разглеждаме, са резултат именно от случаен избор.

Трябва да се отбележи, че в реалния процес на избор е много трудно да се осъществи неговата случайност, а пък още по-трудно е да се убедим, че сме я постигнали. Това означава да премахнем систематичните отклонения в процеса на измерване и да оставим да действуват само случайни фактори (и то едни и същи в целия процес на получаване на извадката). На по-обща език това означава да осигурим условия за *повторяемост на резултатите от експеримента*. Това е едно от главните умения на добрия експериментатор.

Сега да предположим, че все пак сме осигурили условията за случаен избор. Нека елементите на извадката да са подредени в редица по възходящите стойности на \bar{x} (съгласно правилата на случайния избор това не е ограничение на общността на разглеждане, тъй като отделните компоненти са независими), и нека n_x е броят на елементите, за които $\bar{x} < x$. Тогава функцията:

$$W_n(x) = n_x / n \quad (6)$$

може да се разглежда като емпирична функция на вероятностно разпределение на случайната величина \bar{x} . Тази функция е стъпаловидна и тя нараства с $1/n$ всеки път, когато \bar{x} стане равно на някоя от стойностите на така подредените елементи на извадката. Тази функция (6) се нарича разпределение на извадката. Очевидно, $W_n(x)$ е приближение към функцията на разпределение на популацията $F(x)$, към която се стреми при неограничено нарастване на n .

Всяка функция на елементите на извадката (2) е случайна величина. Такива случайни величини се наричат статистики. Най-известният пример за статистика е средното аритметично на извадката:

$$\tilde{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) \quad (7)$$

8.2. Оценки на параметри. Свойства на оценките

Една типична задача на анализа на данни е следната: Предполагаме, че е известен общият вид на вероятностната плътност на разпределението на популацията. От експеримента е получена една извадка, от която трябва да се получи числовата стойност на един (или повече) параметър/ри. Поради крайния обем на извадката резултатът не може да бъде точен. Така възниква задачата за оценка на параметри.

Тъй като оценяваната стойност на параметъра се получава с помощта на извадката, то очевидно това е някаква статистика от специфичен вид; такава статистика се нарича оценка: $S = S(x_1, \dots, x_n)$. Казва се още, че статистиката S служи за оценка на параметъра λ или пък че S е оценка за λ .

Една оценка се нарича неотместена, ако за извадка с произволен обем нейното математическо очакване е равно на оценявания параметър:

$$E\{S(x_1, \dots, x_n)\} = \lambda \quad (\forall n) \quad (8)$$

Една оценка се нарича състоятелна, ако тя се сходя по вероятност към стойността на оценявания параметър (т.е. при неограничено нарастване на обема на извадката точността на оценката също неограничено нараства). Или, за състоятелните оценки е в сила:

$$\lim_{n \rightarrow \infty} \sigma(S) = 0 \quad (9)$$

Накрая, не всички оценки са еднакво ефективни. За сравняване на ефективността на две оценки за един и същ параметър може да се използва съотношението:

$$\eta = \frac{\sigma^2(S_1)}{\sigma^2(S_2)} \quad (10)$$

По-ефективната оценка има при един и същ обем на извадката по-малка дисперсия. По-нататък ще видим, че понякога е възможно да се установи точна долна граница на дисперсията за всевъзможните оценки на даден параметър. Естествено, ако съществува оценка, за която се достига тази долна граница на дисперсията, то тя е за предпочитане пред останалите, тъй като е по-ефективна от всички тях.

8.3. Оценки за средната стойност и за дисперсията на вероятностното разпределение

Средната стойност и дисперсията на вероятностното разпределение на популацията са най-често оценяваните параметри при физичните експерименти. Ето защо ние искаме да построим подходящи статистики, които да се използват за оценки на тези параметри.

Разглеждаме извадка с обем n , получена чрез случаен избор от дадена популация на случайната величина \bar{x} , описваща се с вероятностна плътност $f(x)$. Ще установим, че средната аритметична стойност на извадката:

$$\tilde{x} = \frac{1}{n}(\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_n) \quad (1)$$

е неотместена и състоятелна оценка за математическото очакване на популацията.

Тъй като \tilde{x} всъщност е функция на случайни величини, тя самата е също случайна величина. Тя е и статистика, тъй като е функция на елементите на извадката. Нейното математическо очакване е:

$$E(\tilde{x}) = \frac{1}{n} E(\sum \bar{x}_i) = \frac{1}{n} n\hat{x} = \hat{x} \quad (2)$$

т.е. то е равно на математическото очакване на популацията (и то за всяко n); следователно, (1) представлява неотместена оценка за средното на популацията.

За да установим, че статистиката (1) е и състоятелна оценка, ние пресмятаме нейната дисперсия:

$$\begin{aligned} \sigma^2(\tilde{x}) &= E\{[\tilde{x} - E(\tilde{x})]^2\} = E\left\{\left[\frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_n}{n} - \frac{n\hat{x}}{n}\right]^2\right\} \\ &= \frac{1}{n^2} E\left\{\left[\sum_{i=1}^n (\bar{x}_i - \hat{x})\right]^2\right\} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n E\{(\bar{x}_i - \hat{x})(\bar{x}_j - \hat{x})\} \end{aligned} \quad (3)$$

Тъй като компонентите на извадката \bar{x}_i са взаимно независими (по определение това е случаен избор), то смесените членове в сумата от вида $E\{(\bar{x}_i - \hat{x})(\bar{x}_j - \hat{x})\}$ са всичките нули при $i \neq j$ (това са всъщност ковариациите). Оставащите диагонални елементи (n на брой) са от вида: $E\{(\bar{x}_i - \hat{x})(\bar{x}_i - \hat{x})\} = E\{(\bar{x}_i - \hat{x})^2\} = \sigma^2(\bar{x}_i)$ и те са всичките равни на дисперсията на популацията $\sigma^2(x)$ (поради свойствата на случайния избор). Оттук получаваме:

$$\sigma^2(\tilde{x}) = \frac{1}{n^2} n\sigma^2(x) = \frac{1}{n} \sigma^2(\bar{x}) \quad (4)$$

Тук \bar{x} е кой да е елемент от извадката (те всичките имат еднакво разпределение и следователно, равни дисперсии). От (4) се вижда, че $\lim_{n \rightarrow \infty} \sigma^2(\tilde{x}) = 0$, с което се установява състоятелността на оценката (1) за математическото очакване на популацията \hat{x} . Ще отбележим още, че стойността на всяко отделно измерване (т.е. всяка случайно избрана стойност \bar{x}_i) е също така неотместена оценка за \hat{x} . Тя обаче не е състоятелна.

Сега да потърсим оценка за дисперсията на популацията. Изхождайки от дефиницията на дисперсията като централен момент от втори ред, определяме (засега) дисперсията на извадката като средно аритметично на средноквадратичните отклонения от средното на извадката:

$$s'^2 = \frac{1}{n} \sum_{i=1}^n (\bar{x}_i - \tilde{x})^2 \quad (5)$$

Сега ние искаме да проверим дали (5) е неотместена и състоятелна оценка за дисперсията на самата популация.

Пресмятаме първо математическото очакване на тази оценка:

$$\begin{aligned} E(s'^2) &= \frac{1}{n} E \left\{ \sum_{i=1}^n [\bar{x}_i - \tilde{x}]^2 \right\} = \frac{1}{n} E \left\{ \sum_{i=1}^n [\bar{x}_i - \hat{x} - (\tilde{x} - \hat{x})]^2 \right\} \\ &= \frac{1}{n} E \left\{ \sum_{i=1}^n [\bar{x}_i - \hat{x}]^2 \right\} - \frac{2}{n} E \left\{ \sum_{i=1}^n (\bar{x}_i - \hat{x})(\tilde{x} - \hat{x}) \right\} + \frac{1}{n} E \left\{ \sum_{i=1}^n [\tilde{x} - \hat{x}]^2 \right\} \\ &= \frac{1}{n} n \sigma^2(\bar{x}) - \frac{2}{n} E \left\{ \sum_{i=1}^n (\bar{x}_i - \hat{x}) \frac{1}{n} \sum_{j=1}^n [\bar{x}_j - \hat{x}] \right\} + \frac{1}{n} n \left[\frac{\sigma^2(\bar{x})}{n} \right] \end{aligned} \quad (6)$$

Тези преобразования се правят по аналогия с извършеното в предишната точка. Във втория член остават ненулеви само диагоналните членове ($i = j$), тъй като отново недиагоналните са нули (ковариации на независими поради правилата на случайния избор величини).

Окончателно:

$$E(s'^2) = \sigma^2(\bar{x}) - \frac{2}{n^2} n \sigma^2(\bar{x}) + \frac{\sigma^2(\bar{x})}{n} = \sigma^2(\bar{x}) \left(\frac{n-1}{n} \right) \quad (7)$$

Вижда се, че $E(s'^2) \neq \sigma^2(\bar{x})$, т.е. оценката (5) *не е неотместена* оценка за дисперсията на популацията (нейното отместване впрочем клони към нула при $n \rightarrow \infty$; такива оценки се наричат асимптотично неотместени). Същевременно става ясно, че следната оценка:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{x}_i - \tilde{x})^2 \quad (8)$$

е неотместена оценка за дисперсията на популацията.

Може да се установи (по аналогичен начин), че (8) е и състоятелна оценка за дисперсията на популацията.

Множителят $(n-1)^{-1}$ в дефиницията (8) изглежда донякъде странен. За да се установи неговият смисъл, си представяме, че $n=1$. Тогава средното аритметично на извадката е равно на стойността на единствения ѝ елемент, и според първоначалната дефиниция за дисперсията (5) тя би станала 0 (т.е. безкрайно точно измерване?), което очевидно не може да е вярно. Според неотместената оценка (8) в случая на извадка с обем единица дисперсията на популацията остава неопределена ($0/0$), т.е. тя не може да бъде оценена. Това наистина е така. Причината е, че информацията, съдържаща се в стойността на измерването \bar{x}_1 вече е използвана за оценка на \hat{x} - средната стойност на разпределението на популацията.

Друга гледна точка за обяснението на този множител е, че при определянето на оценката (5), респ. (8), елементите $\{\bar{x}_1, \dots, \bar{x}_n\}$ не са вече всичките независими – между тях се е появила една връзка с определянето на средното им аритметично чрез (1). По такъв начин, ефективният брой на елементите в извадката се намалява с един. В такъв случай се казва, че броят на степените на свобода на оценката s^2 е равен на $n-1$.

Оценката (1) се нарича още емпирично средно, а (8) – емпирична дисперсия на извадката. Положителният квадратен корен от (8) се нарича емпирично средноквадратично отклонение (за извадката).

Може да се установи, че всяко допълнително уравнение (връзка) между $\{\bar{x}_i\}$ - елементите на извадката, намалява броя на степените ѝ на свобода с единица. Ние ще използваме този резултат по-нататък без доказателство.

VI. Метод на максималното правдоподобие

9. Функция на правдоподобие

Досега ние разглеждахме въпроса за оценка на параметри. Бяха посочени някои желателни свойства на оценките, но без да се каже как да се построят оценки с такива свойства в конкретните случаи. Оценки бяха получени само за такива (наистина много важни) величини като математическото очакване и дисперсията, но техният вид като че ли пада от небето. Сега ние ще разгледаме една по-обща задача:

Нека разглеждаме една n -мерна случайна величина: $x = (\bar{x}_1, \dots, \bar{x}_n)$; $\{\bar{x}_i, i = 1, \dots, n\}$ са нейните компоненти, n на брой, които са също така случайни величини. Популацията в този случай се състои от всевъзможните експерименти, при които се получава някаква конкретна, фиксирана стойност на x . (Например, x може да е цял спектър, т.е. последователност от еквиливантни измервания, при което i е номерът на съответния канал, а $\{\bar{x}_i, i = 1, \dots, n\}$ са случайни величини, описващи резултатите от измерванията в съответните канали).

Предполагаме още, че съвместното вероятностно разпределение на величините $\{\bar{x}_i, i = 1, \dots, n\}$ напълно се определя от стойностите на параметрите $\lambda = (\lambda_1, \dots, \lambda_p)$, p на брой. Вероятностната плътност на разпределение на случайната величина x се записва така: $f(\bar{x}; \lambda)$.

Нека сме извършили единичен експеримент, подчиняващ се на правилата за случаен избор (все едно, сме измерили един спектър, но с n компоненти, който се описва с числата $\{\bar{x}_i^{(j)}, i = 1, \dots, n\}$). На този експеримент може да се съпостави едно число:

$$dP^{(j)} = f(x^{(j)}; \lambda) dx \quad (1)$$

Числото dP показва - след като резултатът от експеримента е вече известен - с каква вероятност този резултат (т.е. $x^{(j)}$ да попадне в многомерната област $(x, x + dx)$) е могъл да бъде очакван. То има характер на апостериорна вероятност. За извадка с обем N (независими елементи), тази вероятност се определя като произведение:

$$dP = \prod_{j=1}^N f(x^{(j)}; \lambda) dx \quad (2)$$

при това dP е функция на параметрите $\{\lambda\}$. Съществуват ситуации, в които е известно, че популацията може да се опише с един от два възможни и различни набора от параметри $\{\lambda_1\}$ и $\{\lambda_2\}$. Образуваме отношението:

$$Q = \frac{\prod_{j=1}^N f(x^{(j)}; \lambda_1)}{\prod_{j=1}^N f(x^{(j)}; \lambda_2)} \quad (3)$$

Резултатът от един конкретен (случаен) избор (измерване) може да се изрази така: наборът от параметри $\{\lambda_1\}$ е Q пъти по-вероятен от набора $\{\lambda_2\}$. Изразът (3) се нарича отношение на правдоподобие, а функцията:

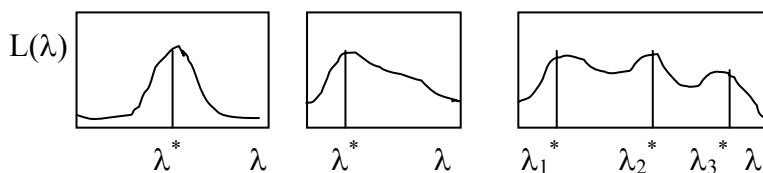
$$L = \prod_{j=1}^N f(x^{(j)}; \lambda) \quad (4)$$

се нарича функция на правдоподобие.

Много важно е да се разбере разликата между така определената функция L и обикновената (априорна) вероятност. Именно, при априорната вероятност (*a priori* = преди експеримента) ние предполагаме, че имаме някакви зададени, фиксирани стойности на параметрите $\{\lambda\}$ и като знаем вероятностната плътност $f(x; \lambda)$, ние можем да определим (да предскажем) вероятността да се получи някакъв резултат (например, че $\bar{x} = x^*$), т.е. тогава λ са параметри, а x - аргументи. При апостериорната вероятност резултатите от измерването са *вече* известни (това са векторите $x^{(j)}$, които са фиксирани), и ние можем по тези резултати от експеримента да оценяваме обратно величините λ , които са сега аргумент на функцията на правдоподобие L (а фиксираните $x^{(j)}$ в този случай са, напротив, параметри).

Разглеждайки конструкцията на понятието отношение на правдоподобие, лесно може да се стигне до заключението, че най-голямо доверие заслужават онези стойности на параметрите λ , при които се реализира максимум на функцията на правдоподобие L . В това собствено се състои и принципът на максималното правдоподобие (англ. *maximum likelihood principle*).

На фигурата са показани някои най-характерни възможни профили на функцията на правдоподобие (за един параметър).



В първия случай ние сме в най-благоприятна ситуация (функцията на правдоподобие е симетрична камбановидна крива). Несъмнено, стойността λ^* , съответстваща на максимума

на $L(\lambda)$, може да се приеме като най-добра оценка на параметъра λ . Функцията на правдоподобие задава едно разпределение (тя е пропорционална на вероятностната плътност при съответната нормировка), чието средно-квадратично отклонение е мярка за неопределеността на оценката λ^* за параметъра λ .

Във втория случай (на унимодално асиметрично разпределение) най-вероятната стойност λ^* все още е най-добра оценка за параметъра λ . Дисперсията обаче, въпреки че запазва своя математически смисъл, не е вече така удовлетворителна оценка за описание на поведението на функцията около нейния максимум. В този случай е по-добре заедно с λ^* да се пресмятат и повисшите моменти или дори цялата съвместна вероятностна плътност L .

Може накрая да се случи и така, че $L(\lambda)$ да има няколко локални максимума. В този случай обикновено се отдава предпочитание на най-големия максимум на функцията на правдоподобие. Но пресмятанията, свързани с намирането на глобалния максимум, могат да бъдат много сложни (и продължителни), особено при многопараметрични задачи (да си спомним, че известните теореми от анализа се отнасят за локални, а не за глобални екстремуми). Вероятността за появата на множество локални максимума на $L(\lambda)$ корелира с възможността за получаване на нееднозначно решение на физическата задача, от която всъщност сме тръгнали. Това може да се случи най-често при:

- малък брой наблюдения;
- лоша статистическа точност на измерванията;
- недобре определени начални предположения за вида на зависимостта $f(x; \lambda)$, която е често само приблизително известна.

В такива случаи е необходимо, очевидно, крайно внимание при интерпретацията на резултатите.

10. Метод на максималното правдоподобие

За определянето на положението на максимума на $L(\lambda)$ могат да бъдат използвани стандартните методи на математическия анализ: да се диференцира спрямо λ , да се приравнят производните на нула и да се решат възникващите при това уравнения (или системи уравнения). Диференцирането обаче на произведения е трудоемко и затова се предпочита следният подход:

$$l = \ln L = \sum_{j=1}^N \ln f(x^{(j)}; \lambda) \quad (5)$$

се въвежда като логаритмична функция на правдоподобие. Очевидно l и L имат максимуми при едни и същи стойности на λ , така че е достатъчно да диференцираме $l(\lambda)$, което е много по-лесно.

И така, трябва да се реши системата:

$$\frac{\partial l}{\partial \lambda_i} = \sum_{j=1}^N \frac{\partial}{\partial \lambda_i} \ln f(x^{(j)}; \lambda) = \sum_{j=1}^N \frac{\frac{\partial f(x^{(j)}; \lambda)}{\partial \lambda_i}}{f(x^{(j)}; \lambda)} = 0, \quad i = 1, \dots, p \quad (6)$$

Тази система от (изобщо нелинейни) алгебрични уравнения се нарича система уравнения на правдоподобие.

По този начин, чрез принципа на максималното правдоподобие, задачата за оценка на неизвестни стойности на физични параметри се свежда до една задача от алгебрата - решаване на системи от уравнения относно неизвестните $\{\lambda_i, i = 1, \dots, p\}$.

Пример: Разглеждаме важния случай за оценка на средната стойност при многократни измервания. Тези измервания (компонентите на извадката) могат да имат различна точност (напр. да са получени чрез различни методи, с различна апаратура и пр.). Предполагаме, че грешките в измерванията са разпределени нормално, т.е. всяко измерване е извадка с обем 1 от популация със средна стойност λ , (една и съща за всички измервания) и (различни) дисперсии σ_j^2 . Тогава вероятностната плътност за $\bar{x}^{(j)}$ е:

$$f(x^{(j)}; \lambda) = \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{(x^{(j)} - \lambda)^2}{2\sigma_j^2}} \quad (7)$$

Нека сме направили N измервания. Съответната функция на правдоподобие е:

$$L = \prod_{j=1}^N \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{(x^{(j)} - \lambda)^2}{2\sigma_j^2}} \quad (8)$$

а логаритмичната функция:

$$l = -\sum_{j=1}^N \frac{(x^{(j)} - \lambda)^2}{2\sigma_j^2} + const \quad (9)$$

Уравнението на правдоподобие е едно и има вида:

$$\frac{dl}{d\lambda} = \sum_{j=1}^N \frac{(x^{(j)} - \lambda)}{\sigma_j^2} = 0 \quad (10)$$

Неговото решение е:

$$\lambda^* = \frac{\sum_{j=1}^N \frac{x^{(j)}}{\sigma_j^2}}{\sum_{j=1}^N \frac{1}{\sigma_j^2}} \quad (11)$$

Следователно, резултатът, следващ от принципа на максималното правдоподобие, може да се изкаже така: най-правдоподобната оценка за средната стойност на популацията при многократни измервания е средното от отделните измервания, взети с тегла, обратно пропорционални на техните дисперсии. Забележете още, че в този конкретен пример функцията на правдоподобие е произведение от Гаусови функции с общ център; следователно ние се намираме в най-благоприятната ситуация от нарисуваните по-горе профили (симетрична камбановидна функция).

Важно е да се отбележи, че методът на максималното правдоподобие е конструктивен метод – той дава възможност да се построяват алгоритми за оценка на произволни параметри във всяка конкретна ситуация.

Методът на максималното правдоподобие (ММП) има и други положителни страни. Както казахме, съществува връзка между отместването на една оценка и нейната дисперсия (т.нар. неравенство за информацията или неравенство на Крамер-Рао (Kramer-Rao)). То поставя долна граница на дисперсията на една оценка, основана на извадка с краен обем. Очевидно, оценките, при които се постига тази долна граница, са за предпочитане пред останалите, тъй като те са най-ефективни (наричат се още оценки с минимална дисперсия). Доказва се, че когато обемът на извадката N клони към безкрайност, оценките, получени чрез метода на максималното правдоподобие (ММП) имат следните (асимптотични) свойства:

- те са неотместени;
- имат минимална дисперсия;
- функцията на правдоподобие е асимптотично нормална.

Последното дава възможност да използваме средноквадратичното отклонение на ММП-оценката $\sigma(\lambda^*)$ като оценка за вероятността истинската стойност на оценявания параметър λ да

лежи в някакъв доверителен интервал, напр. $P\{\lambda \in [\lambda^* - 1\sigma(\lambda^*), \lambda^* + 1\sigma(\lambda^*)]\} = 0.68,$ и пр.,
 $P\{\lambda \in [\lambda^* - 2\sigma(\lambda^*), \lambda^* + 2\sigma(\lambda^*)]\} = 0.95$

което, разбира се, представлява количествена мярка за увереността, с която можем да използваме тези оценки.

VII. Проверка на статистически хипотези

11. Обща теория

При статистическия анализ на експериментални данни ние се стремим да оценяваме стойностите на параметри, които имат определен физически смисъл. Тези параметри не могат да се считат за съвършено неизвестни. Напротив, много често ние имаме конкретни очаквания за стойностите на тези параметри – въз основа на предварителни ориентировъчни експерименти, някакви модели или пък от теорията. Така че целта на измерванията, които провеждаме, е всъщност проверката на дадена *хипотеза*.

Статистическа хипотеза се нарича всяко твърдение относно статистическите свойства на изучаваните случайни величини.

В много случаи хипотезата, подлежаща на проверка, се състои от предположение за вероятностната плътност на разглежданата случайна величина. Една хипотеза се нарича *проста*, ако тя напълно характеризира вероятностната плътност, т.е. тя предполага напълно определени стойности за всички параметри, от които зависи разпределението на всички случайни величини в дадения експеримент. Хипотезата е *сложна*, ако общият функционален вид на вероятностното разпределение е известен, но точната стойност поне на един от неговите параметри не е определена.

Примери:

- проста хипотеза: $f(x)$ има стандартно нормално разпределение (тук всички параметри са определени – математическото очакване е 0, а дисперсията – 1);
- сложна хипотеза: $f(x)$ има нормално разпределение със средна стойност напр. 5.6 (тук дисперсията остава неуточнена).

Хипотезата, която се проверява в даден момент, се нарича *нулева*. Всяка друга хипотеза относно същия експеримент се нарича *алтернативна*.

Как се проверяват статистическите хипотези? При осъществяване на измерванията ние можем да получим оценки (чрез извадката) за стойността на един или друг параметър на вероятностното разпределение на популацията, който представлява интерес за нас. Тъй като всяка проста нулева хипотеза дава пълното вероятностно разпределение в пространството на извадките, то от нея може да се пресметне априорната вероятност за получаване на именно тази стойност на оценката, която сме определили въз основа на извадката. Така на всеки резултат от измерването ние можем да съпоставим едно число – вероятността той да се осъществи, *ако*

нулевата хипотеза е вярна. Ясно е, че ако тази вероятност е голяма, то и вероятността хипотезата ни да е правилна е също голяма, и обратно. Проблемът обаче се състои в това, че поради статистическия характер на измерванията са възможни по принцип всякакви резултати (с определени вероятности). Така че ние не можем да дадем прост отговор на простия въпрос: вярна ли е в крайна сметка проверяваната хипотеза или не?

Тъй като все пак е необходимо да вземем решение (което може да има много важни последици за по-нататъшните ни действия), ние постъпваме по следния начин:

- предполагаме, че нулевата хипотеза H_0 е вярна;
- предварително (преди експеримента) определяме едно (обикновено малко) число α , което се нарича ниво на значимост;
- построяваме една критична област S от условието: $P(\bar{x} \in S | H_0) = \alpha$. Това е вероятността резултатът да попадне в критичната област S при условие, че H_0 е вярна (забележете, че критичната област S не се определя еднозначно от числото α);
- извършваме експеримента и проверяваме дали резултатът от измерването x^* попада в критичната област;
- Ако $x^* \in S$, ние отхвърляме хипотезата H_0 за нивото на значимост α ;
- в противния случай ние не можем да изкажем противоположното твърдение, т.е. не можем да приемем, че хипотезата H_0 е вярна. Ние можем да кажем само, че H_0 е съвместима с резултатите от нашия експеримент (при ниво на значимост α).

Нивото на значимост α е мярка за *риска* нулевата хипотеза H_0 да е вярна, макар че сме я отхвърлили. Такава грешка се нарича грешка от първи род. Подобна грешка е неизбежна поради вероятностния характер на измерването (иначе трябва да положим $\alpha = 0$; тогава цялото пространство на извадките ще бъде извън критичната област и никакво решение относно валидността на нулевата хипотеза не може да бъде взето). Все пак ние имаме свободата да избираме нивото на значимост α . Този избор се прави в зависимост от евентуалните последици от възможната грешка от първи род в конкретния случай - колкото те са по-значителни, толкова по-малко трябва да бъде α . В практиката на научните изследвания обикновено се използват стойностите $\alpha = 0.05, 0.01, 0.001$ и съществуващите таблици на статистическите разпределения са съобразени с тези стойности.

Има още една възможност да приемем грешно решение, именно да отхвърлим някоя вярна алтернативна хипотеза поради това, че x^* не попада в критичната област. Такава грешка се

нарича грешка от втори род. Вероятността за грешка от втори род зависи от конкретната алтернативна хипотеза, която е погрешно отхвърлена.

Близо до ума е, че един критерий за проверка на хипотези ще бъде особено полезен, ако при зададена вероятност α за грешка от първи род критичната област бъде определена така, че грешката от втори род да е минимална. Критерии, които отговарят на условието за минимизиране на вероятността за отхвърляне на *дадена* вярна проста алтернативна хипотеза при зададено ниво на значимост α се наричат най-мощни критерии. Може да съществуват и равномерно най-мощни критерии; това са такива критерии, които са най-мощни спрямо *всяка възможна* алтернативна хипотеза.

Директната проверка за принадлежност на резултата от експеримента към критичната област е неудобна в многомерния случай, тъй като тогава критичната област може да има много сложна конфигурация (тя е област в многомерно пространство). Вместо това се построява някаква статистика T , която е функция на компонентите на извадката. Тогава проверката $x^* \in S$ се заменя с проверката: $T(x^*) \in U(x^*) \equiv T(S(x^*))$. Последното равенство се отнася до образа на критичната област чрез статистиката T . Тази проверка е много по-лесна, защото се свежда обикновено до проверката за принадлежност на *едно* реално число $T(x^*)$ към някакъв интервал. Такава статистика T се нарича статистика, на основата на която се строи критерият (за проверка на нулевата хипотеза). Разбира се, не всяка статистика T е подходяща за построяване на критерий за проверка.

Общата теория на критериите излиза извън рамките на този курс. Засега ще се ограничим с тези най-общии бележки и ще разгледаме няколко примера за най-често използвани конкретни критерии.

12. Критерий за принадлежност към стандартно нормално разпределение

Нулевата хипотеза е, че *измерваната случайна величина \bar{x} има стандартно нормално разпределение*:

$$H_0 : \bar{x} \in N(0,1) \quad (1)$$

Нека смятаме, че сме направили извадка от n елемента. Интуицията ни подсказва, че за статистика, лежаща в основата на критерия за проверка на хипотезата (1) е добре да вземем средното на извадката \tilde{x} . Ако H_0 е вярна, \tilde{x} трябва да има разпределение $N(0, 1/n)$.

Нека за конкретната извадка сме получили резултат $\tilde{x} = x^*$. Тъй като математическото очакване на разпределението $N(0, 1/n)$ е 0, то ясно е, че ако $|x^*| \gg 0$, то проверяваната хипотеза едва ли е вярна.

На практика постъпваме по следния начин:

предполагаме, че (1) е вярна и избираме ниво на значимост α . От свойствата на нормалното разпределение ние знаем вероятността за получаване на резултат, по-голям от някое фиксирано x' :

$$P(|\tilde{x}| > x' : H_0) = 2[1 - \psi_0(x'\sqrt{n})] \quad (2)$$

където:

$$\psi_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x du \exp(-u^2/2) \quad (3)$$

(2) е всъщност площта на областта $(-\infty, -x') \cup (x', +\infty)$. И така, ние определяме границите $\pm x'_\alpha$ на критичната област от уравнението:

$$\alpha = 2[1 - \psi_0(x'_\alpha\sqrt{n})] \quad (4)$$

Избраната от (4) критична област е симетрична; тя се състои от интервалите:

$$S: (-\infty, -x'_\alpha) \cup (x'_\alpha, \infty) \quad (5)$$

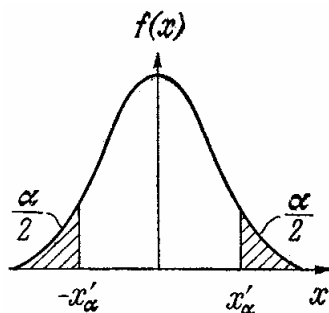
Добре е да имаме предвид, че за стандартните стойности на нивото на значимост α е в сила:

$$\begin{aligned} 0.05 &= 2[1 - \psi_0(1.96)] \\ 0.01 &= 2[1 - \psi_0(2.58)] \\ 0.001 &= 2[1 - \psi_0(3.29)] \end{aligned} \quad (6)$$

И така, ако например нашият резултат е $\tilde{x}^* \sqrt{n} = 2.1$, тогава заключението ни може да се изкаже така: *хипотезата (1) (че изходната случайна величина \bar{x} се описва със стандартно нормално разпределение) се отхвърля за ниво на значимост $\alpha = 0.05$. Ние обаче не можем да отхвърлим тази хипотеза (при този резултат), ако сме избрали ниво на значимост $\alpha = 0.01$ (тъй като тогава резултатът ни не попада в критичната област и рискът да отхвърлим хипотеза, която е вярна, е по-голям от планирания). Както вече казахме, не можем да изкажем и противоположното съждение. Тогава правилната формулировка на заключението ни би трябвало да бъде, че *резултатите от експеримента не противоречат на хипотезата за стандартно нормално разпределение при ниво на значимост $\alpha = 0.01$.**

За голямо съжаление, дори и резултатите ни да не могат да отхвърлят нулевата хипотеза даже и за много голямо ниво на значимост, това не ни позволява да кажем, че нашата хипотеза е вярна (вярна може да бъде всъщност някоя алтернативна хипотеза, която също не може да бъде

отхвърлена при тези стойности на измерването. В конкретния случай това може да бъде друго разпределение със средна стойност близка до 0). Такава ситуация е възможна например при измервания с ниска точност (с големи грешки), които са съвместими с много широко множество от различни хипотези. В крайната ситуация, един експеримент с напълно неопределен резултат е съвместим с всички възможни хипотези (но той не може да отдели нито една от тях спрямо останалите)!



В разглеждания случай критичната област изглежда така (защрихованата площ):

Всяка от защрихованите площи има големина $\alpha/2$. В случая имаме т.нар. двустранен критерий, тъй като критичната област (5) е двусвързана. Понякога ни интересуват едностранни критерии, когато и знакът на отклонението е важен (например когато контролираме температурата в някаква система и желаем тя да не надминава някоя стойност). В такъв случай ние проверяваме отклоненията само в едната посока и гледаме дали се изпълнява неравенството:

$$P(\bar{x} \geq x'_\alpha) < \alpha \quad (7)$$

Това неравенство определя друга критична област, която е само в дясната страна на Гаусовата крива и има площ α ; тя съответствува на едностранен критерий за проверка на хипотезата. Ако резултатът попадне в такава критична област, отхвърля се хипотезата, че измерваната величина има нормално разпределение с неотрицателна (а не нулева, вж по-горе) средна стойност и дисперсия 1 за ниво на значимост α .

13. F-критерий за равенство на дисперсиите

Проблемът за сравняване на дисперсии възниква обикновено във връзка с повторни измервания на една и съща случайна величина, да речем с различни уреди или по различни методики. Друг пример е влошаването на точността на показанията на някакъв уред с течение на времето поради различни причини. Може да възникне въпросът дали това се е случило. Това би могло да се установи, ако имаме две серии (извадки) от измервания, правени с този уред в различни моменти от време на едни и същи еталонни образци. Тогава средните стойности са равни (ако няма систематични грешки), но дисперсиите са, изобщо казано, различни.

И така, предполагаме, че имаме две извадки от съвкупности с нормално разпределение, с обеми съответно N_1 и N_2 . Нулевата хипотеза е: *дисперсията на двете серии измервания е еднаква*.

Статистиката, въз основа на която се строи критерият, се избира така:

$$Z = \frac{s_1^2}{s_2^2} \quad (1)$$

където s_1^2 и s_2^2 са емпиричните дисперсии на извадките N_1 и N_2 . Ясно е, че ако двете дисперсии са равни, Z ще е близко до 1. Известно е, че ако разпределението на популацията е нормално, статистиките:

$$\begin{aligned} X_1^2 &= \frac{(N_1 - 1)s_1^2}{\sigma_1^2} \in \chi^2(f_1) \quad f_1 = N_1 - 1 \\ X_2^2 &= \frac{(N_2 - 1)s_2^2}{\sigma_2^2} \in \chi^2(f_2) \quad f_2 = N_2 - 1 \end{aligned} \quad (2)$$

имат разпределение χ^2 със съответния брой степени на свобода. Ако нулевата хипотеза е вярна, то $\sigma_1^2 = \sigma_2^2$ (това е всъщност дисперсията на популацията, която в този случай е една и съща). Тогава:

$$Z = \frac{f_2 X_1^2}{f_1 X_2^2} \quad (3)$$

т.е. Z е отношение на две статистики с разпределение χ^2 .

Вероятностната плътност на разпределението χ^2 е:

$$f(\chi^2) = \frac{1}{\Gamma\left(\frac{f}{2}\right) 2^{\frac{f}{2}}} (\chi^2)^{\frac{1}{2}(f-2)} e^{-\frac{\chi^2}{2}} \quad (4)$$

Сега ние търсим вероятността:

$$V(Q) = P\left(\frac{X_1^2}{X_2^2} < Q\right) \quad (5)$$

Тя ще се получи с интегриране в триъгълната област:

$$V(Q) = \int_{x>0} \int_{y>0} \int_{x/y < Q} dx dy f_{f_1}(x) f_{f_2}(y) \quad (6)$$

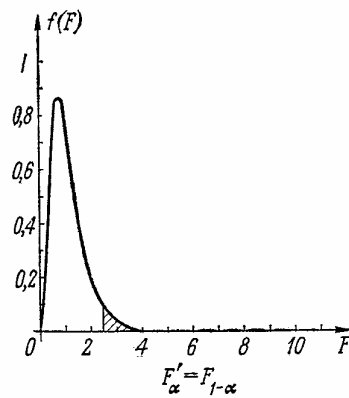
Този интеграл води до:

$$V(Q) = \int_0^Q dF f(F) \quad (7)$$

където за вероятностната плътност $f(F)$ имаме:

$$f(F) = \frac{\Gamma(f_1/2 + f_2/2)}{\Gamma(f_1/2)\Gamma(f_2/2)} \left(\frac{f_1}{f_2} F\right)^{f_1/2-1} \left(1 + \frac{f_1}{f_2} F\right)^{-(f_1/2+f_2/2)} \quad (8)$$

Това разпределение се нарича разпределение на Фишер (*Ernst Sigismund Fischer, австрийски математик*). То има два параметъра – двете степени на свобода f_1 и f_2 . То е донякъде подобно по форма на разпределението χ^2 - има ненулева плътност само за $F > 0$. При $f_2 > 2$ математическото му очакване е: $E(Z) = f_1/(f_2 - 2)$.



Обобщавайки, хипотезата за равенство на дисперсиите на две извадки се проверява така:

- избира се нивото на значимост α ;
- определяме границата на критичната област (заштрихованата на фигурата) от условието: $P(F > F'_\alpha) = \alpha$. Ясно е, че F'_α е квантилът на разпределението на Фишер на ниво $1 - \alpha$;
- извършваме измерванията; определяме емпиричните дисперсии s_1^2 , s_2^2 и пресмятаме Z от (1). За определеност, избираме като първа извадка тази с по-голяма емпирична дисперсия, така че винаги да е изпълнено: $Z = \frac{s_1^2}{s_2^2} \geq 1$.
- Ако $Z > F'_\alpha$, то ние се намираме в критичната област и нулевата хипотеза се отхвърля за нивото на значимост α . Тогава можем да заключим, че $\sigma_1^2 > \sigma_2^2$, тъй като критерият тук е едностранен.

14. Т-критерий за сравняване на средни

14.1. Т-критерий за сравняване на средна стойност с константа

Първо ще разгледаме случая на една случайна величина \bar{x} с нормално разпределение. Нулевата хипотеза е, че *математическото очакване на популацията е $\hat{x} = \lambda$, $\lambda = const.$* Правим извадка с обем N ; нека средното аритметично на извадката е \tilde{x} . Дисперсията на това средно е:

$$\sigma^2(\tilde{x}) = \sigma^2(\bar{x})/N \quad (1)$$

Съгласно централната гранична теорема, за достатъчно голяма извадка средното аритметично \tilde{x} клони към случайна величина с нормално разпределение $N(\hat{x}, \sigma^2)$. Следователно, съответната нормирана случайната величина ще има стандартно нормално разпределение:

$$\bar{y} = \frac{(\tilde{x} - \hat{x})}{\sigma(\tilde{x})} \in N(0,1) \quad (2)$$

Ако $\sigma(\tilde{x})$ би била известна, разглежданата задача би се свела до познатата задача за проверка на хипотезата за принадлежност към стандартно нормално разпределение. Проблемът е, че обикновено $\sigma(\tilde{x})$ не е известна; вместо нея ние разполагаме само с една нейна оценка чрез извадката:

$$s^2(\bar{x}) = \frac{1}{N-1} \sum_{j=1}^N (\bar{x}_j - \tilde{x})^2 \quad (3)$$

Съответната оценка за $\sigma(\tilde{x})$ ще бъде:

$$s^2(\tilde{x}) = \frac{s^2(\bar{x})}{N} = \frac{1}{N(N-1)} \sum_{j=1}^N (\bar{x}_j - \tilde{x})^2 \quad (4)$$

(правете разлика между \bar{x} и \tilde{x} !). Сега, ние се интересуваме от разпределението на случайната величина:

$$\bar{t} = \frac{(\tilde{x} - \hat{x})}{s(\tilde{x})} = \frac{(\tilde{x} - \hat{x})\sqrt{N}}{s(\bar{x})} \quad (5)$$

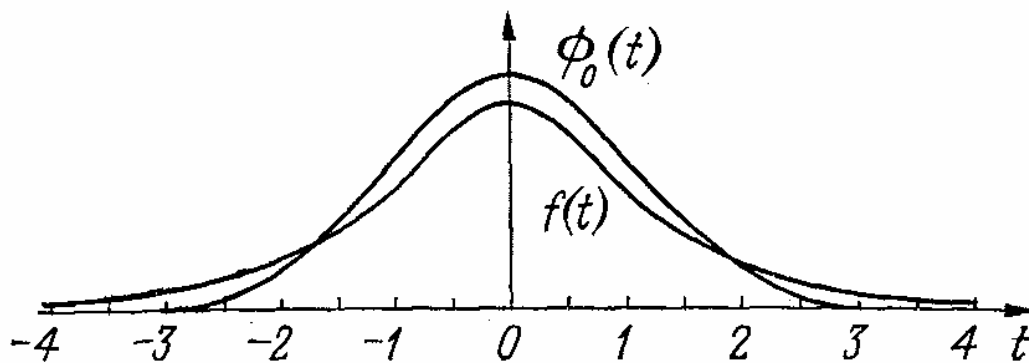
Очевидно, това разпределение ще се отличава от разпределението на \bar{y} (което е $N(0,1)$) поради замяната $\sigma(\tilde{x}) \rightarrow s(\tilde{x})$.

Без ограничение на общността можем временно да считаме, че $\hat{x} = 0$ (това води до проста трансляция по абсцисната ос на търсеното разпределение). Тъй като $(N-1)s^2(\bar{x}) \equiv fs^2(\bar{x}) \in \chi_f^2$ ($f = N-1$ е броят на степените на свобода), то ясно е, че величината \bar{t} представлява частно на две случайни величини: тази в числителя има нормално разпределение, а тази в знаменателя е

квадратен корен от величина с разпределение χ^2 . Вероятностната плътност на (5) при условие, че $\hat{x} = 0$ е:

$$f(t) = \frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\sqrt{\pi}\sqrt{f}} \left(1 + \frac{t^2}{f}\right)^{-\left(\frac{f+1}{2}\right)} \quad (6)$$

Това разпределение се нарича разпределение на Стюдънт (Student). (Заб. Student е псевдоним на William Sealey Gosset, английски химик и математик). То е симетрично (четна функция на t) и зависи от един параметър – броя на степените на свобода $f = N - 1$. При $f \rightarrow \infty$ то клони към нормално. На фигурата е показано разпределението на Стюдънт $f(t)$ в сравнение с нормалното разпределение $\Phi_0(t)$.



T-критерият (известен още като критерий на Стюдънт) за средната стойност на една случайна величина се прилага така:

- Приемаме за вярна нулевата хипотеза, че математическото очакване на популацията е $\hat{x} = \lambda$;
- Избираме ниво на значимост α ;
- намираме границите на критичната област $\pm t'_\alpha$ (разпределението е симетрично и тук имаме двустранен критерий, така че границите са две противоположни числа) от

$$\text{условието: } \int_{t'_\alpha}^{\infty} dt f(t) = \frac{\alpha}{2};$$

- правим измерванията, получаваме извадката и пресмятаме статистиката:

$$|t^*| = \frac{|\tilde{x} - \lambda| \sqrt{N}}{s(\bar{x})};$$

- ако $|t^*| > t'_\alpha$, то хипотезата $\hat{x} = \lambda$ се отхвърля за нивото на значимост α .

Този критерий може да се обобщи и да се използва за сравняването на средните стойности на две извадки.

14.2. Т-критерий за сравняване на средните на две извадки

Нека от две популации X и Y са направени две извадки с обеми N_1 и N_2 . Хипотезата, която ще проверяваме, е, че $\hat{x} = \hat{y}$.

Средноаритметичните по извадка имат приблизително нормално разпределение (съгласно централната гранична теорема). Техните дисперсии са:

$$\begin{aligned}\sigma^2(\tilde{x}) &= \sigma^2(\bar{x})/N_1 \\ \sigma^2(\tilde{y}) &= \sigma^2(\bar{y})/N_2\end{aligned}\quad (1)$$

Емпиричните оценки за дисперсиите на средноаритметичните по извадка са:

$$\begin{aligned}s^2(\tilde{x}) &= \frac{1}{N_1(N_1-1)} \sum_{j=1}^N (\bar{x}_j - \tilde{x})^2 \\ s^2(\tilde{y}) &= \frac{1}{N_2(N_2-1)} \sum_{j=1}^N (\bar{y}_j - \tilde{y})^2\end{aligned}\quad (2)$$

Знаем, че разликата:

$$\Delta = \tilde{x} - \tilde{y}\quad (3)$$

е също приблизително нормално разпределена и оценката за нейната дисперсия е:

$$s^2(\Delta) = s^2(\tilde{x}) + s^2(\tilde{y})\quad (4)$$

Ако нашата хипотеза е вярна ($\hat{x} = \hat{y}$), то, очевидно, $\hat{\Delta} = 0$ и ще е в сила:

$$\bar{t} = \frac{\Delta}{\sigma(\Delta)} \in N(0,1)\quad (5)$$

Тогава веднага можем да пресметнем вероятността нулевата хипотеза да е вярна, ако $\sigma(\Delta)$ е известно. Ние обаче не знаем "истинското" средноквадратично отклонение $\sigma(\Delta)$; вместо него разполагаме само с оценката $s^2(\Delta)$ (4).

Обикновено се предполага, че X и Y са от една и съща генерална съвкупност (това във всеки случай е в съгласие с предположението ни за верността на нулевата хипотеза). Тогава:

$$\sigma^2(\tilde{x}) = \sigma^2(\tilde{y})\quad (6)$$

Следователно:

$$s^2 = \frac{(N_1-1)s^2(\bar{x}) + (N_2-1)s^2(\bar{y})}{(N_1-1) + (N_2-1)}\quad (7)$$

което е претегленото средно на $s^2(\bar{x})$ и $s^2(\bar{y})$, е най-добрата оценка за дисперсията на популацията ((7) всъщност е равноправно сумиране на всички индивидуални дисперсии).

В такъв случай:

$$s^2(\Delta) = s^2(\tilde{x}) + s^2(\tilde{y}) = \frac{s^2}{N_1} + \frac{s^2}{N_2} = \frac{N_1 + N_2}{N_1 N_2} s^2. \quad (8)$$

Доказва се, че величината:

$$\bar{t} = \frac{|\Delta|}{s(\Delta)} \quad (9)$$

се подчинява на разпределението на Student с $f = N_1 - 1 + N_2 - 1$ степени на свобода.

Този критерий се прилага по-нататък напълно аналогично на критерия на Стюдънт от предния раздел.

15. Критерий за съгласие χ^2

15.1. Сравняване на емпиричната хистограма с дадена вероятностна плътност

Досега разгледахме критерии за проверки на хипотези, които се отнасят до определянето на един или повече параметри на популацията. Такива критерии се наричат параметрични критерии. Друг клас критерии сравнява директно вероятностната плътност на хипотетичното разпределение с това на резултатите от извадката. Това са така наречените критерии за съгласие. Те са непараметрични критерии, тъй като се отнасят не до някой конкретен параметър, а до вероятностната плътност като цяло. Тук ще разгледаме най-важния от тях – критерия за съгласие χ^2 .

Нека $f(x)$, $F(x)$ са съответно вероятностната плътност и функцията на разпределение на популацията на изследваната случайна величина \bar{x} .

Цялата област на изменение на случайната величина \bar{x} се разделя на r подинтервала: $\xi_1, \xi_2, \dots, \xi_r$. Чрез интегриране на $f(x)$ в тези интервали се получава вероятността за попадане на резултатите от наблюденията на \bar{x} в ξ_k :

$$p_k = P(\bar{x} \in \xi_k) = \int_{\xi_k} f(x) dx, \quad \sum_{k=1}^r p_k = 1 \quad (1)$$

Нека направим извадка с обем n и да я сортираме в интервалите $\xi_1, \xi_2, \dots, \xi_r$. Нека n_k е броят на елементите на извадката, попадащи в интервала ξ_k . Ясно е, че: $\sum_{k=1}^r n_k = n$.

Нулевата ни хипотеза тук е, че *разпределението на случайната величина \bar{x} е с вероятностна плътност $f(x)$* . Тогава трябва да очакваме да е изпълнено приблизителното равенство:

$$n_k \cong np_k, \quad k=1, \dots, r \quad (2)$$

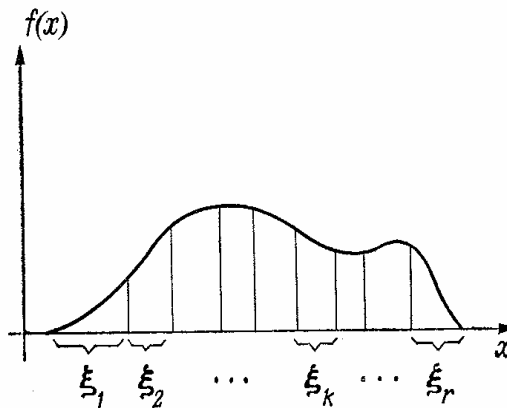
Като мярка за отклоненията построяваме статистиката:

$$X^2 = \sum_{k=1}^r \frac{(n_k - np_k)^2}{np_k} \quad (3)$$

Статистиката X^2 се подчинява (асимптотично) на разпределение χ^2 с $r-1$ степени на свобода. Това твърдение няма да доказваме тук. Само ще отбележим, че намаляването на броя на степените на свобода с единица се дължи на ограничението:

$$\sum_{k=1}^r n_k - n \sum_{k=1}^r p_k = 0 \quad (4)$$

Ние вече знаем, че такива ограничения намаляват броя на степените на свобода.



Прилагане на критерия:

- Избира се ниво на значимост α ;
- Правят се n наблюдения: x_1, x_2, \dots, x_n на случайната величина \bar{x} . Разбиването на дефиниционната област на \bar{x} на подинтервали $\{\xi\}$ се прави така, че:
 - броят им r да е достатъчно голям, така че да се описва добре профилът на разпределението;
 - броят на елементите във всяко ξ_k да е достатъчно голям, $n_k \gg 1$, за да бъде X^2 близо до асимптотичното си разпределение;
- пресмята се статистиката X^2 от (3) и се сравнява с границата на критичната област χ'_α , която се определя от условието: $\int_{\chi'_\alpha}^{\infty} f(\chi^2) d\chi^2 = \alpha$ (за $r-1$ степени на свобода). Ако

$X^2 > \chi'_\alpha$, хипотезата, че вероятностната плътност на \bar{x} е $f(x)$, се отхвърля за нивото на значимост α .

15.2. Сравняване на вероятностните плътности на две извадки

Критерият за съгласие може да се обобщи за сравняване на вероятностните плътности на две случайни величини \bar{x} и \bar{y} . В такъв случай нулевата ни хипотеза е, че *двете вероятности плътности са равни*: $f(x) = g(y)$. За проверка на тази хипотеза ние извършваме n_x наблюдения на величината \bar{x} и n_y наблюдения на \bar{y} . След това разпределяме двете серии наблюдения в едно и също множество интервали $\{\xi\}$. Статистиката (3) се модифицира така:

$$X^2 = \sum_{k=1}^r \frac{[n(x)_k - n(y)_k]^2}{n(x)_k + n(y)_k} \quad (5)$$

Величината X^2 се подчинява (асимптотично) на разпределение χ^2 с r степени на свобода.

15.3. Проверка на сложни хипотези

Казаното в предишния раздел е валидно при проста хипотеза, т.е. когато $f(x)$ е напълно известна. Може обаче да е известен общият вид на $f(x)$, но стойностите на някои параметри да не са известни, т.е. ние трябва да проверим една сложна хипотеза. Сега ние ще покажем как неизвестните стойности на параметрите на разпределението могат да се получат от наличните наблюдения, като използваме метода на максималното правдоподобие:

Тук отново разпределяме наблюденията n в интервалите $\xi_1, \xi_2, \dots, \xi_r$. Нека неизвестните параметри са $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$, $p < r$. Тогава вероятностите p_k са функции на λ : $p_k = p_k(\lambda)$. Функцията на правдоподобие за тази извадка е:

$$L(\lambda) = \frac{n!}{\prod_{k=1}^r n_k!} \prod_{k=1}^r [p_k(\lambda)]^{n_k} \quad (5)$$

Този израз се получава направо от съвместната вероятностна плътност за наблюдаване на n_k елемента в интервала ξ_k , $k = 1, \dots, r$ (това е всъщност известното полиномно разпределение).

Логаритмичната функция на правдоподобие е:

$$l(\lambda) = \ln(n!) - \sum_{k=1}^r \ln(n_k!) + \sum_{k=1}^r n_k \cdot \ln[p_k(\lambda)], \quad (5')$$

а системата уравнения на правдоподобие е:

$$\frac{\partial l}{\partial \lambda_i} = \sum_{k=1}^r \frac{n_k}{p_k} \frac{\partial p_k}{\partial \lambda_i} = 0, \quad i = 1, \dots, p \quad (6)$$

Решавайки (6), ние можем да определим най-правдоподобните оценки за стойностите на неизвестните параметри $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$.

Уравненията (6) налагат върху компонентите на X^2 още p на брой ограничения, които намаляват броя на степените на свобода. В такъв случай:

$$X^2 = \sum_{k=1}^r \frac{(n_k - np_k(\lambda))^2}{np_k(\lambda)} \underset{n \rightarrow \infty}{\in} \chi_{r-p-1}^2 \quad (7)$$

VIII. Метод на най-малките квадрати

16. Общи бележки

Методът на най-малките квадрати (МНК, least-squares) е създаден от Лъожандър (Legendre) и Гаус (1805-1809) първоначално въз основа на интуитивни съображения. В първоначалния си вид това правило е изглеждало така:

Резултатите от многократните наблюдения $\{y_j, j = 1, \dots, n\}$ на неизвестната величина x се разглеждат като сума от истинската величина x и някаква добавъчна грешка:

$$y_j = x + \varepsilon_j \quad j = 1, \dots, n \quad (1)$$

Правилото гласи, че *стойността на неизвестната величина x трябва да се определи от условието за минимум на сумата от квадратите на грешките*:

$$\sum_{j=1}^n \varepsilon_j^2 = \sum_{j=1}^n (y_j - x)^2 = \min_x \quad (2)$$

В много случаи това правило може да бъде получено от метода на максималното правдоподобие (ММП, който е развит доста по-късно), което вече служи за сериозна обосновка на МНК. Но дори и в случаите, когато МНК не може да се получи чрез ММП, МНК-резултатите притежават редица благоприятни свойства (в сравнение с други методи за оценка). Поради това МНК е най-често използваният от всички статистически методи за обработка на данни и оценка на параметри.

Трябва да се разграничават следните случаи на употреба на МНК:

- при преки наблюдения - когато ние директно измерваме интересуващата ни величина;
- при косвени наблюдения - когато наблюдаваните величини са свързани с оценяваните параметри не пряко, а чрез някакви (известни или предполагаеми) функционални зависимости. С други думи, наблюдаваните величини са някакви функции - линейни (ЛМНК, linear least-squares) или нелинейни (НМНК, non-linear least squares) на търсените параметри;
- накрая, говорим за наблюдения с ограничения - когато между неизвестните параметри съществуват (или се предполагат) съотношения от определен функционален тип (пак се разграничават случаите на линейни и нелинейни ограничения);
- по-нататък, ограниченията могат да бъдат строги (ограничения от тип равенства) (напр. ограничението сумата от измеряемите ъгли на един триъгълник да е 180^0) или пък

ограничения от тип неравенства (напр. площите на спектралните линии да са неотрицателни).

Ние ще разглеждаме последователно различните варианти на МНК, започвайки от простите.

17. Метод на най-малките квадрати при преки наблюдения

Този случай е най-прост, но ние го разглеждаме специално, за да въведем някои означения и да очертаем основната схема на МНК.

Нека са извършени n *независими* измервания на една (неизвестна) величина x . Измерените стойности $\{y_j, j = 1, \dots, n\}$ съдържат грешки от измерванията:

$$y_j = x + \varepsilon_j \quad j = 1, \dots, n \quad (1)$$

За грешките ε_j предполагаме, че са *случайни величини, разпределени нормално около нулата*:

$$\varepsilon_j \in N(0, \sigma_j^2) \quad j = 1, \dots, n \quad (2)$$

Това предположение се оправдава в много случаи от централната гранична теорема.

Вероятността в резултат на измерването да получим стойност в интервала $(y_j, y_j + dy)$ е:

$$f_j dy = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left[-\frac{(y_j - x)^2}{2\sigma_j^2}\right] dy \quad (3)$$

Сега можем да конструираме функцията на правдоподобие и логаритмичната функция на правдоподобие. Те са, съответно:

$$L = \prod_{j=1}^n \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left[-\frac{(y_j - x)^2}{2\sigma_j^2}\right] \quad (4)$$

$$l = \ln L = -\frac{1}{2} \sum_{j=1}^n \frac{(y_j - x)^2}{\sigma_j^2} + \text{const}(x) \quad (5)$$

Следователно, условието за максимум на $l = \ln L$ съвпада с условието:

$$M = \sum_{j=1}^n \frac{(y_j - x)^2}{\sigma_j^2} \equiv \sum_{j=1}^n \frac{\varepsilon_j^2}{\sigma_j^2} = \min_x \quad (6)$$

което е точно условието на МНК. Виждаме, че в този най-прост случай МНК се получава непосредствено от ММП.

В интерес на единството на по-нататъшните означения ще въведем величините:

$$g_j \stackrel{\text{def}}{=} \frac{1}{\sigma_j^2} \quad j = 1, \dots, n \quad (7)$$

Те се наричат тегла на измерванията $\{y_j, j = 1, \dots, n\}$. При това получаваме:

$$M = \sum_{j=1}^n g_j \varepsilon_j^2 \quad (8)$$

Вижда се, че теглото на всеки от членовете в сумата (8) е обратно пропорционално на дисперсията на съответното измерване; следователно, по-малко точните измервания участвуват с по-малко тегло при определянето на резултата x .

Самият резултат (т.е. оценката за стойността на неизвестната величина x), разбира се, може да се получи в този случай непосредствено чрез пресмятане на минимума на функционала M :

$$\frac{\partial M}{\partial x} = 2 \sum_{j=1}^n \frac{(x - y_j)}{\sigma_j^2} = 2 \sum_{j=1}^n g_j (x - y_j) = 0 \quad (9)$$

Оттук:

$$\tilde{x} = \frac{\sum_{j=1}^n g_j y_j}{\sum_{j=1}^n g_j} \quad (10)$$

това е претегленото средно аритметично на отделните измервания.

Ако всички дисперсии биха били равни $\{\sigma_j^2 = \sigma^2, j = 1, \dots, n\}$ (което би съответствало точно на условията за случаен избор от една и съща популация), тогава \tilde{x} би съвпаднало точно със средното аритметично на извадката. Изразът (10) е по-общ, доколкото той отчита евентуалната нееднаква точност на измерванията.

Оценката на търсената величина x , т.е. \tilde{x} , е случайна величина, тъй като тя се получава чрез извадката. На нас ни е необходима оценка и за нейните статистически характеристики, на първо място за нейната дисперсия, която зависи от свойствата на първичните наблюдения y , съдържащи случайните величини ε . Използвайки известните ни свойства за дисперсията на сума от независими случайни величини, за дисперсията на \tilde{x} получаваме:

$$\sigma^2(\tilde{x}) = \left(\sum_{j=1}^n \frac{1}{\sigma_j^2} \right)^{-1} = \left(\sum_{j=1}^n g_j \right)^{-1} \quad (11)$$

(при равни дисперсии на измерванията получаваме известния ни резултат $\sigma^2(\tilde{x}) = \frac{\sigma^2}{n}$).

Сега, тъй като грешките $\{\varepsilon_j\}$ са случайни величини, техните средни стойности $\{\tilde{\varepsilon}_j\}$ могат да бъдат оценени обратно чрез извадката $\{y_j\}$ в съответствие с вече полученото решение \tilde{x} :

$$\tilde{\varepsilon}_j = y_j - \tilde{x}, \quad j = 1, \dots, n \quad (12)$$

Оценките $\tilde{\varepsilon}_j$ са също случайни величини и имат разпределение $\tilde{\varepsilon}_j \in N(0, \sigma_j^2)$ (същото като това на ε_j). Следователно:

$$\frac{\tilde{\varepsilon}_j}{\sigma_j} \in N(0,1) \quad (13)$$

Но тогава ние знаем, че в точката на решението:

$$\tilde{M} = \sum_{j=1}^n \frac{\tilde{\varepsilon}_j^2}{\sigma_j^2} = \sum_{j=1}^n \frac{(y_j - \tilde{x})^2}{\sigma_j^2} = \sum_{j=1}^n g_j (y_j - \tilde{x})^2 \in \chi_{n-1}^2 \quad (14)$$

има χ^2 разпределение с $n - 1$ степени на свобода (една степен се отнема от замяната $x \rightarrow \tilde{x}$). Това свойство на величината M може да се използва за проверка на първоначалната ни хипотеза, а именно, че (припомняме):

- $\{y_j = x + \varepsilon_j, \quad j = 1, \dots, n\}$ са наблюдения на неизвестната величина x ;
- $\{\varepsilon_j \in N(0, \sigma_j^2) \quad j = 1, \dots, n\}$

И така, статистиката (14) може да се използва за проверка на горната хипотеза. Как ставаше това? Ако за избраното ниво на значимост α , $\tilde{M} = \sum_{j=1}^n g_j \tilde{\varepsilon}_j^2 > \chi_{1-\alpha}^2$, то ние се намираме в критичната област и трябва да отхвърлим хипотезата. Това означава или че (някои от) y_j не са наблюдения на неизвестната величина x , или че грешките $\{\varepsilon_j \notin N(0, \sigma_j^2) \quad j = 1, \dots, n\}$. Първата възможност може да се дължи на някакво грубо въздействие върху апаратурата (смущение, "изхвърчали" точки и др.), когато резултатите от измерването не може да се считат за наблюдения само на неизвестната величина x . Дори само едно $y_j \neq x + \varepsilon_j$ води до много голямо M , $M \gg \chi^2$ и по-нататъшната проверка става безсмислена. В такъв случай се постъпва така: взема се измерването с най-голямо отклонение ($\tilde{\varepsilon}_j = \max\{\tilde{\varepsilon}\}$) и му се приписва тегло нула, $g_j = 0$, т.е. това измерване се изключва от процеса на оценка. След това построяваме нова МНК-оценка за x , \tilde{x} . Ако за нея $M_{\tilde{x}}$ е значително по-малко, то трябва да изхвърлим измерването y_j и да продължим с останалите наблюдения.

Понякога обаче е възможно първият пункт от изказаната по-горе хипотеза да е в сила, но грешките от измерванията да не са разпределени нормално; в частност, наблюденията могат да

имат отместване спрямо нулата (напр. "пъзлене" на апаратурата с времето и пр.). Това вече говори за *систематични грешки* в измерванията. В такъв случай оправдано от статистическа гледна точка е (ако не могат да се "набедят" само 1-2 измервания като неточни) да не се правят опити за МНК-оценки и позитивни заключения за резултатите от измерванията, докато не бъдат извършени по-нататъшни експерименти.

18. Метод на най-малките квадрати при косвени наблюдения

18.1. Линеен случай

Сега предполагаме, че имаме няколко неизвестни величини: $x = (x_1, \dots, x_r)$. Обикновено ние не измерваме пряко самите величини, които ни интересуват, а някакви техни функции. Означаваме тези функции с η и предполагаме засега, че те са линейни:

$$\eta_j = a_{j0} + a_{j1}x_1 + \dots + a_{jr}x_r \quad j = 1, \dots, n \quad (1)$$

В матричен вид системата (1) се записва така:

$$\eta = a_0 + Ax \quad (2)$$

Тук η и a_0 са n -мерни вектори (стълбове), а A – $n \times r$ матрица.

Ние пак смятаме, че всяко измерване съдържа грешка и че $\{\varepsilon_j \in N(0, \sigma_j^2) \quad j = 1, \dots, n\}$.

Самите резултати от измерванията означаваме с y :

$$y = \eta(x) + \varepsilon \quad (3)$$

Предполагаме, че измерванията са *независими*. Следователно, ковариационната матрица на y (същата е и на ε , защото $\eta(x)$ не са случайни величини) е диагонална и има вида:

$$C_y = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ \dots & \dots & \dots \\ 0 & 0 & \sigma_n^2 \end{pmatrix} \quad (4)$$

Обратната матрица на C_y се нарича тегловна матрица на измерванията. В нашия случай (ур.(4)) тя се получава лесно:

$$G_y \stackrel{def}{=} C_y^{-1} = \begin{pmatrix} g_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & g_n \end{pmatrix} \quad (5)$$

$$g_j = 1 / \sigma_j^2, \quad j = 1, \dots, n$$

Заместваме (3) в (2) и получаваме:

$$y = a_0 + Ax + \varepsilon \quad (6)$$

Това е система линейни уравнения спрямо $x = (x_1, \dots, x_r)$. Ще търсим решение на тази система в духа на метода на максималното правдоподобие (тъй като тази система е преопределена при брой на неизвестните по-малък от този на измерванията, $r < n$, което е обичайният случай. Ако $r = n$, решението е единствено и няма какво повече да се прави). По предположение грешките ε_j (а от (3) това следва и за тези на y_j) са разпределени нормално, т.е.:

$$f(y_j) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left[-\frac{(y_j - \eta_j)^2}{2\sigma_j^2}\right] = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left[-\frac{\varepsilon_j^2}{2\sigma_j^2}\right] \quad j = 1, \dots, n \quad (7)$$

Функцията на правдоподобие, респ. логаритмичната функция на правдоподобие ще бъдат:

$$L = \prod_{j=1}^n f(y_j) = (2\pi)^{-\frac{n}{2}} \left(\prod_{j=1}^n \sigma_j^{-1} \right) \exp\left(-\frac{1}{2} \sum_{j=1}^n \frac{\varepsilon_j^2}{\sigma_j^2}\right) \quad (8)$$

$$l = \ln L = -\frac{n}{2} \ln(2\pi) - \sum_{j=1}^n \ln \sigma_j - \frac{1}{2} \sum_{j=1}^n \frac{\varepsilon_j^2}{\sigma_j^2}$$

Очевидно, ММП изисква да търсим максимум на l , който съвпада с минимума на M , където:

$$M = \sum_{j=1}^n \frac{\varepsilon_j^2}{\sigma_j^2} = \sum_{j=1}^n \frac{(y_j - a_{j0} - \sum_{k=1}^r a_{jk} x_k)^2}{\sigma_j^2} = \min_{\{x_k, k=1..r\}} \quad (9)$$

Или, в матричен вид M се записва така (този запис е валиден дори и когато измерванията y не са независими, т.е. тегловната матрица няма простия диагонален вид (5)):

$$M = \varepsilon^T G_y \varepsilon = (y - a_0 - Ax)^T G_y (y - a_0 - Ax) \quad (10)$$

Минимумът на M се постига, когато едновременно са изпълнени условията:

$$\frac{\partial M}{\partial x_i} = 0 \quad i = 1, \dots, r \quad (11)$$

Диференцирайки (10), се получава:

$$2A^T G_y (y - a_0 - Ax) = 0 \quad (12)$$

(последният резултат може да се провери чрез диференциране по компоненти). При $r \leq n$ това матрично уравнение (еквивалентно на система от линейни уравнения) може да се реши спрямо x . За целта то се записва във вида:

$$(A^T G_y A)x = -A^T G_y (y - a_0) \quad (13)$$

Тогава решението му е:

$$\tilde{x} = (A^T G_y A)^{-1} A^T G_y (y - a_0) \quad (14)$$

По такъв начин най-добрите ММП-оценки за неизвестните величини x се изразяват чрез измерванията y , тяхната ковариационна матрица $C_y = G_y^{-1}$ и известните коефициенти (A, a_0) на функциите $\eta(x)$.

За съгласуване на означенията с по-нататъшните раздели въвеждаме съкращението:

$$c \stackrel{def}{=} y - a_0 \quad (15)$$

което не зависи от x . Тогава решението (14) се записва така:

$$\tilde{x} = (A^T G_y A)^{-1} A^T G_y c \quad (16)$$

Тук случаят на преки измервания се получава като частен случай.

Сега ще се спрем на въпроса за влиянието на грешките от измерванията върху оценките на неизвестните параметри x . Тъй като (16) дава линейна връзка между неизвестните x и измерванията y , ние можем да използваме закона за разпространение на грешките за определяне на дисперсиите (и ковариациите) на x .

Нека $C_{\tilde{x}}$ е ковариационната матрица на x . Тогава имаме:

$$C_{\tilde{x}} = [(A^T G_y A)^{-1} A^T G_y] G_y^{-1} [(A^T G_y A)^{-1} A^T G_y]^T \quad (17)$$

За опростяване на израза вземаме предвид известното свойство на матриците $(AB)^T = B^T A^T$ и факта, че матриците G_y , G_y^{-1} , $(A^T G_y A)$ са симетрични (т.е. съвпадат със своите транспонирани). Резултатът е:

$$C_{\tilde{x}} = (A^T G_y A)^{-1} \quad (18)$$

И тъй, получихме проста форма за ковариационната матрица на най-добрите МНК-оценки \tilde{x} за неизвестните x . Квадратните корени от диагоналните елементи на (18) могат да се разглеждат като мерки за неопределеностите ("грешките") на неизвестните x , макар и те да не се измерват непосредствено в експеримента. Недиагоналните елементи са, разбира се, мярка за корелациите на x_i , x_j по двойки.

По-нататък, МНК-оценката за x , т.е. (16) може да се използва за определяне на най-добрата (в смисъл на МНК) оценка за вектора на грешките ε :

$$\tilde{\varepsilon} = c - A\tilde{x} \quad (19)$$

Тъй като грешките ε имат нормално разпределение, то статистиката M :

$$M = \tilde{\varepsilon}^T G_y \tilde{\varepsilon} \quad (20)$$

се подчинява на разпределение χ^2 с $n - r$ степени на свобода.

Пример: Полиномната апроксимация като линейна задача на МНК

Нека t е една независима променлива, а величините, които се измерват, са полиноми на t :

$$\eta = \sum_{i=0}^{r-1} x_i t^i \quad (1)$$

Нека сме извършили n измервания на η за $t = t_j$, $j = 1, \dots, n$. Това са:

$$y_j = \eta_j + \varepsilon_j \quad j = 1, \dots, n \quad (2)$$

Предполагаме, че освен y ние знаем и тяхната ковариационна матрица C_y (най-често предполагаме, че y са независими; тогава е достатъчно да познаваме $\{\sigma^2(y_j) \quad j = 1, \dots, n\}$).

Системата линейни уравнения за x ще има вида:

$$\eta = a_0 + Ax \quad (3)$$

Очевидно, сега ще имаме: $a_0 = 0$ и:

$$A = \begin{pmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{r-1} \\ 1 & t_2 & t_2^2 & \dots & t_2^{r-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & t_n & t_n^2 & \dots & t_n^{r-1} \end{pmatrix} \quad (4)$$

Това е задачата за прекарване на най-добрия полином от степен $r-1$ през n точки. Тази задача е най-популярното приложение на МНК.

18.2. Нелинеен случай

Досега предполагаме, че функционалните съотношения между измеряемите величини и неизвестните параметри са линейни. Сега ще се освободим от това ограничение и ще разгледаме общия случай.

И тъй, имаме:

$$\eta_j = \eta_j(x) = 0 \quad j = 1, \dots, n \quad (1)$$

Идеята е, както много често се прави, нелинейният случай да се сведе към линеен чрез развитие на нелинейните функции в ред на Тейлър (Taylor) и ограничаването на това разложение до линейни членове включително.

Нека $x_0 = (x_{10}, \dots, x_{r0})$ е някакво начално приближение към стойностите на неизвестните параметри x . Тогава:

$$\eta_j(x) = \eta_j(x_0) + \sum_{k=1}^r \left. \frac{\partial \eta_j}{\partial x_k} \right|_{x_k = x_{k0}} (x_k - x_{k0}) + O(x_k - x_{k0})^2, \quad j = 1, \dots, n \quad (2)$$

Означаваме:

$$\begin{aligned} \xi &= x - x_0 \\ a_{jk} &= \left. \frac{\partial \eta_j}{\partial x_k} \right|_{x_0} \\ A &= \begin{pmatrix} a_{11} & \dots & a_{1r} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nr} \end{pmatrix} \end{aligned} \quad (3)$$

Предполагаме, както и преди, че за фактическите резултати от измерванията y е в сила:

$$y_j = \eta_j + \varepsilon_j \quad j = 1, \dots, n \quad (4)$$

Тогава:

$$y = \eta(x_0) + A\xi + \varepsilon \quad (5)$$

Системата линейни уравнения (5) е напълно аналогична на линейния случай. Решението ѝ е:

$$\tilde{\xi} = (A^T G_y A)^{-1} A^T G_y c, \quad (6)$$

където сме означили:

$$c = y - \eta(x_0). \quad (7)$$

Това обаче сега е МНК-оценката за вектора на корекциите ξ . Съответно, ковариационната матрица на $\tilde{\xi}$ ще бъде:

$$C_{\tilde{\xi}} = (A^T G_y A)^{-1} \quad (9)$$

По същия начин (както за линейния случай) може да се получи оценка за уточнените грешки $\tilde{\varepsilon}$.

Тъй като началното приближение x_0 е фиксиран вектор, за вектора $x_1 = x_0 + \tilde{\xi}$ е в сила:

$$C_{x_1} = C_{\tilde{\xi}} = (A^T G_y A)^{-1} \quad (10)$$

Намереното решение x_1 на линейната задача е по-добро приближение към търсените стойности на неизвестните параметри x отколкото началното приближение x_0 . Затова е разумно да заменим x_0 с $x_1 = x_0 + \tilde{\xi}$, да развием $f(x, \eta)$ около x_1 и да повторим цялата процедура. Този итерационен процес може да се повтаря, докато поправките $\tilde{\xi}$ станат толкова малки, че уточненията за поредната стъпка да станат незначителни. В такъв случай можем да приемем стойностите, получени на последната итерация, за решение на нелинейния проблем.

Тук не е показано, че такава процедура е сходяща и води към единствено решение. Нещо повече, това в общия случай не може и да се докаже (засега?) Сходимостта трябва да се проверява следователно във всеки конкретен случай.

Ясно е, че функциите η не трябва да се отличават силно от линейни функции в околността на точките x_0, x_1, \dots . Или пък, алтернативно, ако η са силно нелинейни, началното приближение x_0 трябва да е достатъчно близо до решението \tilde{x} . Това е същността на изкуството за прилагане на МНК в нелинейния случай.

18.3. Алгоритъм на МНК за измервания без ограничения

Тук ще представим последователността на пресмятанията, необходими за получаването на решението на задачата за оценка на параметри по метода на най-малките квадрати за измервания без ограничения.

Дадени са:

- n измервания, $\{y_j, j = 1, \dots, n\}$;
- тяхната ковариационна матрица C_y ;
- n на брой измеряеми величини $\{\eta_j, j = 1, \dots, n\}$, които са известни функции на (евентуално) някакви независими променливи и на неизвестните параметри $\{x_k, k = 1, \dots, r\}$: $\eta = \eta(x)$.

Дадено е също едно начално приближение за стойностите на неизвестните параметри: $x = x_0$ (r -мерен вектор).

Търсим, разбира се, оценки за стойностите на неизвестните параметри $\{x_k, k = 1, \dots, r\}$ и за тяхната ковариационна матрица C_x .

Следва схемата на итерационния процес за решаване на МНК-проблема:

0. Означаваме номера на итерацията с s и полагаме $s = 1$. Пресмятаме тегловната матрица на измерванията: $G_y = C_y^{-1}$;
1. Пресмятаме матрицата $A = \left(\frac{\partial \eta}{\partial x} \right) \Big|_{x=x_{s-1}}$ и вектора $c = y - \eta(x_{s-1})$;
2. Пресмятаме $C_{x_s} = (A^T G_y A)^{-1}$ - ковариационната матрица на неизвестните в това приближение;
3. Определяме $\tilde{\xi} = C_{x_s} A^T G_y c$, $\tilde{\varepsilon} = c - A \tilde{\xi}$ - вектора на корекциите и оценките за грешките на измерванията;
4. $x_s = x_{s-1} + \tilde{\xi}$ - неизвестните параметри в това приближение;
5. $M_s = \tilde{\varepsilon}^T G_y \tilde{\varepsilon}$.

6. Ако е постигната сходимост (вж. по-долу), приключваме пресмятанята. Текущото решение се съдържа в x_s, C_{x_s} . В противен случай, увеличаваме номера на итерацията $s = s + 1$ и се връщаме към точка 1.

Критериите за постигане на сходимост на итерационната процедура могат да са следните:

- $|M_s - M_{s-1}| < \delta M_s$: (в две последователни итерации стойностите на статистиката M се различават незначително); това е белег, че минимумът на M е практически достигнат;
- $\|\tilde{\xi}\| < \delta$ (нормата на вектора на корекциите за дадена итерация е незначителна; тук

нормата е мярка за дължината на вектора, напр. Евклидовата норма е: $\|a\| = \sqrt{\sum_{k=1}^r a_k^2}$);

това означава, че корекциите вече са достатъчно малки.

Тези два критерия могат да се прилагат поотделно или в някаква комбинация помежду им.

Освен това трябва да се предвидят и резервни изходи:

- изход при достигане на някакъв предварително определен максимален брой итерации, независимо от това дали е постигната сходимост;
- изход при разходимост ($M_s > M_{s-1}$), тъй като, както отбелязахме, МНК-процедурата не е гарантирано сходяща; това се случва нерядко при лоши начални приближения.

19. Свойства на решенията, получени чрез метода на най-малките квадрати

Досега ние разглеждахме (и получавахме) МНК-решения на задачата за статистическа оценка на параметри само като приложение на метода на максималното правдоподобие за решаването на линейни и линеаризуеми задачи. Ние считахме, че грешките на наблюденията са разпределени нормално. При това положение е лесно да се построи функцията на правдоподобие, а оттам изискването на МНК за минимум на сумата на квадратите на отклоненията се получава непосредствено и съвпада с резултата от максимизацията на функцията на правдоподобие.

Ако се освободим от предположението (което е най-спорният пункт), че резултатите от наблюденията са разпределени нормално около истинските стойности, то все още е възможно да се използват съотношения от типа на МНК като “рецепти” за построяване на оценки на неизвестните параметри. Може да изглежда, че в такъв случай няма достатъчно основания от теоретичен характер за използването на МНК. Съществува обаче една теорема (на Гаус-

Марков), която установява, че дори и в такива случаи ($\varepsilon_j \notin N(0, \sigma_j^2)$) МНК-решенията притежават ред благоприятни свойства.

По-долу ще изброим всички свойства на решенията, получени чрез МНК в двата случая. Първо ще дадем едно определение:

Съвкупността от предположения:

$$\begin{aligned} y &= Ax + \varepsilon \\ \text{cov}(\varepsilon) &= \sigma^2 I \\ \det(A^T A) &\neq 0 \end{aligned} \quad (1)$$

се нарича общ линеен модел (тук I е единичната матрица от ранг n).

Сега ще сумираме *свойствата на МНК-оценката* \tilde{x} в рамките на общия линеен модел:

1. МНК-оценката \tilde{x} е неотместена;
2. Ако $\lim_{n \rightarrow \infty} \frac{1}{n} (A^T A)$ е неизродена матрица, то \tilde{x} е състоятелна оценка, а векторът $\sqrt{n}(\tilde{x} - x)$ е асимптотично нормален;
3. МНК-оценката \tilde{x} притежава минимална дисперсия в класа на всички възможни неотместени линейни (спрямо y) оценки за параметрите x ; при това векторите \tilde{x} и $\tilde{\varepsilon} = y - A\tilde{x}$ са некорелирани (теорема на Гаус-Марков);
4. Математическото очакване на величината $M = \tilde{\varepsilon}^T C_y^{-1} \tilde{\varepsilon}$ е $E(M) = n - r$ (т.е. е равно на броя на степените на свобода на задачата).

Ако в допълнение на горните предположения векторът на грешките ε е с нормално разпределение ($\varepsilon \in N(0, \sigma^2)$), то МНК-решението има следните *допълнителни свойства*:

5. МНК-оценката \tilde{x} съвпада с оценката по метода на максималното правдоподобие;
6. \tilde{x} има нормално разпределение с математическо очакване равно на x и с ковариационна матрица: $C_x = (A^T G_y A)^{-1}$;
7. Случайната величина (статистиката) $M = \tilde{\varepsilon}^T C_y^{-1} \tilde{\varepsilon}$ има χ^2 -разпределение с $n - r$ степени на свобода;
8. Оценките \tilde{x} и $s^2 = \frac{1}{n - r} \tilde{\varepsilon}^T \tilde{\varepsilon}$ са достатъчни оценки съответно за x и за σ^2 (т.е. за вектора на неизвестните параметри и за неговите дисперсии);
9. Векторите \tilde{x} и $\tilde{\varepsilon} = y - A\tilde{x}$ са независими.

Очевидно, условието 9) е по-силно от условието 3b); също така условието 7) е по-силно от условието 4) - 4) се отнася само до математическото очакване на M , а 7) - до цялата вероятностна плътност на M .

Когато грешките ε на измерванията са разпределени нормално, ние можем да използваме статистиката M за проверка на нулевата хипотеза по критерия χ^2 . Именно, ако за едно избрано ниво на значимост α се случи $M = \tilde{\varepsilon}^T C_y^{-1} \tilde{\varepsilon} > \chi_{1-\alpha}^2(n-r)$, то ние трябва да отхвърлим хипотезата, от която сме тръгнали (за даденото ниво на значимост). Освен от неизбежната възможност за грешка от I-ви род (т.е. отхвърляне на все пак вярна хипотеза), рискът за което е известен (той е точно равен на α), отказът ни от нулевата хипотеза може да бъде обусловен от някоя от следните причини:

- a) предполагаемата функционална зависимост $f(x, \eta) = 0$ между измеряемите величини и неизвестните параметри не се реализира (т.е. говорим за *грешен модел*, *грешна моделна функция*). Или видът на тази функция не е правилно избран с оглед на конкретния експеримент, или някои константи в модела, които сме считали за точно известни, са зададени неточно. В такъв случай е необходимо моделът да се преразгледа (евентуално да се допълни), както и стойностите на константите (понякога положението се подобрява, ако те се разглеждат също като неизвестни параметри, подлежащи на оценка);
- b) функционалната зависимост $f(x, \eta) = 0$ е правилна, но нейното Тейлорово разложение с точност само до линейни членове не е достатъчно, за да представи адекватно тази функция в разглежданата област от изменение на параметрите;
- c) началното приближение x_0 е твърде далеч от истинското решение x . По-добро начално приближение може да доведе до по-приемлива стойност на M (ако сме попаднали на локален минимум). Очевидно, този пункт е тясно свързан с предишния, тъй като всяка развиваема в ред на Тейлър функция в достатъчно близка околност на центъра може да бъде представена с линейни членове с произволна точност. От тази гледна точка трябва да отбележим, че си струва да се пожертвуват усилия (и време) за търсене на добро начално приближение, тъй като иначе биха се употребили повече на брой итерации за достигане на сходимост. От друга страна, изчислителното време за една итерация е значително по-голямо от това за пресмятане на M при $x = x_0 = fix$;
- d) Ковариационната матрица на измерванията $C_y = G_y^{-1}$, която често се основава само на груби оценки или даже на догадки, е невярна. В тази връзка едно сериозно подобрене

на оценката на C_y може да се постигне чрез извършването на многократни измервания при същите условия (т.е. да получим извадка) и да оценим C_y чрез емпиричните дисперсии и ковариации на основата на тази извадка.