Софийски Университет Св.Климент Охридски
Факултет по математика и информатика
Вероятности, Операционни изследвания и Статистика

доц. ДИМИТЪР Л. ВЪНДЕВ

Записки
по
Теория на вероятностите

СОФИЯ, януари, 2002

Съдържание

Увод

Цел на тези записки е да се даде едно допълнително пособие на студентите от всички специалности във ФМИ, което да ги снабди със сведенията, отсъстващи в стандартните български учебници. Материалът може да бъде намерен на сървера на ФМИ във вид удобен за разглеждане със стандартен броузер, и в .pdf формат за четене с Acrobat Reader (значително по-добро качество и удобен за печат).

Тези части, които са отразени в учебника на (Димитров и Янев 1990) са дадени в максимално съкратен вид.

Основният материал е използуван във Факултета по Математика и Информатика при четене на курсове и за специалностите Приложна математика и информатика.

Направени са няколко допълнения свързани с условните разпределения и условно математическо очакване. Вместо на лекции, материалът е разделен на теми. Те са по-малки и почти отговарят на въпросите от конспекта.

От литературата по вероятности, дадена накрая, специално искаме да отбележим знаменитата книга на (Фелър 1979), и хубаво написаните учебници на (Гнеденко 1965), претърпяли много издания (с леко различно съдържание).

Авторът е много благодарен на колегите си от бившата катедра Вероятности и Статистика, които си направиха труда да прочетат внимателно първия вариант на записките и да отбележат многобройните грешки. Специални благодарности дължа и на студента Панайот Добриков за положения труд в същото направление.

Това е текущ вариант на записките. Той все още съдържа много непълноти и затова постоянно се променя. Моля Ви да използувате последния вариант.

Тема 1

Аксиоматика на теория на вероятностите

В тази лекция си поставяме следните цели:

- да разгледаме генезиса на понятието вероятност;
- да въведем събития и действия с тях;
- да определим вероятностно пространство;
- да дадем примери за прости вероятностни пространства.

1.1 Емпирични основи

Историята ни учи, че основите на понятието "шанс" са твърде стари. Това, което хората първо са забелязали, е устойчивостта на средната аритметична с нарастването на броя наблюдения. В миналото например, мерките за дължина са се определяли с "усредняване". В Англия, една от популярните мерки за дължина се е определяла като средна дължина на ходилото на първите 30 човека излизащи от черквата в неделя сутринта, в древния Египет - като общата дължина на определен брой семена от свещенно растение.

Честотна вероятност

Пример 1.1 *Хвърляме монета многократно. Каква е честотата на получените ези?*

Нека означим общия брой хвърляния с n, а броят на получените ези с m. Тогава честотата m/n на поява на ези би трябвало да клони към едно постоянно число:

$$\text{Честотна вероятност} = \lim_{n\to\infty} \frac{m}{n} = \lim_{n\to\infty} \frac{\text{Брой на благоприятните изходи}}{\text{Брой на извършените опити}} \tag{1.1}$$

Така, ако монетата е правилна и хвърляме честно, би трябвало броят на езитата разделен на броя на опитите да клони към половина. Ако монетата не е правилна, граничната вероятност ще се окаже друго число, което е естествено да интерпретираме като оценка на шанса при един опит да получим ези.

Класическа вероятност

Първите опити да се построи математически модел са свързани с понятието равен "шанс". Предполага се, че даден опит има краен брой изходи, които са равноправни. При провеждане на опита се случва някой от тези изходи, при това всеки от тях може да се случи с еднакъв "шанс". Най-простите примери за такава концепция са свързани с хазартните игри, където се хвърлят зарове или използуват добре разбъркани тестета карти.

Пример 1.2 *Хвърляме зар. Каква е вероятността да получим четно число?*

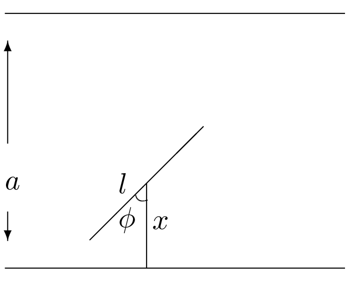
Рецептата е проста. Достатъчно е да преброим благоприятните изходи и разделим това число с броя на всички изходи:

$$\text{Класическа вероятност} = \frac{\text{Брой на благоприятните изходи}}{\text{Брой на всички възможни изходи}} \tag{1.2}$$

Така за нашата задача отговорът трябва да бъде $3/6 = 1/2$.

Геометрична вероятност

Пример 1.3 *(Задача на Бюфон) Хвърляме игла върху раирана покривка. Каква е вероятността иглата да пресече раето?*

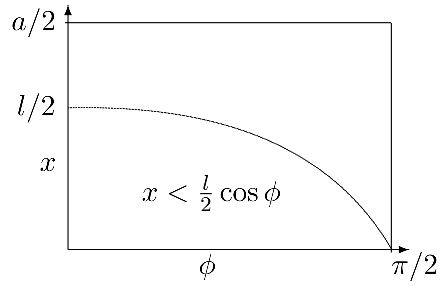


За да решим задачата трябва да формализираме условията. Нека означим с l дължината на иглата и с a - разстоянието между райетата. За простота ще сметнем, че широчината на едно рае е 0. Да означим с x разстоянието от средата на иглата до по - близкото рае,а с ϕ - острия ъгъл, които иглата сключва с перпендикуляра към същото рае.

Фиг. 1.1: Иглата на Бюфон 1

Тогава имаме $0 \leq x \leq a/2$ и $0 \leq \phi \leq \pi/2$. Това са всички възможности. Благоприятните (когато иглата пресече райето) се определят от неравенството: $(l/2)cos\phi > x$. Рецептата е проста:

$$\text{Геометрична вероятност} = \frac{\text{Площ на благоприятните изходи}}{\text{Обща площ}}$$



Така, ако $l < a$, задачата се свежда до пресмятането на интеграла

$$p = \frac{2l}{a\pi}\int_0^{\frac{\pi}{2}} \cos\alpha d\alpha = \frac{2l}{a\pi}.$$

Фиг. 1.2: Иглата на Бюфон 2

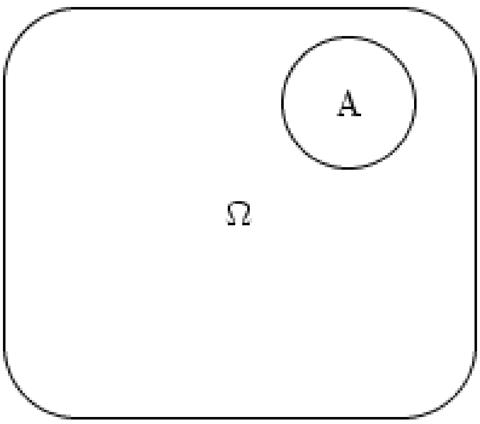
1.2 Аксиоматика

Теория на вероятностите става строга математическа теория едва след въвеждането в 1939 г. от А.Н.Колмогоров на следната аксиоматика, основана на теория на мярката (теория на интеграла).

Алгебра на събитията

Елементарно събитие е първично понятие – нещо като точка в геометрията.

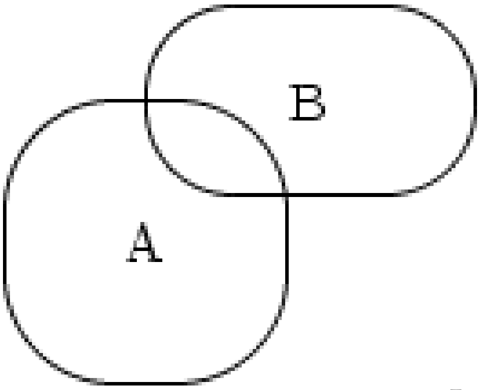
Множеството от всички елементарни събития наричаме "достоверно събитие" и означаваме с Ω. Празното множество бележим с $\emptyset$ и наричаме "невъзможно събитие".



Всички събития са подмножества на Ω и с тях могат да се правят обичайните в теория на множествата действия. В теория на вероятностите събитието A има смисъла на логическото твърдение *сбъднало се е някое от елементарните събития в* A. Със събитията могат да се правят обичайните за множествата действия:*допълнение, обединение, сечение*, които обаче носят други имена.

Фиг. 1.3: A в Ω

Допълнението $\Omega \setminus A$ на множеството A в Ω означаваме с $\overline{A}$ и наричаме допълнително събитие (или отрицание) на събитието A.



Сечението на множествата A, B означаваме с $A \cap B$ и казваме, че са се сбъднали съвместно събитията A и B.
Обединението на множествата A, B означаваме с $A \cup B$ и казваме, че се е сбъднало поне едно от събитията A и B. За краткост това се произнася сбъднало се е A или B.

Фиг. 1.4: A и/или B

Когато $A \subset B$ казваме, че събитието A "влече" събитието B.

Операциите със събития удовлетворяват обичайните свойства на операциите с множества. Те лесно се разпространяват и върху безкраен брой събития. Изпълнени са и т.н. закони на Де Морган:

$$\overline{\cup_k A_k} = \cap_k \overline{A_k}, \qquad \overline{\cap_k A_k} = \cup_k \overline{A_k} \tag{1.3}$$

За удобство са въведени и някои производни определения и операции:

- означаваме с $AB = A \cap B$;
- събитията A и B наричаме несъвместими, ако $AB = \emptyset$;
- за несъвместими събития вместо $A \cup B$ използуваме знака събиране - пишем $A + B$;
- означаваме с $A \Delta B = \overline{A}B + A\overline{B}$.

За да си осигурим възможността да правим всичките тези операции ще поискаме множеството от събития да го допуска.

Определение 1.1 *Семейство $\mathfrak{A}$ от подмножества на Ω се нарича булова алгебра, ако удовлетворява следните три условия:*

1. $\Omega \in \mathfrak{A}$;
2. *ако* $A \in \mathfrak{A}$,*то* $\overline{A} \in \mathfrak{A}$;
3. *ако* $A, B \in \mathfrak{A}$,*то* $A \cup B \in \mathfrak{A}$.

Веднага се вижда от 1.3, че буловата алгебра от множества е затворена и относно операциите $\cap, \Delta, +$. Тя обаче не е длъжна да бъде затворена относно операции с безкраен брой множества.

Определение 1.2 *Булова алгебра $\mathfrak{A}$, която е затворена относно изброимите операции обединение и сечение, се нарича булова σ-алгебра – ако $A_k \in \mathfrak{A}(k = 1, 2, \ldots)$, то $\cup_k A_k \in \mathfrak{A}$ и $\cap_k A_k \in \mathfrak{A}$.*

Определение 1.3 *Двойката* $(\Omega, \mathfrak{A})$*, където* $\mathfrak{A}$ *е булова* σ*-алгебра, се нарича измеримо пространство. Елементите на* $\mathfrak{A}$ *се наричат случайни събития.*

σ-алгебрите притежават някои универсални свойства. Например, сечение на произволен брой σ-алгебри е σ-алгебра. Това ни дава възможност да определим лесно минималната σ-алгебра съдържаща семейството множества $\mathfrak{F}$ като сечение на всички σ-алгебри, съдържащи семейството $\mathfrak{F}$. Ще означаваме тази σ-алгебра $\sigma(\mathfrak{F})$.

Вероятностно пространство

Определение 1.4 *Реалната функция* P*, определена върху елементите на буловата* σ*-алгебра* $\mathfrak{A}$*, се нарича вероятност, ако удовлетворява условията:*

1. *неотрицателност:* $\mathsf{P}(A) \geq 0, \forall A \in \mathfrak{A}$*;*
2. *нормираност:* $\mathsf{P}(\Omega) = 1$*;*
3. *адитивност: Ако* $A_i \cap A_j = \emptyset$*, то* $\mathsf{P}(A_1 + A_2 + \ldots) = \mathsf{P}(A_1) + \mathsf{P}(A_2) + \ldots$*.*

Определение 1.5 *Тройката* $(\Omega, \mathfrak{A}, \mathsf{P})$ *наричаме вероятностно пространство.*

От аксиомите 1.4 лесно следват следните свойства на случайните събития.

- $\mathsf{P}(\emptyset) = 0$;
- $\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A)$;
- *непрекъснатост в* $\emptyset$. Ако $A_i, i = 1, 2, \ldots$ е намаляваща редица от събития, т.е. $A_{i+1} \subset A_i$ и $\cap_i A_i = \emptyset$, то $lim_i \, \mathsf{P}(A_i) = 0$.

Да се върнем към примерите. Във пример 1.2 Ω се състои от 6 елемента, $\mathfrak{A}$ е множеството от всички подмножества на това крайно множество. Вероятността се определя просто – всички елементарни събития са равновероятни.

Значително по сложна е ситуацията при примера 1.3. Тук ролята на Ω се поема от множеството от всички точки (α, x) в правоъгълника $0 \leq \alpha \leq \pi/2$ и $0 \leq x \leq a/2$. σ-алгебрата $\mathfrak{A}$ се състои от измеримите по Лебег подмножества на този правоъгълник, т.е. тези на които можем да мерим лице или площ. Вероятността е относителната площ, заемана от тях в правоъгълника. Елементарните събития в това пространство притежават нулева вероятност.

Понякога се случва елементарните събития да не са събития, т.е. да не са елементи на $\mathfrak{A}$. За да избегнем тази и други неприятности обикновено попълваме $\mathfrak{A}$ добавяйки към нея всички подмножества на събития с нулева вероятност. Ако означим това семейство от множества с $\mathfrak{N}(\mathsf{P})$, то прието е вместо с $\mathfrak{A}$ да се работи с попълнената σ-алгебра $\sigma(\mathfrak{A}, \mathfrak{N}(\mathsf{P}))$. Разбира се, такова попълнение зависи от вероятността P. От тук нататък, когато вероятността във вероятностното пространство е фиксирана, ще предполагаме, че $\mathfrak{A} = \sigma(\mathfrak{A}, \mathfrak{N}(\mathsf{P}))$.

Със задачата за монетата (пример 1.1) ще се заемем подробно в специална лекция. Тук само ще конструираме едно възможно пространство от елементарни събития за нея. Ако записваме резултата от един експеримент с число – нула при получаване на тура и единица – при получаване на ези, получаваме Ω като множеството от всички безкрайни двоични редици. (Предполагаме, че хвърляме неограничен брой пъти).

Ще докажем една лема, която има самостоятелно значение.

Лема 1.1 *Нека вероятността* P *е зададена върху буловата алгебра* $\mathfrak{F}$ *и е адитивна. Тогава, необходимо и достатъчно условие да е непрекъсната в* $\emptyset$*, е тя да е* σ*-адитивна върху* $\mathfrak{F}$.

Доказателство: *Необходимост.* Нека $A = \sum_{n=1}^{\infty} A_n$ и $A, A_n \in \mathfrak{F}$. От адитивността на P за $\forall n$ следва равенството

$$\mathsf{P}(A) = \mathsf{P}(A_1) + \mathsf{P}(A_2) + \ldots + \mathsf{P}(A_n) + \mathsf{P}(\sum_{i=n+1}^{\infty} A_i). \tag{1.4}$$

От непрекъснатостта следва, че $\mathsf{P}(\sum_{i=n+1}^{\infty} A_i) \to 0$. Следователно, P е σ-адитивна върху $\mathfrak{F}$.

Достатъчност. Нека $B_{n+1} \subset B_n \in \mathfrak{F}$ и $\bigcap B_n = \emptyset$. Да положим $A_n = B_n \setminus B_{n+1}$. Тъй като $\forall n, B_n = \sum_{i=n}^{\infty} A_i$ и P е σ-адитивна върху $\mathfrak{F}$, получаваме, че

$$\mathsf{P}(B_n) = \sum_{i=n}^{\infty} \mathsf{P}(A_i) \longrightarrow 0$$

като остатък на сходящ се ред. Значи P е непрекъсната. □

Пример 1.4 *Вероятност, която не е непрекъсната в* $\emptyset$.

Да разгледаме множеството $\Omega = (0, .5]$. Върху всички отворени отляво и затворени отдясно интервали ($0 \leq a \leq b \leq .5$) определяме вероятността така:

$$\mathsf{P}((a,b]) = \begin{cases} b-a, & a > 0 \\ .5 + b, & a = 0 \end{cases}$$

Множеството от крайните обединения на такива интервали $\mathfrak{F}$ е буловата алгебра. Проверете го. Проверете, че P е зададена коректно, но не е непрекъсната в $\emptyset$.

1.3 *Теорема за продължението

Следната знаменита теорема се нарича теорема за продължение на вероятности върху σ-алгебра.

Теорема 1.1 *Ако една вероятност, зададена върху буловата алгебра, е непрекъсната в* $\emptyset$*, то тя е продължима еднозначно върху* $\sigma(\mathfrak{F})$.

Доказателство: Виж в приложението.□

Граница на редица събития

В буловата σ-алгебра $\mathfrak{A}$ сме в състояние да въведем граница на събития и така, че тя да се окаже събитие. Означаваме:

$$A^* = \limsup_{n\to\infty} A_n = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k, \qquad A_* = \liminf_{n\to\infty} A_n = \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k \tag{1.5}$$

Интерпретацията на така определените гранични събития е следната:

- A^* - се състои от тези елементарни събития, които влекат безкраен брой елементи A_n;
- A_* - се състои от тези елементарни събития, които влекат всички елементи A_n от дадено място нататък;

Наистина, $\omega \in A^*$ може да се интерпретира така: $\forall n \exists k > n$ такова, че $\omega \in A_k$. Аналогично, $\omega \in A_*$ може да се интерпретира така: $\exists n \forall k > n$ такова, че $\omega \in A_k$. Очевидно при това $A_* \subset A^*$.

Метрично пространство на събитията

Нека определим за всеки две събития от една алгебра $\mathfrak{F}$ числото $\rho(A,B) = \mathsf{P}(A\Delta B)$. Лесно се проверява, че то е неотрицателно и удовлетворява неравенството на триъгълника. От никъде, обаче не следва, че ако $\rho(A,B) = 0$, то $A = B$, както това става в крайно вероятностно пространство с класическа вероятност.

Нека $\rho(A,B)$ е определено на σ-алгебра с σ-адитивна вероятност. Проверете, че сходимостта с това разстояние е еквивалентна на условието: $\mathsf{P}(A^* \setminus A_*) = \rho(A^*, A_*) = 0$.

За да може $\rho(A,B)$ да стане разстояние (или метрика) е необходимо да разгледаме множеството от класове еквивалентни събития $\tilde{\mathfrak{A}}$, (казваме че $A \sim B$, ако $\rho(A,B) = 0$). Това множество е също σ-алгебра, но освен това става (компактно) метрично пространство (зависещо от вероятността P).

Тема 2

Условна вероятност

В тази лекция си поставяме следните цели:

- да определим понятието условна вероятност;
- да докажем знаменитите формули за пълната вероятност и на Бейс;
- да определим условни вероятностни пространства;
- да дадем примери и контрапримери.

2.1 Формула на пълната вероятност

Определение 2.1 *Нека $B \in \mathfrak{A}$ и $\mathsf{P}(B) > 0$. За всяко събитие $A \in \mathfrak{A}$ ще наречем числото*

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

условна вероятност на събитието A при условие събитието B.

Лесно е да се види, че ако фиксираме условието B, условната вероятност притежава всичките свойства на безусловната. Събитието B (и всички съдържащи го събития) притежава условна вероятност 1. Събитията влечащи B повишават своята вероятност, а несъвместимите с B стават "невъзможни". Така върху същата σ-алгебра е породена нова вероятност отразяваща факта за настъпването на събитието B. Нека я означим с P_B.

Определение 2.2 *Казваме, че събитията $(H_1, H_2, \ldots, H_n)$ образуват пълна група в (или крайно разделяне γ на) Ω, когато събитията са несъвместими ($H_iH_j = \emptyset, \forall i \neq j$) и изчерпват достоверното събитие ($H_1 + H_2 + \cdots + H_n = \Omega$). Прието е събитията от пълната група да се наричат хипотези.*

Нека е зададена пълната група събития $(H_1, H_2, \dots, H_n)$. Изпълнена е следната формула за пълната вероятност:

$$\mathsf{P}(A) = \sum_{i=1}^{n} \mathsf{P}(A|H_i)\,\mathsf{P}(H_i). \tag{2.1}$$

Доказателство: Следва лесно от очевидното равенство:

$$A = AH_1 + AH_2 + \cdots + AH_n,$$

определения 2.1 (на условна вероятност) и 1.4 (на вероятност). □

Пример 2.1 *На лов за патици. Едно младо семейство отишло на лов за патици. Те хвърлили чоп кой да стреля по първата патица. Пита се каква е вероятността да са я улучили, ако е известно, че:*

1. *чоп са хвърляли честно с правилна монета;*
2. *мъжът улучва средно в един от пет изстрела;*
3. *мъжът улучва средно два пъти по-често от жената.*

Решение. Да означим с A събитието мъжът да е стрелял пръв. Според условие 1. $P(A) = 1/2$. Да означим с B събитието патицата да е улучена. Според условие 2. $P(B/A) = 1/5$. Според условие 3. $P(B/A) = 2 * P(B/\overline{A})$. Следователно, $P(B) = P(A) * P(B/A) + P(\overline{A}) * P(B/\overline{A}) = 3/20$. □

Теорема 2.1 *(Формула за умножение на вероятности) Вярна е следната формула:*

$$\mathsf{P}(A_1A_2\dots A_n) = \mathsf{P}(A_1)\,\mathsf{P}(A_2|A_1)\,\mathsf{P}(A_3|A_1A_2)\dots\mathsf{P}(A_n|A_1A_2\dots A_{n-1}). \tag{2.2}$$

Доказателство: Ще докажем твърдението по индукция. За $n = 2$ то е очевидно следствие от определение 2.1 на условна вероятност. Нека то е изпълнено за някое n. Тогава да приложим същото определение за събитията $B = A_1A_2\dots A_n$ и A_{n+1}:

$$\mathsf{P}(A_1A_2\dots A_{n+1}) = \mathsf{P}(BA_{n+1}) = \mathsf{P}(B)\,\mathsf{P}(A_{n+1}|B)$$

$$= \mathsf{P}(A_1)\,\mathsf{P}(A_2|A_1)\,\mathsf{P}(A_3|A_1A_2)\dots\mathsf{P}(A_n|A_1A_2\dots A_{n+1}).$$

□

Пример 2.2 *Пак на лов за патици. Каква е вероятността да се улучат три патици подред от първия път, ако договорката между двамата е била, който улучи пръв, да стреля докато улучва.*

Решение. Да означим с B_i събитието "улучена е i-тата патица" и $B = B_1B_2B_3$. Съгласно формулата за пълна вероятност имаме:

$$P(B) = P(A) * P(B/A) + P(\overline{A}) * P(B/\overline{A}).$$

Съгласно формулата за умножение имаме:

$$P(B/A) = P(B_1B_2B_3/A) = P(B_1/A)P(B_2/AB_1)P(B_3/AB_1B_2) = (P(B_1/A))^3.$$

Тук неявно предполагаме, че вероятностите за улучване са едни и същи и не зависят от резултатите от предходните опити (едно доста съмнително предположение). Така окончателно получаваме $P(B) = 1/2 * (1/5)^3 + 1/2(1/10)^3 = (9/2)10^{-3} = 0.45\%$ □

2.2 Формула на Бейс

Следната знаменита формула на Бейс намира широко приложение в статистиката:

$$\mathsf{P}(H_k|A) = \frac{\mathsf{P}(A|H_k)\,\mathsf{P}(H_k)}{\sum_{i=1}^{n} \mathsf{P}(A|H_i)\,\mathsf{P}(H_i)}. \tag{2.3}$$

Доказателство: Следва от определение 2.1 на условна вероятност. □

Пример 2.3 *И пак на лов за патици. Един съсед бил добре запознат с тяхната уговорка и наблюдавал иззад баира, как с три изстрела семейството сваля три патици. Запитал се той, кой ли всъщност е стрелял или колко по-голяма е вероятността да се е паднало на мъжа да стреля.*

Решение. Ще използуваме същите означения. Тъй като $P(B_i/A) = 2 * P(B_i/\overline{A})$, то $P(B/A) = 8 * P(B/\overline{A})$. Тогава

$$P(\overline{A}/B) = \frac{1/2P(B/\overline{A})}{1/2P(B/A) + 1/2P(B/\overline{A})} = \frac{P(B/\overline{A})}{8 * P(B/\overline{A}) + P(B/\overline{A})} = 1/9,$$

$$P(A/B) = 1 - P(\overline{A}/B) = 8/9.$$

Забележете, че тук даже не се нуждаем от точната величина на условната вероятност $P(B/A)$. Достатъчни са само предположения 1 и 3, описани в примера 2.1. Оказва се 8 пъти по-вероятно да се е паднало на мъжа да стреля пръв. □

2.3 Условни вероятностни пространства

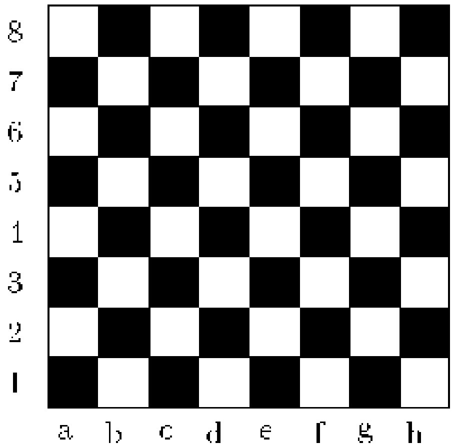
Както вече отбелязахме, всяко крайно разделяне на пространството γ на непресичащи се събития води до появата на серия от условни вероятности, определени върху $\mathfrak{A}$. Това ни дава възможност да препишем формула (2.1) във вида:

$$\mathsf{P}(A) = \sum_{B\in\gamma} \mathsf{P}(B)\,\mathsf{P}_B(A). \tag{2.4}$$

Горното равенство се разбира като равенство на вероятностни мерки, т.е. то е изпълнено за всяко събитие $A \in \mathfrak{A}$.

Понякога, обаче, е удобно да разглеждаме условните вероятности като съсредоточени само върху събитието – условие. Тогава трябва да определим подходящо σ-алгебра $\mathfrak{A}_B$ на B и разглеждаме вероятностното пространство $(B, \mathfrak{A}_B, \mathsf{P}_B)$. Това се прави тривиално, когато условието B има ненулева вероятност: $\mathfrak{A}_B = \{A \bigcap B, A \in \mathfrak{A}\}$, $\mathsf{P}(A) = \mathsf{P}(A)/\,\mathsf{P}(B), A \in \mathfrak{A}_B$. Такива вероятностни пространства наричаме условни.

Пример 2.4 *Нека разгледаме една квадратна шахматна дъска.*



Да свържем с нея вероятностно пространство съгласно формулата за геометрична вероятност. Тъй като дъската е разделена на 64 еднакви квадратчета, вероятността една случайна точка да попадне в такова квадратче е $1/64$. Ние, обаче, ще променим тези вероятности като ги заменим на числата $\{p_{i,j}, i = 1, 2, \ldots, 8, \quad j = a, b, \ldots, h$. Ясно е, че стига да са изпълнени условията $p_{i,j} > 0$ и $\sum p_{i,j} = 1$, ние ще получим едно добре определено вероятностно пространство.

Фиг. 2.1: Шахматна дъска

В това пространство има три различни естествени разделяния:

- γ_1 - на колони $j = a, b, \ldots, h$;
- γ_2 -на редове $i = 1, 2, \ldots, 8$;
- γ_3 -на два цвята - бели и черни.

Нека означим с $p_{i,.} = \sum_j p_{i,j}, \quad p_{.,j} = \sum_i p_{i,j}$. Така, когато се случи, например, събитието $\{j = b\} \in \gamma_1$, получаваме условното вероятностно пространство, свързано с колоната: $\{b\}$. В него има 8 полета с означения:
$1b, 2b, \ldots, 8b$ и условни вероятности, съответно, $p_{1,b}/p_{.,b}, p_{2,b}/p_{.,b}, \ldots p_{8,b}/p_{.,b}$.

Възможна е и обратната интерпретация. Нека вземем един набор от различни вероятностни пространства: $\{(\Omega_i, \mathfrak{A}_i, \mathsf{P}_i), i = 1, 2, ...\}$

Сега можем да образуваме директната сума от множества: $\Omega = \Omega_1 + \Omega_2 + ...$ Да определим на Ω формалната сума от $\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2 + ...$ С всеки набор от вероятности $\{p_i\}, (p_i > 0, \sum p_i = 1)$ е свързана една вероятностна мярка на $\mathfrak{A}$, определена по формулата (2.1).

Така формулата за пълна вероятност съответствува на едно разпадане на вероятностното пространство в "претеглена сума" от условни вероятностни пространства.

2.4 Измерими разделяния и булови σ-подалгебри

С всяко крайно разделяне γ на пространството Ω е свързана еднозначно определена булова подалгебра $\mathfrak{A}_\gamma \subset \mathfrak{A}$. Тя е крайна и съдържа точно 2^N елемента, ако разделянето γ се състои от N непразни множества. Изобщо казано е вярна следната

Теорема 2.2 *Съществува взаимно - еднозначно съответствие между крайните булови алгебри и крайните разделяния на пространството* Ω.

Доказателство: 1.Да съпоставим на всяко разделяне булова алгебра. Нека $\gamma = \{H_1, H_2, \ldots H_n\}$. Нека означим с N множеството $\{1, 2, \ldots, n\}$ и нека $I \subset N$. Да разгледаме семейството от множества от вида: $A_I = \sum_{i \in I} H_i$. Те образуват булова подалгебра на $\mathfrak{A}$. Наистина, $A_\emptyset = \emptyset$, $A_N = \Omega$, $A_{\overline{I}} = \overline{A_I}$, $A_{I \cup J} = A_I \cup A_J$. Да означим тази подалгебра $\mathfrak{A}_\gamma$.
2. Обратно, нека $\mathfrak{B}$ е крайна (състояща се от краен брой множества) булова подалгебра $\subset \mathfrak{A}$. Ще наречем атом на $\mathfrak{B}$ всяко множество от вида:

$$A(\omega) = \cap_{B \in \mathfrak{B}} B^{(\omega)}, \qquad B^{(\omega)} = \begin{cases} B, & \omega \in B \\ \overline{B}, & \omega \in \overline{B} \end{cases}$$

Така на всяко $\omega \in \Omega$ съпоставяме единствено множество $A(\omega) \in \mathfrak{B}$, което го съдържа - значи множеството от атоми образува разделяне. Взаимната еднозначност на това съответствие предоставяме на читателя. □

Това съответствие поражда интересна връзка между операциите определени върху подалгебри и разделяния. На тривиалната алгебра $\{\emptyset, \Omega\}$ отговаря тривиалното разделяне $\nu = \{\Omega\}$.

Определение 2.3 *Казваме, че δ е по-ситно (или по фино) от γ($\gamma \le \delta$) ако $\mathfrak{A}_\gamma \subset \mathfrak{A}_\delta$.*

Това означава, че елементите на разделянето γ са по-груби, те се състоят от по няколко елемента на разделянето δ. И двете множества – това на буловите алгебри и това на разделянията – са частично наредени. Но тъй като съществуват max, min те са и решетки. В частност,

$$max(\mathfrak{A}_\delta, \mathfrak{A}_\gamma) = \sigma(\mathfrak{A}_\delta, \mathfrak{A}_\gamma) = \mathfrak{A}_{max(\delta,\gamma)}.$$

Разделянето $max(\delta, \gamma)$ се състои от всевъзможни сечения на елементи от двете разделяния.

$$min(\mathfrak{A}_\delta, \mathfrak{A}_\gamma) = \mathfrak{A}_\delta \bigcap \mathfrak{A}_\gamma = \mathfrak{A}_{min(\delta,\gamma)}.$$

Разделянето $min(\delta, \gamma)$ се състои от събития, които могат да се представят като обединения на елементи, както от едното, така и от другото разделяне.

Благодарение на това двете структури - на булови алгебри и разделяния стават изоморфни. Този изоморфизъм може да се разшири и за σ-подалгебри, но това е възможно само след въвеждането на фиксирана вероятност. Тогава всички σ-подалгебри се попълват с подмножествата на събития с нулева вероятност и се разглеждат само класовете еквивалентни събития. Когато σ-алгебрата $\mathfrak{A}$ е породена от някакво изброимо семейство множества, което разделя точките във вероятностното пространство Ω, теорема 2.2 може да се обобщи.

Теорема 2.3 *Съществува взаимно - еднозначно съответствие между σ-подалгебрите на $\mathfrak{A}$ и измеримите разделяния на Ω.*

Един такъв пример ще видим в теорема 8.2 по-късно за схемата на Бернули.

Тема 3

Независимост

Независимостта е най-фундаменталното понятие на теорията на вероятностите. Макар че, както ще видим нататък, тя е някакъв еквивалент на декартовото произведение на множества, или на правото произведение на алгебри, т.е. в математически смисъл едва ли привнася нещо ново, независимостта в действителност е основата на тази теория. Това е понятието, което прави теорията незаменима, когато има нужда от математическо моделиране на явления с непредсказуем изход.

Независимостта, като строго понятие от математиката, се оказва неимоверно близка до нормалните, езикови или човешки представи за същото – кога едно събитие оказва (или не) някакво влияние върху възможността друго събитие да настъпи.

Независимостта има и недостатъци — изискванията са толкова строги, че стават непроверяеми. Когато казваме, че две събития са независими, ние влагаме много повече вяра, от колкото бихме могли да проверим.

3.1 Независимост на събития

Регистрацията на настъпване на дадено случайно събитие променя състоянието на вероятностното пространство – вече е невъзможно настъпването на елементарни събития извън (не влечащи) това събитие. Тази ситуация е отразена в изменението на вероятността на другите събития – условната им вероятност не винаги е същата като безусловната.

В някои, редки случаи, обаче настъпването на някои събития не оказва такова влияние върху шансовете на други събития.

Тук ще дадем формално определение на понятието независимост. Ще се убедим, че в тази си формулировка, то изключва някаква причинно следствена връзка между явленията, които наричаме независими.

Определение 3.1 *Казваме че събитията A, B са независими, ако е изпълнено равенството $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$. Ще бележим независимите събития $A \perp\!\!\!\perp B$.*

От това определение веднага следва, че условната вероятност на всяко от двете събития при условие, че е настъпило другото, е равна на безусловната му вероятност. С други думи, вероятността да настъпи събитието A не зависи от това, дали е настъпило или не, събитието B.

Определение 3.2 *Казваме, че разделянията γ и δ са независими, ако $\forall A \in \gamma$ и $\forall B \in \delta$ имаме $A \perp\!\!\!\perp B$. Ще бележим независимите разделяния $\gamma \perp\!\!\!\perp \delta$.*

3.2 Независимост в съвкупност

Определение 3.3 *Казваме че събитията $\{A_k, \quad k = 1, 2, \dots, n\}$ са независими в съвкупност, ако вероятността на всяко от тях не зависи от това дали се е случила някоя комбинация от останалите събития.*

От това определение следва силно опростяване на формулата за умножение (2.2), когато събитията са независими в съвкупност:

$$\mathsf{P}(A_1 A_2 \dots A_n) = \mathsf{P}(A_1)\,\mathsf{P}(A_2)\dots\mathsf{P}(A_n). \tag{3.1}$$

Нещо повече, тя е изпълнена и за произволна комбинация от тези събития.

Пример 3.1 *Да разгледаме следното вероятностното пространство състоящо се от 4 равновероятни елементарни събития: $\{\omega_i, i = 1, 2, 3, 4\}$. Тогава събитията $A = \{\omega_1, \omega_2\}, B = \{\omega_1, \omega_3\}, C = \{\omega_1, \omega_4\}$ са независими две по две, но не са независими в съвкупност.*

Наистина, $\mathsf{P}(A) = \mathsf{P}(B) = \mathsf{P}(C) = \frac{1}{2}$, $\mathsf{P}(AB) = \mathsf{P}(BC) = \mathsf{P}(AC) = \frac{1}{4}$, но $\mathsf{P}(ABC) = \mathsf{P}(\{\omega_1\}) = \frac{1}{4} \neq \frac{1}{8}$.

3.3 Произведение на вероятностни пространства

Нека са зададени две вероятностни пространства: $(\Omega_1, \mathfrak{A}_1, \mathsf{P}_2)$ и $(\Omega_2, \mathfrak{A}_2, \mathsf{P}_2)$. Да образуваме декартовото произведение Ω на двете множества Ω_1 и Ω_2:

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2), \quad \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}.$$

Нека разгледаме Ω и в него множеството от всички "правоъгълници" $\Pi = \{A_1 \times A_2, A_1 \in \mathfrak{A}_1, A_2 \in \mathfrak{A}_2\}$. Вероятността върху "правоъгълниците" се определя просто: $\mathsf{P}(A_1 \times A_2) = \mathsf{P}_1(A_1) * \mathsf{P}_2(A_2)$. Множеството Π очевидно не е булова алгебра.

Теорема 3.1 *Множеството $\mathfrak{F}$ от крайни суми на непресичащи се "правоъгълници" е булова алгебра.*

Доказателство: Трябва да проверим аксиомите. В доказателството за краткост ще използваме означението C (понякога с индекси) за означаване на правоъгълници.
1. Аксиомата $\Omega \in \mathfrak{F}$ е очевидно изпълнена: $\emptyset = \emptyset_1 \times \emptyset_2 = \Omega_1 \times \emptyset_2 = \emptyset_1 \times \Omega_2, \Omega = \Omega_1 \times \Omega_2$.;
2. Ако $A, B \in \mathfrak{F}$, то $A \cap B \in \mathfrak{F}$.
Това следва непосредствено от определението на декартово произведение:

$$A_1 \times A_2 \bigcap B_1 \times B_2 = A_1 \bigcap B_1 \times A_2 \bigcap B_2 = C \in \mathfrak{F}. \tag{3.2}$$

От (3.2) следва:

$$(\sum_i A^i \times B^i)\bigcap(\sum_j A^j \times B^j) = (\sum_i C_1^i)\bigcap(\sum_j C_2^j) = \sum_{i,j} C_1^i \cap C_2^j, \tag{3.3}$$

което изразява сечението на два произволни елемента от $\mathfrak{F}$. Това доказателство тривиално се разширява за краен брой елементи на $\mathfrak{F}$.
3. Ако $A \in \mathfrak{F}$, то $\bar{A} \in \mathfrak{F}$.
Допълнението на един правоъгълник се получава просто като сума на три непресичащи се правоъгълника:

$$\overline{A \times B} = \bar{A} \times B + A \times \bar{B} + \bar{A} \times \bar{B} = C_1 + C_2 + C_3. \tag{3.4}$$

Формули (3.3) и (3.4) и законите на Де Морган дават възможност да изразим лесно и допълнението на произволен елемент от $\mathfrak{F}$:

$$\overline{\sum_i A_i \times B_i} = \bigcap_i \overline{A_i \times B_i} = \bigcap_i (C_1^i + C_2^i + C_3^i).\square \tag{3.5}$$

Определението на P се разширява просто от Π върху $\mathfrak{F}$:

$$\mathsf{P}(\sum_{i=1}^n A_1^i \times A_2^i) = \sum_{i=1}^n \mathsf{P}(A_1^i \times A_2^i).$$

Означаваме σ-алгебрата породена от правоъгълниците $\mathfrak{A} = \sigma(\Pi) \supset \mathfrak{F}$. Вероятността P на $\mathfrak{F}$ се оказва непрекъсната в $\emptyset$ и можем да използуваме теоремата 1.1 (на Каратеодори), която ни дава съществуване и единственост на вероятността P върху $\mathfrak{A}$. Така полученото вероятностно пространство наричаме произведение и бележим с: $(\Omega_1, \mathfrak{A}_1, \mathsf{P}_1) \times (\Omega_2, \mathfrak{A}_2, \mathsf{P}_2)$.

Лема 3.1 *Ако* $C = \sum_{i=1}^\infty C_i, \quad C, C_i \in \Pi,$ *то* $\mathsf{P}(C) = \sum_{i=1}^\infty (C_i)$.

Доказателство: Да отбележим, че ако един правоъгълник $A = A_1 \times A_2$ се раздели на сума (от непресичащи се правоъгълници) $A = B + C$, то това може да стане само по два начина – чрез разделяне на A_1 или A_2. Във всеки от двата случая имаме: $\mathsf{P}(A) = \mathsf{P}(B) + \mathsf{P}(C)$. По индукция това свойство се прехвърля и върху безкрайно разделяне на правоъгълници. Да го наречем *частична σ-адитивност.* $\square$

Теорема 3.2 *Вероятността* P *определена върху буловата алгебра* $\mathfrak{F}$ *е непрекъсната в нулата.*

Доказателство: Да разгледаме намаляваща редица от елементи $B_n \in \Pi$, такава, че $\bigcap B_n = \emptyset$. Да означим с $A_n = B_n \setminus B_{n+1}$. Съгласно теорема 3.1 имаме $A_n \in \mathfrak{F}$. Да отбележим, че поради предположението имаме $B_1 = \sum_{n=1}^\infty A_n$. Тъй като множествата A_n са непресичащи се, за $\forall n$ е изпълнено равенството:

$$\mathsf{P}(B_1) = \mathsf{P}(A_1) + \mathsf{P}(A_2) + \ldots + \mathsf{P}(A_n) + \mathsf{P}(B_{n+1}).$$

Тъй като B_1, $A_n, n = 1, 2, ...$ са крайни суми от "правоъгълници", остана да пренаредим елементите им така, че да получим последователни разделяния на елементите на B_1. Тогава можем да се възползуваме от отбелязаната частична σ-адитивност на P. $\square$

3.4 Независимост на σ-алгебри и разделяния

Определение 3.4 *Казваме, че дадени σ-алгебри $\{\mathfrak{A}_i, i = 1, ..., n\}$ са независими две по две (в съвкупност), ако събитията от всеки набор $\{A_i, A_i \in \mathfrak{A}_i, i = 1, ..., n\}$ са независими две по две (в съвкупност).*

В пространството $(\Omega_1, \mathfrak{A}_1, \mathsf{P}_1) \times (\Omega_2, \mathfrak{A}_2, \mathsf{P}_2)$, построено в предната секция има две независими σ-подалгебри:

- $\mathfrak{A}_1 = \{A_1 \times \Omega_2, A_1 \in \mathfrak{A}_1\}$;
- $\mathfrak{A}_2 = \{\Omega_1 \times A_2, A_2 \in \mathfrak{A}_2\}$.

Пример 3.2 *Хвърляне на два зара. В пример 3.2 вероятностното пространство е произведение на две пространства:*

- *резултат от първи зар $\Omega_1 = \{1, 2, ..., 6\}$;*
- *резултат от втори зар $\Omega_2 = \{1, 2, ..., 6\}$;*

Пример 3.3 *Нека сега разгледаме отново пример 2.4. Нека всички полета на шахматната дъска са равновероятни.*

Сега е лесно да се види, че трите разделяния: γ_1 - на колони, γ_2 -на редове,γ_3 -на два цвята - бели и черни са независими две по две, но не са независими в съвкупност.

Пример 3.4 *В задачата на Бюфон (пример 1.3) вероятностното пространство също се оказва произведение на две пространства:*

$$\Omega_1 = \{0 \leq \alpha \leq \pi/2\} \text{ и } \Omega_2 = \{0 \leq x \leq a/2\}.$$

Тема 4

Случайни величини

В тази лекция си поставяме следните цели:

- да определим случайна величина (сл.в.);
- да определим разпределение на сл.в.;
- да дадем примери за сл.в..

Случайните събития представляват най-простия пример за модел на наблюдение със случаен (неопределен отнапред) изход. Често се налага да на практика наблюденията да бъдат всъщност измервания – резултатът от експеримента да се записва с число. Модел на такива експерименти са случайните величини.

4.1 Прости случайни величини

Сл.в. са числови функции определени върху множеството от елементарни събития Ω, като точното определение зависи от това кои са случайните събития в Ω.

Определение 4.1 *Нека е зададена пълната група събития (разделянето) $(H_1, H_2, \ldots, H_n)$. Ще казваме, че е определена проста сл.в., ако $\xi(\omega) = x_i, \forall\omega \in H_i, i = 1, 2, \ldots, n$.*

Пример 4.1 *Нека разгледаме вероятностното пространство от пример 3.2: 'Хвърляне на два зара.' То се състои от 36 равновероятни елементарни събития. Да ги означим с $w_{i,j}, i, j = 1, 2, \ldots, 6)$. Да определим на това вероятностно пространство 2 сл.в. $\xi(w_{i,j}) = i, \eta(w_{i,j}) = j$.*

Намерете пълните групи на двете сл.в.

Теорема 4.1 *Линейна комбинация, произведение, минимум, максимум и функция на прости сл.в. е проста сл.в.*

Доказателство: Нека пълната група събития $(H_1, H_2, \ldots, H_n)$ съответствува на сл.в. ξ, пълната група събития $(G_1, G_2, \ldots, G_m)$ на сл.в. η, а техните стойности са съответно $\{x_1, x_2, \ldots x_n\}$ и $\{y_1, y_2, \ldots y_m\}$.

Първо ще докажем, че събитията $\{H_iG_j,\, i=1,2,\ldots,n,\, j=1,2,\ldots,m\}$ образуват пълна група.

$$H_iG_j \bigcap H_kG_l = H_iH_k \bigcap G_jG_l = \emptyset, \text{ когато } i \neq k \text{ или } j \neq l.$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} H_i \bigcap G_j = \sum_{i=1}^{n} H_i \bigcap \sum_{j=1}^{m} G_j = \sum_{i=1}^{n} H_i\Omega = \sum_{i=1}^{n} H_i = \Omega.$$

Тогава сл.в. $\alpha\xi+\beta\eta$, $\xi\eta$, $min(\xi,\eta)$, $max(\xi,\eta)$ ще приемат стойности, съответно, $\alpha x_i+\beta y_j$, x_iy_j, $min(x_i,y_j)$, $max(x_i,y_j)$ за всяко елементарно събитие от събитието H_iG_j от пълната група от събития. Функцията $f(\xi,\eta)$, съответно, ще приема стойности $f(x_i,y_i)$ върху същото събитие. $\Box$

Определение 4.2 *Казваме, че простите сл.в. ξ и η са независими, ако е независимо всяко от събитията на едната пълна група с всяко от събитията на другата пълна група. Бележим това с $\xi \perp\!\!\!\perp \eta$.*

Покажете, че в примера с двата зара сл.в. $\xi \perp\!\!\!\perp \eta$.

Определение 4.3 *Ще казваме, че функцията $\xi(.)$, определена на Ω с стойности в R, е сл.в., ако $L_x(\xi) = \{\omega : \xi(\omega) < x\} \in \mathfrak{A}, \forall x \in R^1$.*

От това определение в частност следва, че множествата $L_{a,b}(\xi) = \{\omega : a \leq \xi(\omega) < b\} \in \mathfrak{A}, \forall a,b \in R^1$. Наистина, $L_{a,b}(\xi) = L_a(xi) \cap \overline{L}_b(\xi)$.

Теорема 4.2 *Всяка сл.в. може да се представи като граница на редица от прости сл.в.*

Доказателство: Нека ξ е ограничена сл.в., т.е. с множество от стойности: $-C \leq \xi(\omega) < C$. За всяко n можем да определим крайно разделяне на пространството: $H_i = \{a_{i-1} \leq \xi(\omega) < a_i\}$, където числата $a_i = -C + i * \frac{2C}{n}$. Лесно се проверява, че множествата H_i образуват разделяне: $H_i \in \mathfrak{A}$, $H_i \cap H_j = \emptyset$, $\cup H_i = \Omega$. Да определим простата сл.в. $\xi_n(\omega) = a_i, \quad \omega \in H_i$. Тогава $|\xi(\omega) - \xi_n(\omega)| \leq \frac{2C}{n}$. Значи $\xi(\omega) = \lim \xi_n(\omega)$.

Сега ще покажем, че всяка сл.в. може да се представи като граница на редица от ограничени сл.в. ξ_n. Да означим

$$\xi_n(\omega) = \begin{cases} \xi(\omega), & L_{-n,n} = \{\omega : -n \leq \xi(\omega) < n\} \\ 0, & \{\xi(\omega) < -n\} \cup \{n \leq \xi(\omega)\} \end{cases}$$

Тъй като $\cup_n L_{-n,n} = \Omega$, за всяко $\omega \in \Omega$ съществува N такова, че $\xi_n(\omega) = \xi(\omega)$ при $n > N$. Значи $\xi(\omega) = \lim \xi_n(\omega)$. $\Box$

Теорема 4.3 *Линейна комбинация, произведение и измерима функция на сл.в. е сл.в.*

Доказателството на тази теорема изисква известна математическа подготовка и ще бъде дадено в последната секция.

4.2 Функция на разпределение и плътност

Определение 4.4 *Ще наричаме функция на разпределение на сл.в. ξ функцията $F(x) = \mathsf{P}(w : \xi(w) < x)$.*

Функцията $F(x)$ е монотонно ненамаляваща и непрекъсната от ляво. Освен това $F(-\infty) = 0, \quad F(\infty) = 1$. В термини на разпределението си сл.в. се класифицират лесно.

Определение 4.5 *Случайната величина, която приема стойности $x_1, x_2, x_3, \ldots$ с вероятности съответно $p_1, p_2, p_3, \ldots$ се нарича дискретна.*

Естествено $\sum p_i = 1$ и $p_i \geq 0$. Тогава функцията на разпределение има само скокове в точките x_i, навсякъде другаде е константа. В точката x_i скокът ѝ е равен точно на числото p_i.

Определение 4.6 *Ако съществува функция $f(x) \geq 0$ такава, че за всяко x $\int_{-\infty}^{x} f(x)dx = F(x)$, ще наричаме сл.в. непрекъсната, а функцията $f(x)$ наричаме плътност.*

Тогава, разбира се $F(x)$ е непрекъсната, т.е. няма никакви скокове, и плътността е производна на функцията на разпределение $f(x) = F'(x)$ за почти всяко x. Тук почти всяко означава всяко с изключение на изброимо много.

Естествено е, че за да бъде една функция плътност на случайна величина, тя трябва да отговаря на две изисквания:

- неотрицателност $f(x) \geq 0$ и
- нормираност - $\int_{-\infty}^{\infty} f(x)dx = 1$.

.

4.3 *Разпределения на сл.в и вектори

Да означим с 2^B множеството от всички подмножества на множеството B.

Теорема 4.4 *Ако $f(\cdot) : A \longrightarrow B$, то съответствието "пълен праобраз" $f^{-1}(\cdot) : 2^B \longrightarrow 2^A$ прехвърля всяка σ-алгебра на B в σ-алгебра на A.*

Доказателство: Само проверяваме аксиомите на σ-алгебра.$\square$

Напомняме, че σ-алгебрата на Бореловите множества в R^n е минималната σ-алгебра, съдържаща всички крайни интервали: $\cap_i \{a_i \leq x_i < b_i), i = 1, 2, \ldots, n$. Нека я означим с $\mathfrak{B}$. Елементите ѝ ще наричаме Борелови множества.

Нека ξ е сл.в. Тогава очевидно всички множества от елементарни събития от вида: $\xi^{-1}(A) = \{w : \xi(w) \in A\}$, където A е интервала (x, y) от реалната права ще са събития. Произволен набор от сл.в. може да се разглежда като векторна сл.в.

Теорема 4.5 *Ако ξ е сл.в., то съответствието "пълен праобраз" $\xi^{-1}(.)$ прехвърля $\mathfrak{B}$ в σ-подалгебра $\mathfrak{A}_\xi \subset \mathfrak{A}$, за елементите на която е определена вероятността P.*

Доказателство: Следствие от теорема 4.4.□

Теорема 4.6 *Ако определим вероятност върху* $\mathfrak{B}$ *по формулата* $Q(B) = P(\xi^{-1}(B))$, *то тройката* $(R^1, \mathfrak{B}, Q)$ *става вероятностно пространство.*

Доказателство: За да докажем, че тройката $(R^1, \mathfrak{B}, Q)$ става вероятностно пространство е достатъчно да проверим аксиомите за вероятност. □

Определение 4.7 *Функцията* $f(\cdot) : R^n \longrightarrow R^m$ *се нарича Борелова, ако пълният праобраз на Борелово множество е Борелово множество.*

Поради теорема 4.4 проверката за това дали една функция е Борелова е тривиална – достатъчно е да проверим дали пълният праобраз на интервал е Борелово множество.

Теорема 4.7 *Борелова функция от (векторна) сл.в. е (векторна) сл.в.*

Доказателство: Достатъчно е да разгледаме съответствието $x = f(\xi(\omega))$. Пълният му праобраз може да се запише като композицията $\xi^{-1}(f^{-1}(.))$. □
Доказателство: (Теорема 4.3)
Така за да я докажем е достатъчно да проверим, че функциите $x+y, xy$ са Борелови. Да разгледаме, например, $x+y$ и множеството: $A = \{x+y < c\} \subset R^2$. Да разгледаме изброимото обединение на двумерни интервали:

$$B = \cup_i(\{x : x < x_i\} \cap \{y : y < y_i\}). \tag{4.1}$$

Тук с (x_i, y_i) сме означили всички двойки рационални числа такива, че $x_i + y_i < c$. Очевидно $A \supset B$. Обратно, за всяка двойка $(x, y) : x + y < c$ интервалът $(x, c - y)$ е непразен и в него съществува рационално число – да го означим с a. Имаме $a+y < c$. Сега интервалът $(y, c - a)$ е непразен и в него съществува рационално число – да го означим с b. При това са изпълнени и трите неравенства $a + b < c$, $x < a$, $y < b$. □

Теорема 4.8 *Съществува взаимно еднозначно съответствие между функциите на разпределение и възможните вероятности, определени на* $\mathfrak{B}$ *в* R^1.

Доказателство: Следва от това, че всяка вероятност се определя от своите стойности върху пораждащите Бореловата алгебра интервали.□

Тема 5

Числови характеристики на сл.в.

В миналата лекция се запознахме с разпределенията на сл.в. Сега ще разгледаме някои числови характеристики, които се пресмятат само по разпределението. Първо ще определим най - популярната числова характеристика математическото очакване (м.о.) и ще изведем основните му свойства.

5.1 Математическо очакване.

Определение 5.1 *Математическо очакване* $\mathbf{E}\,\xi$ *на простата сл.в.* ξ *приемаща стойности* $x_1, x_2, \dots, x_n$ *върху събитията от пълната група* $(H_1, H_2, \dots, H_n)$ *определяме като* $\mathbf{E}\,\xi = \sum_{k=1}^{n} x_k \, \mathsf{P}(H_k)$.

От това определение се вижда веднага, че числото $\mathbf{E}\,\xi$ зависи само от стойностите на ξ и вероятностите, с които те се приемат, а не зависи от това за кои точно елементарни събития и каква пълна група това става. Т.е. то не зависи от това в какво вероятностно пространство е реализирана сл.в.

Теорема 5.1 *За прости сл.в. математическото очакване притежава следните свойства:*

1. *монотонност - Ако* $\xi < \eta$, *то* $\mathbf{E}\,\xi < \mathbf{E}\,\eta$;
2. *линейност -* $\mathbf{E}\,(\alpha\xi + \beta\eta) = \alpha\mathbf{E}\,\xi + \beta\mathbf{E}\,\eta$;
3. *мултипликативност - Ако* $\xi \perp\!\!\!\perp \eta$, *то* $\mathbf{E}\,\xi\eta = \mathbf{E}\,\xi\mathbf{E}\,\eta$;

Доказателство: Ще използуваме означенията на теорема 4.1. Нека пълната група събития $(H_1, H_2, \dots, H_n)$ съответствува на сл.в. ξ, пълната група събития $(G_1, G_2, \dots, G_m)$ на сл.в. η, а техните стойности са съответно $\{x_1, x_2, \dots x_n\}$ и $\{y_1, y_2, \dots y_m\}$. Тогава

$$\mathbf{E}\,\xi\eta = \sum_{i,j} x_i y_j \, \mathsf{P}(H_i G_j) = \sum_{i,j} x_i y_j \, \mathsf{P}(H_i) \, \mathsf{P}(G_j) = \sum_{i} x_i \, \mathsf{P}(H_i) \sum_{j} y_j \, \mathsf{P}(G_j) = \mathbf{E}\,\xi\mathbf{E}\,\eta. \square$$

Тези свойства на математическото очакване и особено неговата монотонност позволяват то лесно да се разпространи за произволни неотрицателни сл.в. Може обаче да се окаже, че то е безкрайно.

Математическото очакване на непрекъснатата сл.в. ξ се пресмята като интеграла: $\mathbf{E}\,\xi = \int xf(x)dx$, а това на дискретна като сумата $\mathbf{E}\,\xi = \sum_i x_i p_i$, когато това е възможно. Тези две формули могат да се обединят в една с помощта на интеграла на Лебег-Стилтес: $\mathbf{E}\,\xi = \int xdF(x)$. Той съществува (и е краен), когато е краен интегралът $\mathbf{E}\,\xi = \int |x|dF(x) < \infty$.

Определение 5.2 *Момент от ред k на сл.в.ξ наричаме следната числова величина (когато съществува): обикновен -* $\mathbf{E}\,\xi^k$ *, абсолютен -* $\mathbf{E}\,|\xi|^k$*, централен -* $\mathbf{E}\,(\xi - \mathbf{E}\,\xi)^k$*.*

5.2 Числови характеристики.

Локация

М.о. е най - важната характеристика за положението на стойностите на сл.в. върху числовата ос. За съжаление, както видяхме, тя е определена не за всички сл.в..

Определение 5.3 *Медиана се определя като решение на уравнението: $F(\mu) = \frac{1}{2}$.*

Тя описва положението на средата на разпределението върху числовата ос. Когато решението не е единствено, се взима средата на интервала от решения. В много случаи се използва и положението на други характерни точки от разпределението.

Определение 5.4 *Квантил с ниво α на дадено разпределение F се определя като решение на уравнението:*

$$F(q_\alpha) = \alpha.$$

В статистиката квантилите на вероятности кратни на $1/4$ се наричат квартили, тези - на $1/10$ - децили, а на $1/100$ - процентили. Така $\mu = q_{1/2}$ е втори квартил, пети децил, петдесети процентил.

Определение 5.5 *Мода се определя като най - вероятното число за дискретни сл.в., а за непрекъснати —като координатата на максимума на плътността.*

Естествено, разпределенията могат и да не притежават единствена мода. За симетрични разпределения, очевидно трите характеристики: мода (ако има такава), медиана и м.о. съвпадат.

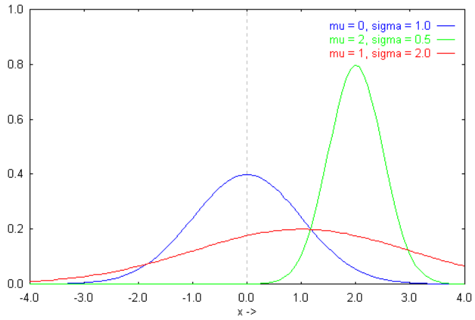
Мащаб

Определение 5.6 *Дисперсия на сл.в. ξ се определя като числото* $\mathbf{D}\,\xi = \mathbf{E}\,(\xi - \mathbf{E}\,\xi)^2$*. Дисперсията може да се окаже и безкрайна.*

Дисперсията е най - важната характеристика на разсейване на стойностите на сл.в. За дискретни и непрекъснати разпределения тя се пресмята по формулите:

$$\mathbf{D}\,\xi = \int (x - \mathbf{E}\,\xi)^2 f(x)dx, \quad \mathbf{D}\,\xi = \sum_i (x_i - \mathbf{E}\,\xi)^2 p_i. \tag{5.1}$$

Фактически вместо дисперсията, както в числовите, така и в аналитичните сметки, се използува *стандартно отклонение.* Това е:



$$\sigma(\xi) = (\mathbf{D}\,\xi)^{1/2} = (\mathbf{E}\,(\xi - \mathbf{E}\,\xi)^2)^{1/2}. \tag{5.2}$$

Тази характеристика се мери в същите физически единици, като ξ и може да бъде съответно интерпретирана. Тук са показани плътности от нормалното семейство с различни стандартни отклонения.

Фиг. 5.1: Различни дисперсии

Колкото по - малка е дисперсията или стандартното отклонение, толкова по - сгъстени са стойностите и по - вероятни са те в центъра на разпределението. За това, когато искаме да се отървем от размерността, например за да сравним разпределенията на две различни сл.в., прилагаме т.н. *центриране и нормиране.* Вместо величината ξ разглеждаме центрираната и нормирана величина

$$\tilde{\xi} = \frac{\xi - \mathbf{E}\,\xi}{\sigma(\xi)}. \tag{5.3}$$

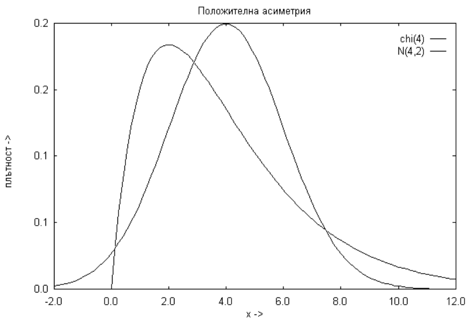
Когато дисперсията е безкрайна за "определяне" на мащаба се използува т.н. интерквартилен размах.

Определение 5.7 *Наричаме интерквартилен размах r разликата между третия и първия квартили:* $r = q_{3/4} - q_{1/4}$.

Форма

Следните две характеристики на разпределенията не зависят от мерните единици, с които са отчитани съответните сл.в., както и от условните начала на скалите. С други думи, те са безразмерни. Те отразяват различията във формата на разпределенията, но не зависят от мащаба и локацията.

Определение 5.8 *Ще наричаме асиметрия на ξ числото (когато съществува):*



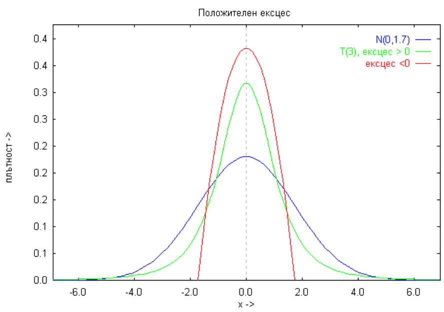
Фиг. 5.2: Положителна асиметрия

$$Ass(\xi) = \frac{\mathbf{E}\,(\xi - \mathbf{E}\,\xi)^3}{\sigma^3(\xi)} = \mathbf{E}\,\tilde{\xi}^3. \tag{5.4}$$

На тази фигура е дадено сравнение на положително асиметрична плътност с плътността на стандартния нормален закон, която е симетрична и има асиметрия 0. Положителната асиметрия се характеризира с "по - тежка" дясна опашка на разпределението.

При асиметричните разпределения се променят обикновено и взаимните положения на модата, медианата и математическото очакване. За разпределения с положителна асиметрия те се нареждат в посочения ред, а за тези с отрицателна — в обратния. Това правило, разбира се, е верно само за унимодални разпределения с проста аналитична форма на плътността. Вижте също семейството на Бета - разпределенията 11.3.

Определение 5.9 *Ще наричаме ексцес на ξ числото (когато съществува):*



Фиг. 5.3: Положителен ексцес

$$Ex(\xi) = \frac{\mathbf{E}\,(\xi - \mathbf{E}\,\xi)^4}{\sigma^4(\xi)} - 3 = \mathbf{E}\,\tilde{\xi}^4 - 3. \tag{5.5}$$

Тук е представено разпределение с положителен ексцес. То има по - дълги и тежки опашки от нормалното (с ексцес 0). Разпределенията с отрицателен ексцес може изобщо да нямат опашки — например, такова е равномерното в краен интервал.

Изобщо казано, двата параметъра асиметрия и ексцес дават достатъчно пълна картина за формата на разпределението, когато то е унимодално и гладко. Всъщност такива разпределения обикновено принадлежат на семейство, описвано с няколко параметъра.

Особено място сред тези фамилии заема двупараметричната фамилия на нормалните разпределения (вж. определение 7.1), с която скоро ще се запознаем. За всички нормални разпределения и асиметрията и ексцеса са равни на 0. Така тези две характеристики за форма придобиват смисъл на характеристики за *отклонение от нормалността.*

5.3 Неравенства

Ще започнем с някои прости свойства на дисперсията.

$$\mathbf{D}\,\xi \geq 0; \tag{5.6}$$

$$\mathbf{D}\,\xi = 0 \qquad \Leftrightarrow \qquad \mathsf{P}(\xi = const) = 0; \tag{5.7}$$

$$\mathbf{D}\,\xi = \quad \mathbf{E}\,\xi^2 - (\mathbf{E}\,\xi)^2; \tag{5.8}$$

$$\mathbf{D}\,\alpha\xi = \alpha^2\mathbf{D}\,\xi; \quad \mathbf{D}\,(\xi + c) = D\xi$$

$$\mathbf{D}\,(\xi + \eta) = D\xi + D\eta, \qquad \text{ако} \qquad \xi \perp\!\!\!\perp \eta.$$

Доказателство:

- $(\xi - \mathbf{E}\,\xi)^2 \geq 0 \Rightarrow \mathbf{D}\,\xi = \mathbf{E}\,(\xi - \mathbf{E}\,\xi)^2 \geq 0$
- Следва от неравенството на Чебишов.
- $\mathbf{E}\,(\xi - \mathbf{E}\,\xi)^2 = \mathbf{E}\,\xi^2 - 2\mathbf{E}\,\xi\mathbf{E}\,\xi + (\mathbf{E}\,\xi)^2$.
- $\mathbf{D}\,(\xi + \eta) = \mathbf{D}\,\xi + 2\mathbf{E}\,(\xi - \mathbf{E}\,\xi)(\eta - \mathbf{E}\,\eta) + \mathbf{D}\,\eta$. $\square$

Неравенство на Чебишов

Следното знаменито неравенство се използува много често. За всички $r > 0$ и $\epsilon > 0$ е в сила неравенството:

$$\mathsf{P}(|\xi| > \epsilon) < \frac{\mathbf{E}\,|\xi|^r}{\epsilon^r} \tag{5.9}$$

Доказателство: Да означим с I_A сл.в. (индикатор на събитието A)

$$\mathbf{E}\,|\xi|^r = \mathbf{E}\,|\xi|^r I_{|\xi|>\epsilon} + \mathbf{E}\,|\xi|^r I_{|\xi|\leq\epsilon} \geq \epsilon^r \mathbf{E}\,I_{|\xi|>\epsilon} = \epsilon^r\,\mathsf{P}(|\xi| > \epsilon).\square$$

От това неравенство следва $\Rightarrow$ на твърдение (5.7):

$$0 = \mathsf{P}(|\xi - \mathbf{E}\,\xi| > 1/n) \Rightarrow \mathsf{P}(\xi = \mathbf{E}\,\xi) = 1.$$

Неравенства за моментите

От уравненията (5.6) и (5.8) веднага следва следното неравенство за моментите:

$$(\mathbf{E}\,\xi)^2 \leq \mathbf{E}\,\xi^2, \tag{5.10}$$

което е частен случай на по - общото:

$$(\mathbf{E}\,|\xi|^r)^{1/r} \leq (\mathbf{E}\,|\xi|^k)^{1/k}, \quad \text{при всички } r < k. \tag{5.11}$$

От това неравенство следва, че ако съществува (т.е. е краен) моментът от ред k на една сл.в., то съществуват моментите от всеки ред $r < k$.

5.4 *Линейни пространства сл.в.

М.о. може, разбира се, да се определи и абстрактно. Всяка неотрицателна сл.в. ξ^+ може да се представи като граница на монотонно нарастваща редица от прости сл.в. $\xi^+ = \lim_n \xi_n =\uparrow \xi_n$. Следователно, винаги съществува границата $\mathbf{E}\,\xi = \lim \mathbf{E}\,\xi_n$, възможно равна на ∞. Всяка сл.в. ξ може да се представи като сума $\xi = \xi^+ + \xi^-$. Така, ако и двете м.о. са крайни, можем спокойно да определим $\mathbf{E}\,\xi = \mathbf{E}\,\xi^+ + \mathbf{E}\,\xi^-$.

Класовете еквивалентни с вероятност 1 сл.в. (к.е.сл.в.), зададени на дадено фиксирано вероятностно пространство $(\Omega, \mathfrak{A}, \mathsf{P})$ образуват линейно пространство. Нека го означим с $\mathfrak{M} = \mathfrak{M}(\Omega, \mathfrak{A}, \mathsf{P})$. Да разгледаме подмножествата $\mathbf{L}^r \subset \mathfrak{M}, (1 \le r \le \infty)$ от сл.в. с краен r-ти абсолютен момент. Тъй като в тях работи същото отношение на еквивалентност, к.е.сл.в. от съответните множества образуват пълни линейни нормирани пространства, ако за норма във всяко от тях служи: $||\xi||_r = (\mathbf{E}\,|\xi|^r)^{1/r}$. При това е от неравенството (5.11) следва включването:

$$\mathbf{L}^\infty \subset \mathbf{L}^r \subset \mathbf{L}^p \subset \mathbf{L}^1 \subset \mathfrak{M}, \quad 1 < p < r < \infty.$$

Особен интерес представлява пространството $\mathbf{L}^2$ от сл.в. с крайна дисперсия. То е Хилбертово и скаларното произведение в него е $(\xi, \eta) = \mathbf{E}\,\xi\eta$. Така центрираните сл.в., ако са независими, са ортогонални в $\mathbf{L}^2$. Обратното твърдение не винаги е вярно.

Нека $\mathfrak{F} \subset \mathfrak{A}$ и разделянето, което й съответствува е γ. Да разгледаме подмножеството на $\mathbf{L}^2$ от сл.в., такива, че $\mathfrak{A}_\xi \subset \mathfrak{F}$. Те очевидно образуват линейно подпространство на $\mathbf{L}^2$. Проекторът върху това подпространство да означим с $\mathbf{E}_\gamma$. Ще го наричаме условно м.о. при условие разделянето γ, или σ-алгебрата, която му съответствува. В частност, ако означим с ϵ разделянето, което съответствува на $\mathfrak{A}$ и $\nu = \{\emptyset, \Omega\}$, то $\mathbf{E}_\epsilon = I$, $\mathbf{E}_\nu = \mathbf{E}$. Условното м.о. на всяка сл.в. ξ при условие крайното разделяне γ е проста сл.в. с константни стойности върху елементите му. Условните м.о. удовлетворяват всички свойства на ортогоналните проектори в Хилбертово пространство. Ако $\gamma \perp\!\!\!\perp \delta$, то $\mathbf{E}_\gamma$ и $\mathbf{E}_\delta$ комутират.

Тема 6

Дискретни разпределения

В тази лекция си поставяме за цел да обобщим и разширим понятията си за:

- целочислена сл.в. и нейното разпределение;
- да въведем някои най-срещани разпределения;
- ще покажем как се използуват някои от средствата на анализа за облекчаване на пресмятанията на разпределенията и техните количествени характеристики - моментите.
- да дадем представа за връзката между някои вече определени дискретни разпределения;

6.1 Пораждащи функции

Определение 6.1 *Сл.в. приемаща за стойности натуралните числа наричаме целочислена.*

Целочислените сл.в. са особено удобни за моделиране на реални явления като брой успехи или други бройки. За пресмятане на моментите на целочислени сл.в. служат т.н. пораждащи моментите функции.

Определение 6.2 *Пораждащата функция на целочислена сл.в. ξ се задава с формулата:*

$$p(s) \stackrel{def}{=} \mathbf{E}\, s^{\xi} = \sum_{i=0}^{\infty} s^i\, \mathsf{P}(\xi_i = i) = \sum_{i=0}^{\infty} s^i p_i. \qquad (6.1)$$

Пораждащата функция е удобна защото съществува винаги (при достатъчно малко s, например, когато $s \leq 1$) и е диференцируема при $0 \leq s < 1$. Тя притежава още следните свойства:

- $p(1) = \sum_{i=0}^{\infty} p_i = 1$;
- $p(0) = \mathsf{P}(\xi = 0) = p_0$;

- $p'(1) = \mathbf{E}\,\xi = \sum_{i=0}^{\infty} p_i.$, когато съществува;
- $p''(1) = \mathbf{E}\,\xi(\xi-1) = \mathbf{E}\,\xi^2 - \mathbf{E}\,\xi$, когато съществува.
- Когато $\xi \perp\!\!\!\perp \eta$, $p_{\xi+\eta}(s) = p_\xi(s)p_\eta(s)$.

Теорема 6.1 *Между разпределенията на целочислени сл.в. и пораждащите функции съществува взаимно еднозначно съответствие.*

Доказателство: Следва от формулата: $p^{(n)}(0) = n!p_n$. □

Биномно разпределение

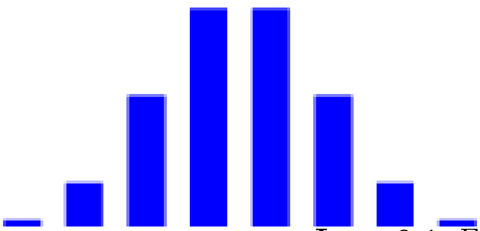
Определение 6.3 *Редица от независими еднакво разпределени случайни величини* $\{\xi_i,\ i = 1, 2, ...\}$, *всяка от които приема две стойности: 1 и 0 с вероятности (съответно)* p *и* $q = 1 - p$, *наричаме схема на Бернули.*

Да разгледаме сумата η_n на n сл.в. от схемата на Бернули. Това е целочислена случайна величина, приемаща стойности от 0 до n. Ние я интерпретираме като *Брой успехи от* n *опита с постоянна вероятност* p *за успех във всеки опит.* Разпределението на тази сл.в. наричаме *биномно*. Вероятността тази сл.в. да приеме стойност k наричаме биномна и означаваме с $b(n, k, p)$.

Теорема 6.2 *Биномните вероятности се пресмятат по формулата:*

$$b(n, k, p) = \binom{n}{k} p^k q^{n-k} \tag{6.2}$$

$$k = 0, 1, \ldots, n$$



Фиг. 6.1: Биномно разпределение $n = 7$

Доказателство: Първо да пресметнем вероятността на събитието

$$W_{\epsilon_1,\epsilon_2,\ldots,\epsilon_n} = \bigcap_{i=1}^{n} \{\xi_i = \epsilon_i\},$$

където $\epsilon_j \in \{0, 1\}, \quad j = 1, 2, \ldots, n$. Тъй като сл.в. са независими и $P(\xi = 1) = p$, получаваме

$$\mathsf{P}(W_{\epsilon_1,\epsilon_2,\ldots,\epsilon_n}) = p^{\sum_{i=1}^{n} \epsilon_i} q^{n-\sum_{i=1}^{n} \epsilon_i} \tag{6.3}$$

Ако означим $\eta_n = \sum_{i=1}^{n} \xi_i$ и $k = \sum_{i=1}^{n} \epsilon_i$, ще получим

$$P(\eta_n = k) = \sum_{\epsilon_1+\epsilon_2+\ldots+\epsilon_n=k} p^k q^{n-k} = p^k q^{n-k} \sum_{\epsilon_1+\epsilon_2+\ldots+\epsilon_n=k} 1.$$

Но от тук следва търсената формула. □

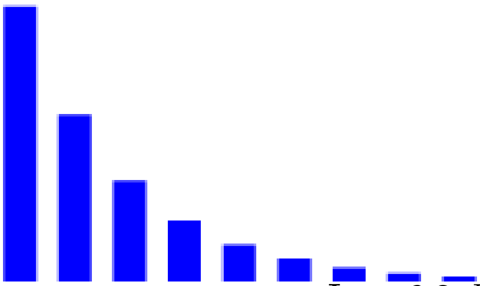
Пораждащата функция на биномното разпределение се пресмята лесно, защото биномната сл.в. η е сума на еднакво разпределени независими сл.в.

$$\mathbf{E}\, s^{\eta} = \mathbf{E} \prod_{i=1}^{n} s^{\xi_i} = (\mathbf{E}\, s^{\xi_1})^n = (ps+q)^n.$$

Геометрично разпределение

Нека разгледаме в ситуацията на независими опити (схема на Бернули) сл.в. ξ — брой успешни опити до първи неуспех.

Определение 6.4 *Казваме, че целочислената сл.в. ξ има геометрично разпределение, ако:*



$$\mathsf{P}(\xi = m) = p^m q,$$

$$m = 0, 1, 2, \ldots.$$

Фиг. 6.2: Геометрично разпределение

Математическото очакване и дисперсията на това разпределение се пресмятат лесно:

$$\mathbf{E}\,\xi = q\sum_{k=0}^{\infty} kp^k = qp\sum_{k=0}^{\infty} kp^{k-1} = qp\frac{d}{dp}(\frac{1}{1-p}) = \frac{p}{q},$$

$$\mathbf{D}\,\xi = \mathbf{E}\,\xi(\xi-1) + \mathbf{E}\,\xi - (\mathbf{E}\,\xi)^2 = q\sum_{k=0}^{\infty} k(k-1)p^k + \frac{p}{q} - (\frac{p}{q})^2 =$$

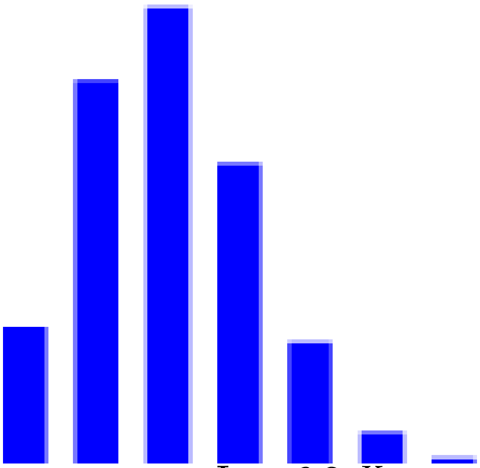
$$qp^2\frac{d^2}{dp^2}(\frac{1}{1-p}) + \frac{p}{q} - (\frac{p}{q})^2 = (\frac{p}{q})^2 + \frac{p}{q}.$$

Хипергеометрично разпределение

Да разгледаме една задача от статистическия качествен контрол. Нека е дадена партида съдържаща N изделия, от които M са дефектни. Правим случайна извадка от $n < N$ изделия. Пита се каква е вероятността точно m от тях да са дефектни.

Оказва се, че разпределението на сл.в. брой дефектни е следното:

Определение 6.5 *Казваме, че целочислената сл.в. ξ има хипергеометрично разпределение, ако:*



$$\mathsf{P}(\xi = m) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}},$$

$m = 0, 1, \ldots, n.$

Фиг. 6.3: Хипергеометрично $N = 100, M = 20, n = 10$

Тази формула се извежда лесно. Броят на всички възможни извадки без връщане очевидно е

$$\binom{N}{n}$$

(смятаме ги за равновероятни). "Благоприятните", тези които съдържат точно m дефектни детайла, могат да се получат чрез комбиниране на извадка от M на m дефектни и извадка от $N - M$ на $n - m$ изправни. Тъй като извадките от здрави и дефектни детайли се комбинират без ограничения, общият брой на "благоприятните" извадки става

$$\binom{M}{m}\binom{N-M}{n-m}.$$

Математическото очакване и дисперсията на това разпределение също се пресмятат лесно:

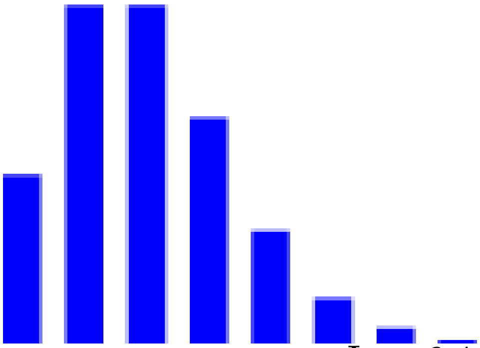
$$\mathbf{E}\,\xi = np, \quad p = \frac{M}{N}, \quad q = 1 - p$$

$$\mathbf{D}\,\xi = npq\frac{N-1}{N-n}.$$

От тези формули се вижда, че това разпределение клони към биномното при голям брой N на детайлите в партидата.

Разпределение на Поасон

Поасоновото разпределение се определя лесно като граница на биномни разпределения, когато $n \to \infty$ така че $np \to \lambda > 0$. Сл.в. може да приема всякакви целочислени стойности:



$$\mathsf{P}(\xi = k) = e^{-\lambda}\frac{\lambda^k}{k!}. \tag{6.4}$$

Фиг. 6.4: Поасоново разпределение

То е особено подходящо за моделиране на броя на случайни редки събития – брой частици на единица обем, брой радиоактивни разпадания за единица време и т.н. Средното и дисперсията му съвпадат: $\mathbf{E}\,\xi = \mathbf{D}\,\xi = \lambda$. Това най-лесно се вижда от пораждащата функция на Поасоновото разпределение, която се пресмята директно:

$$\mathbf{E}\,s^{\eta} = e^{-\lambda}\sum_{k=1}^{\infty}\frac{(\lambda s)^k}{k!} = e^{\lambda(s-1)}.$$

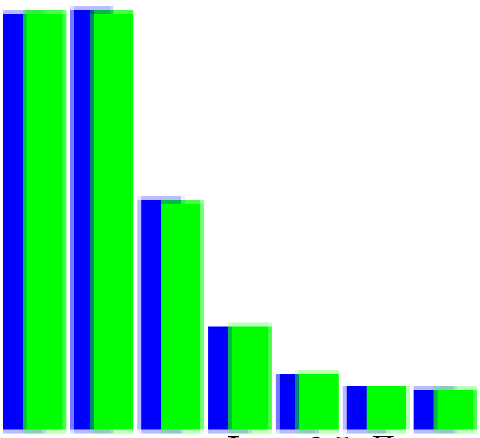
6.2 Близости между дискретни разпределения

Апроксимациите играят съществена роля в теория на вероятностите. Много от получените формули са тежки за пресмятане и се налага те да бъдат замествани с приближени.

Поасоново и биномно

Тук ще разгледаме едно много полезно и старо приближение на биномната вероятност при малки k.

Теорема 6.3 *(Теорема на Поасон) Ако в схемата на Бернули $np_n \to \lambda$, то*



$$b(n,k,p_n) \underset{n\to\infty}{\overset{np_n=\lambda}{\longrightarrow}} \frac{\lambda^k}{k!}e^{-\lambda}.$$

Фиг. 6.5: Поасонова апроксимация $n = 100, p = 0.01$

Доказателство: Да означим $\lambda = np$. Можем да запишем биномната вероятност във формата:

$$b(n,k,p) = \frac{n(n-1)\dots(n-k+1)}{k!}p^k(1-p)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}\epsilon(k,n,\lambda),$$

където

$$\epsilon(k,n,\lambda) = \prod_{i=0}^{k-1}(1+\frac{i}{n}) \quad \times \quad (1-\frac{\lambda}{n})^k \quad \times \quad e^{\lambda}(1-\frac{\lambda}{n})^n. \tag{6.5}$$

Всеки от трите съмножителя на дясната страна клони към 1 при фиксирано k и $np_n \to \lambda$. $\square$

Още по-лесно се доказва тази теорема с помощта на пораждащи функции. Наистина, достатъчно е да покажем, че

$$(ps+q)^n = (1+\frac{(s-1)\lambda}{n})^n \to e^{\lambda(s-1)}.$$

Възниква въпросът колко точно е приближението на Поасон. Ще приведем едно уточнение взето от (А.А.Боровков 1972).

Теорема 6.4 *(Уточнение) Нека* $k \le 1+n/4$ *и* $p < 1/4$. *Тогава*

$$\frac{k}{n}\lambda + \frac{7k(1-k)-8\lambda^2}{12n} \le \ln\epsilon(k,n,\lambda) \le \frac{k}{n}\lambda + \frac{k(1-k)}{n} \tag{6.6}$$

Доказателство: Да логаритмуваме израза в (6.5)

$$\ln\epsilon(k,n,\lambda) = (\sum_{i=0}^{k}\ln(1-\frac{i}{n}) + ((n-k)\ln(1-\frac{\lambda}{n})+\lambda). \tag{6.7}$$

Да се възползуваме от неравенствата за $\ln(1-x)$ при $0 \le x \le 1/4$:

$$-\frac{7x}{6} \le -x(1+\frac{x}{2(1-x)}) \le \ln(1-x) \le -x,$$

$$\ln(1-x) - x \ge -\frac{x^2}{2(1-x)} \ge -\frac{2x^2}{3}.$$

Сега лесно получаваме:

$$\ln\epsilon(k,n,\lambda) \le -\sum_{i=1}^{k-1}\frac{i}{n} - (n-k)p + np = \frac{k}{n}\lambda + \frac{k(1-k)}{n}$$

$$\ln\epsilon(k,n,\lambda) \ge -\frac{7}{6}\sum_{i=1}^{k-1}\frac{i}{n} + kp - n\frac{2p^2}{3} = \frac{k}{n}\lambda + \frac{7k(1-k)-8\lambda^2}{12n}.\square$$

Следователно, относителната грешка при използването на формулата на Поасон за приближение на биномната вероятност не превъзхожда

$$\frac{k}{n}\lambda + \frac{k(1-k)}{n}. \tag{6.8}$$

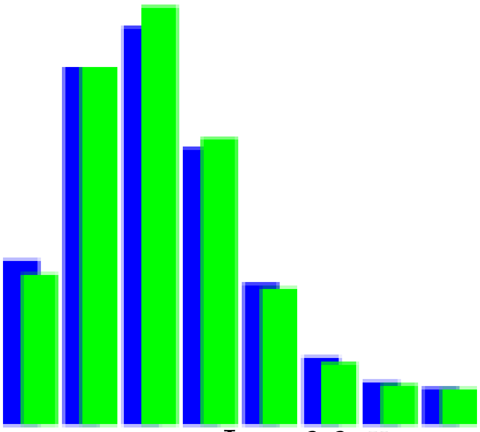
От същите разсъждения следва, че при необходимост може да бъде получено и още по-добро приближение по формулата:

$$b(n,k,p) \cong \frac{\lambda^k}{k!} e^{-\lambda + \frac{k}{n}\lambda + \frac{k(1-k)}{n}}$$

Хипергеометрично и биномно

Вече видяхме, че моментите (м.о. и дисперсията) на хипергеометричното разпределение апроксимират при $N \to \infty$ тези на биномното разпределение. Тук ще покажем, че това е верно и за вероятностите.

Теорема 6.5 *Нека $N \to \infty$ и $M = pN$. Тогава*



$$\frac{\binom{M}{m}\binom{N-M}{M-m}}{\binom{N}{n}} \underset{N\to\infty}{\longrightarrow} b(n,m,p) \tag{6.9}$$

Фиг. 6.6: Хипергеометрично и биномно (N=100,p=0.2)

Доказателство: Да прегрупираме лявата част:

$$\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} = \binom{n}{m} \overbrace{\frac{M}{N}\frac{M-1}{N-1}\cdots\frac{M-m+1}{N-m+1}}$$
$$\underbrace{\frac{N-M}{N-m}\frac{N-M}{N-m-1}\cdots\frac{N-M-n+m+1}{N-n}}$$
$$\longrightarrow \binom{n}{m}\overbrace{p^m}\underbrace{(1-p)^{n-m}}.\square$$

* Случайна сума сл.в.

Нека разгледаме редицата от целочислени независими еднакво разпределени сл.в. $\{\xi_i, i = 1,2,\ldots\}$. Нека освен това ни е зададена независимата от тях сл.в. ν. Нека разгледаме случайната величина η случайна сума от случайни събираеми:

$$\eta = \sum_{i=0}^{\nu} \xi_i$$

и се запитаме какво е нейното разпределение и други числови характеристики. Тази задача се решава изключително елегантно с апарата на пораждащите функции.

Теорема 6.6 *Нека сл.в.* $\{\xi_i, i = 1, 2, \ldots\}$ *са независими с пораждаща функция* $p(s)$. *Нека освен това независимата от тях сл.в.* ν *има пораждаща функция* $q(s)$. *Тогава пораждащата функция на* η *е равна на* $p_\eta(s) = q(p(s))$.

Доказателство: Наистина, да приложим формулата за пълната вероятност и условното м.о.:

$$p_\eta(s) = \mathbf{E}\, s^{\sum_{i=0}^{\nu} \xi_i} = \sum_{k=0}^{\infty} \mathsf{P}(\nu = k)\mathbf{E}\,(s^{\sum_{i=0}^{k} \xi_i} | \nu = i)) =$$

$$\sum_{k=0}^{\infty} \mathsf{P}(\nu = k)\mathbf{E}\,(s^{\sum_{i=0}^{k} \xi_i})) = \sum_{k=0}^{\infty} \mathsf{P}(\nu = k) p(s)^k = q(p(s)).\Box$$

Тема 7

Нормално разпределение

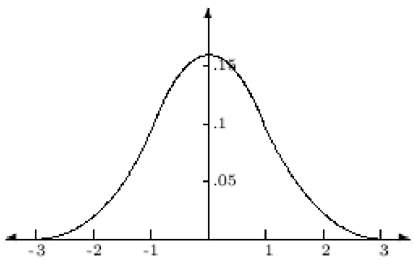
В тази лекция си поставяме следните цели:

- да определим знаменитото нормално разпределение;
- да покажем връзката между непрекъснати и дискретни разпределения;
- на примера на най-простата статистическа задача да илюстрираме начина, по който се строят статистическите разсъждения.

7.1 Нормално разпределение

Определение 7.1 *Казваме, че сл.в. с непрекъснато разпределение е нормално разпределена $N(\mu, \sigma)$, ако нейната плътност има вида:*

$$f(x, \mu, \sigma) = \frac{1}{(2\pi)^{1/2}\sigma} e^{-(x-\mu)^2/2\sigma^2}. \tag{7.1}$$



Фиг. 7.1: Стандартно нормално

Математическото очакване на това разпределение е μ, а дисперсията му е σ^2. *Стандартно* нормално разпределение се нарича разпределението $N(0, 1)$, неговата плътност означаваме с $\phi(x)$.

Нормалното разпределение има голямо значение в теория на вероятностите и математическата статистика, което се дължи на твърдението, известно като Централна Гранична Теорема. То гласи, че разпределението на сума от голям брой независими, еднакво разпределени случайни величини клони към нормално разпределение.

7.2 Теореми на Муавър-Лаплас

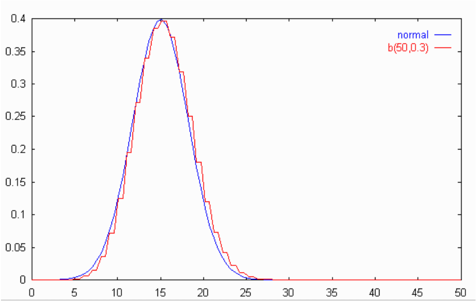
Тук ще се опитаме да покажем как е възникнала формулата за плътността на нормалното разпределение.

7.2.1 Локална теорема

Нека устремим към безкрайност броят на опитите в схемата на Бернули. Да означим

$$x = \frac{k - np}{(npq)^{1/2}}$$

и поискаме това число да остане ”почти постоянно” при $n \to \infty$. Смисълът на това число е ясен - това е центрираната и нормирана стойност на сл.в. брой на успехи. Ясно е, че тогава (при фиксирана вероятност за успех p) също и $k \to \infty$. Да означим $\sigma = (npq)^{1/2}$.



Теорема 7.1 *(Локална теорема на Муавър-Лаплас) Нека p, x са фиксирани. Нека $n \to \infty$ Нека означим $\sigma_n^2 = npq$, $k_n = np + \sigma_n x$. Тогава равномерно в интервала $-\infty < a < x < b < \infty$ е изпълнено съотношението:*

$$\sigma_n b(n, k_n, p) \longrightarrow \phi(x)$$

Фиг. 7.2: Нормална апроксимация на биномно разпределение

Доказателство: За простота ще изпускаме индекса от означенията σ_n, k_n в доказателството. Да логаритмуваме биномната вероятност от лявата страна

$$\ln b(n, k_n, p) = \ln n! - \ln k_n! - \ln(n - k_n)! + k_n \ln p + (n - k) \ln q. \tag{7.2}$$

Ще използуваме представянето на Стирлинг на $\ln n!$:

$$\ln n! = n \ln n + \frac{1}{2} \ln(2\pi n) - n + \alpha(n), \tag{7.3}$$

където $\alpha(n) = O(1/n)$. Да означим $m_n = n - k_n$. Тъй като $m_n, k_n \to \infty$, то от (7.2) и (7.3) следва

$$\begin{aligned} \ln b(n, k, p) - \frac{1}{2} \ln 2\pi \\ = n \ln n - k \ln k - m \ln m + k \ln p + m \ln q - \frac{1}{2} \ln \frac{km}{n} + \beta_n = \\ = -(np + \sigma x) \ln(1 + \frac{xq}{\sigma}) - (nq - \sigma x) \ln(1 - \frac{xp}{\sigma}) - \frac{1}{2} \ln \frac{km}{n} + \beta_n, \end{aligned} \tag{7.4}$$

където $\beta_n = \alpha(n) - \alpha(k_n) - \alpha(m_n) = O(1/n)$, $k = np + \sigma x$ и $m = n - k = nq - \sigma x$. Тук ще се отклоним малко да разгледаме дробта:

$$\frac{km}{n\sigma^2} = \frac{(np + \sigma x)(nq - \sigma x)}{n\sigma^2} = (1 + \frac{(q - p)x}{\sigma} - \frac{pqx^2}{\sigma^2}).$$

Така лесно ще можем да получим израз за третия логаритъм в (7.4) чрез σ.

$$\frac{1}{2} \ln \frac{mk}{n} = \frac{1}{2} \ln \sigma + O(\frac{1}{\sigma}).$$

За първите два логаритъма ще използуваме разложението $\ln(1+x) = x - x^2/2 + O(x^3)$. Заместваме, съкращаваме и получаваме окончателно:

$$\begin{aligned} \ln\sigma + \ln b(n, k_n, p) &= \frac{1}{2}\ln 2\pi + \\ -(np+\sigma x)(\frac{xq}{\sigma} - \frac{x^2q^2}{2\sigma^2})) - (nq - \sigma x)(-\frac{xp}{\sigma} - \frac{x^2p^2}{2\sigma^2})) + \gamma_n \\ &= \frac{1}{2}\ln 2\pi - \frac{x^2}{2} + \gamma_n. \end{aligned}$$

където $\gamma_n = O(\sigma^{-1}) = O(n^{-1/2})$. □

И тази теорема може да се уточни, както теоремата на Поасон, но тук няма да се спираме на това. Тя ни дава възможност да пресмятаме лесно конкретни биномни вероятности. При големи стойности на n това са твърде малки числа.

7.2.2 Интегрална теорема

За да можем да пресмятаме суми от Биномни вероятности си служим със следната интегрална теорема на Муавър - Лаплас.

Теорема 7.2 *Да означим с ξ биномна сл.в. с разпределение $b(n,k,p)$. Тогава при фиксирани b, a, p имаме*

$$\mathsf{P}(b < \frac{\xi - np}{(npq)^{1/2}} < a) = \sum_{k=[np+b(npq)^{1/2}]}^{[np+a(npq)^{1/2}]} b(n,k,p) \underset{n\to\infty}{\longrightarrow} \int_b^a \phi(y)dy.$$

Доказателство: За краен интервал (b, a) тя лесно следва от локалната теорема, а за произволен – от централната гранична теорема за еднакво разпределени събираеми. □

7.3 *Доверителен интервал за вероятност

В този параграф ще се запознаем с един статистически извод — твърдение за неизвестното разпределение на стойностите в изучавано множество от обекти въз основа на ограничената информация получена от една случайна крайна извадка.

Нека си поставим за цел по n наблюдавани стойности да кажем нещо за неизвестната вероятност p на поява на даден признак. Нека с k означим резултата от нашите наблюдения – броят на поява на признака в извадката. Ако предположим, че извадката е случайна и изучаваното множество толкова голямо, че да се пренебрегне възможността от повторение на вече изтеглен обект, то можем да гледаме на резултатите - появата на признака в поредния обект като на независими сл.в. Следователно общия брой появи k е подчинен на биномно разпределение.

Ще използуваме интегралната теорема на Муавър – Лаплас (теорема 7.2).

$$\int_{-x}^{x} \phi(y)dy = \mathsf{P}(-x < \frac{k-np}{(npq)^{1/2}} < x) = \mathsf{P}(|\frac{k}{n} - p| < x\sqrt{pq/n}).$$

Можем да подберем числото $x = 1.96$, тогава вероятността е 0.95. Тъй като $max(pq) = 1/4$, получаваме:

$$0.95 \leq \mathsf{P}(|\frac{k}{n} - p| < \frac{1.96}{2n^{1/2}}).$$

С други думи горното неравенство ни казва: Ако сте наблюдавали определена стойност на k при известен брой на експериментите n, то с вероятност по-голяма от .95 можете да твърдите, че неизвестната стойност на вероятността не е много далече от честотата. При това големината на доверителния интервал е от порядъка на $1/\sqrt{n}$. Смисълът на надежността на вашето твърдение е следния:

При многократно повторение на целия опит и изработване на вашето заключение в 95 процента от случаите вие ще сте прави – неизвестната вероятност ще се покрива от така получения интервал.

7.4 *Формула на Стирлинг

7.4.1 Просто доказателство

Да разгледаме интеграла:

$$n! = \int_0^\infty x^n e^{-x} dx. \tag{7.5}$$

Ще направим смяна на променливите:

$$x^n e^{-x} = n^n e^{-n} e^{-t^2}. \tag{7.6}$$

Тъй като подинтегралната функция има максимум при $x = n > 0$, променливата t ще се мени в интервала $-\infty, \infty$. За да изразим dx ще логаритмуваме 7.6 и ще го диференцираме:

$$n \lg x - x = -t^2 + n \lg n - n, \tag{7.7}$$

$$dx = 2t \frac{x}{x-n} dt \tag{7.8}$$

За да изразим $tx/(x-n)$ само чрез t да отбележим, че

$$t^2 = x - n - n \lg \frac{x}{n} = n(z - \lg(1+z)), \quad z = \frac{x-n}{n}$$

Ще представим $\lg(1+z)$ по формулата на Тейлор до 2-ри ред:

$$\lg(1+z) = z - \frac{z^2}{2} \frac{1}{(1+z\theta(z))^2},$$

където $0 < \theta(z) \leq 1$. Значи (t и z имат еднакъв знак)

$$t^2 = \frac{n}{2} \frac{z^2}{(1+z\theta(z))^2}, \quad t = \sqrt{\frac{n}{2}} \frac{z}{1+z\theta(z)}, \quad z^{-1} + \theta(z) = \sqrt{\frac{n}{2}} t^{-1}. \tag{7.9}$$

Сега можем да изразим $t\frac{x}{x-n}$ чрез t:

$$t\frac{x}{x-n} = t + \frac{t}{z} = \sqrt{\frac{n}{2}} + t(1-\theta(z)). \tag{7.10}$$

Няма нужда да заместваме z, за нас е важно само че $0 < 1-\theta(z) < 1$. Така получаваме за интеграла 7.5:

$$n! = \sqrt{2n}n^n e^{-n}(\int_{-\infty}^{\infty} e^{-t^2}dt + \sqrt{\frac{2}{n}}\int_{-\infty}^{\infty}(1-\theta(z))te^{-t^2}dt.) \tag{7.11}$$

Първият интеграл е добре известният интеграл на Поасон и неговата стойност е $\sqrt{\pi}$, а вторият е ограничен. Така окончателно получаваме:

$$n! = \sqrt{2\pi n}n^n e^{-n}(1+\epsilon_n), \quad \epsilon_n < \frac{C}{\sqrt{n}}.$$

7.4.2 Директно доказателство

Ще предполагаме известно разложението в ред на функцията $\ln(1+x)$:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots (-1)^{n-1}\frac{x^n}{n}\dots \tag{7.12}$$

От него лесно получаваме

$$\ln(\frac{1+x}{1-x}) = \ln(1+x) - \ln(1-x) = 2x(1 + \frac{x^2}{3} + \frac{x^4}{5} + \dots \frac{x^{2m}}{2m+1}\dots) \tag{7.13}$$

Като поставим тук $x = 1/(2n+1)$ получаваме:

$$(n+\frac{1}{2})\ln(1+\frac{1}{n}) = 1 + \frac{(2n+1)^{-2}}{3} + \frac{(2n+1)^{-4}}{5} + \dots \frac{(2n+1)^{-2m}}{2m+1}\dots \tag{7.14}$$

Този израз може да се оцени от двете страни:

$$1 < (n+\frac{1}{2})\ln(1+\frac{1}{n}) \leq 1 + \frac{1}{3(2n+1)^2}(1 + (2n+1)^{-2} + \dots) = 1 + \frac{1}{12n(n+1)}. \tag{7.15}$$

Значи като експоненцираме и разделим на e ще получим неравенствата:

$$1 < e^{-1}(1+\frac{1}{n})^{n+1/2} < e^{\frac{1}{12n(n+1)}} \tag{7.16}$$

От друга страна нека разгледаме редицата $a_n = n!e^n n^{-(n+1/2)}$. Тогава

$$\frac{a_n}{a_{n+1}} = e^{-1}(1+\frac{1}{n})^{n+1/2} \tag{7.17}$$

Значи

$$1 < \frac{a_n}{a_{n+1}} < e^{1/12n(n+1)} = \frac{e^{1/12n}}{e^{1/12(n+1)}}$$

Следователно, $a_n > a_{n+1}$ и $a_n e^{-\frac{1}{12n}} < a_{n+1} e^{-\frac{1}{12(n+1)}}$ и редицата клони към крайна граница a в интервала (0,1), като $a_n e^{-\frac{1}{12n}} < a < a_n$. Значи съществува число θ_n в интервала (0,1), че $a = a_n e^{-\frac{\theta_n}{12n}}$. Следователно,

$$n! = a n^{n+1/2} e^{-n} e^{-\frac{\theta_n}{12n}} \tag{7.18}$$

Остава да се определи константата $a = \sqrt{2\pi}$. Това също може да стане директно, но няма да го правим, тъй като вече го получихме в (7.11).

Тема 8

Схема на Бернули

В тази лекция си поставяме следните цели:

- да разгледаме най - простото безкрайно произведение на вероятностни пространства;
- да въведем понятието изоморфизъм на вероятностни пространства;
- да покажем как възниква понятието информация;
- на примера на най - простата статистическа задача да илюстрираме начина, по който се строят статистическите разсъждения.

8.1 Схема на Бернули

В миналата лекция определихме (определение 6.3) схема на Бернули. Това е редица от независими еднакво разпределени случайни величини $\{\xi_i, i = 1, 2, \ldots\}$, всяка от които приема две стойности: 1 и 0 с вероятности (съответно) p и $q = 1 - p$.

Съгласно теорема 1.1 този набор от случайни величини поражда вероятностна мярка в пространството от безкрайни двоични редици. Да извършим това построение.

Нека означим това пространство с $\mathfrak{E}$ и неговите елементи с $\epsilon = \{\epsilon_1, \epsilon_2, \ldots\}$. За всяко крайно n наборът от първите n случайни величини е векторна случайна величина, която определя в основното вероятностно пространство набор от събития:

$$W_{\epsilon_1,\epsilon_2,\ldots,\epsilon_n} = \{\omega : \{\xi_i = \epsilon_i\}, i = 1, 2, ...n\} = \bigcap_{i=1,2,\ldots n} \{\xi_i = \epsilon_i\}.$$

Определената от тях булова подалгебра W^n се състои от крайни обединения на непресичащи се множества от този тип. W^n е естествено изоморфна на множеството от всички подмножества на крайното множество

$$\mathfrak{E}^n = \prod_{i=1}^{n} \{0, 1\}.$$

От друга страна тези булови алгебри се влагат в пространството $\mathfrak{E}$, като определението на съответните множества се продължава цилиндрично — неупоменатите координати са свободни от ограничение. Там алгебрите образуват нарастваща редица. Тяхното обединение $W = \bigcup W^n$ обаче не е σ-алгебра, макар че остава булова алгебра.

Върху множествата от тази алгебра вероятността се определя просто:

$$\mathsf{P}(W_{\epsilon_1,\epsilon_2,\ldots,\epsilon_n}) = p^k q^{n-k}, \quad k = \sum_{k=1}^{n} \epsilon_i. \tag{8.1}$$

Минималната σ-алгебра, която съдържа W ще означим с $\sigma(W)$. Това че вероятността е непрекъсната върху W следва от нейното определение — всяко ограничение върху нарастващ брой индекси поражда намаляваща към нула редица от вероятности. Така че от теоремата на Каратеодори следва съществуването и единствеността на вероятностната мярка P на $\sigma(W)$. Така получаваме следната теорема:

Теорема 8.1 *Вероятностното пространство* $(\mathfrak{E}, \sigma(W), \mathsf{P})$ *е добре определено и координатите на неговите елементи могат да се разглеждат като схема на Бернули.*

8.2 Изоморфни пространства

Следващата теорема не е необходима за понататъшните разглеждания. Тя обаче дава представа за понятието изоморфизъм в теорията на мярката и интеграла. Ще започнем със следното определение:

Определение 8.1 *Изображението*

$$T : (\Omega', \mathfrak{B}', \mathsf{P}') \longrightarrow (\Omega'', \mathfrak{B}'', \mathsf{P}'')$$

се нарича изоморфизъм на вероятностни пространства, ако

1. T *е определено п.с., т.е. върху достоверно събитие в* Ω'*;*
2. T *е измеримо:* $T^{-1}(B'') \in \mathfrak{B}'$ *за* $\forall B'' \in \mathfrak{B}''$*;*
3. T *запазва вероятността:* $\mathsf{P}'(T^{-1}(B'')) = \mathsf{P}''(B'')$ *за* $\forall B'' \in \mathfrak{B}''$*;*
4. T *е обратимо п.с. в* Ω''*.*

Теорема 8.2 *Вероятностното пространство* $(\mathfrak{E}, \sigma(W), \mathsf{P})$ *е изоморфно на интервала* $[0, 1]$ *с Бореловата* σ*-алгебра и Лебеговата мярка върху нея.*

Доказателство: Ще построим конструктивно съответствието T^{-1} и ще предоставим на заинтересованите да покажат, че то е истински изоморфизъм. Ще построим на полуотворения интервал $[0, 1)$ поредица от разделяния. Първото разделяне w_1 го дели на два подинтервала $[0, q)$ и $[q, 1)$. Разделянето w_{n+1} дели всички получени до момента подинтервали на по два – ляв и десен в същото съотношение q към p. При това левите части се обединени в едно множество, а десните в друго. Така всяко разделяне се състои от две множества - ляво и дясно, а прилагането му удвоява броя на получените до момента полузатворени интервали.

Всяка точка от $[0, 1)$ попада в точно едно от множествата лявото или дясното на разделянето w_n (за всяко n). Означаваме с $\epsilon_n = 0$, когато е попаднала в лявото и $\epsilon_n = 1$ – в дясното. Така получаваме редицата ϵ, съответстващa на точката. За всеки две различни точки от интервала съществува число n такова, че те се различават по първите си $\epsilon_i, i = 1, 2, \ldots, n$. $\square$

По този начин стават изоморфни и сл.в. зададени върху двете пространства. При този изоморфизъм очевидно стойностите им върху подмножества с мярка 0 нямат значение. Така вместо с конкретни функции ние се занимаваме с класове на еквивалентност.

8.3 Количество информация и ентропия

Нека сега разгледаме понятието информация. Естествено е, когато провеждаме един експеримент с отнапред неизвестен изход, да получаваме толкова по - голяма информация от него, колкото по - малка е вероятността на резултата и по - голяма неговата неопределеност. Нека сега да формализираме неопределеността на случайните събития.

Определение 8.2 *Неопределеност (информация от сбъдването) на събитието A наричаме неотрицателна функция Λ определена върху σ-алгебрата от събития във вероятностното пространство със следните свойства:*

1. $\Lambda(\Omega) = 0$;
2. *Ако* $\mathsf{P}(A) < \mathsf{P}(B)$, *то* $\Lambda(A) > \Lambda(B)$;
3. *Ако* A *и* B *са независими, то* $\Lambda(AB) = \Lambda(A) + \Lambda(B)$.

Сред монотонните функции от вероятността P единственото решение изглежда така:

$$\Lambda(A) = \begin{cases} -b \log \mathsf{P}(A), & \mathsf{P}(A) > 0 \\ \infty, & \mathsf{P}(A) = 0. \end{cases}$$

Тук константата $b > 0$ (или основата на логаритъма) може да бъде избрана произволно. Заедно със събитието A очевидно трябва да се разглежда и неговото отрицание $\bar{A}$. Информацията, която получаваме в един експеримент е естествено свързана с двете събития. Така "очакваната" информация би трябвало да се получи по формулата:

$$H(p) = -b(p \log p + q \log q)$$

За да се отстрани неопределеността от b се избира симетричната схема на Бернули $(p = q = \frac{1}{2})$. Тогава функцията $H(p)$ е максимална. Поставя се b такова, че ентропията на един експеримент в тази схема да стане 1. Така избраната единица информация носи названието *бит*.

Нека сега разгледаме всичките възможни (краен брой) изходи на един експеримент, зададени с пълната група събития или измеримото разделяне $\gamma = \{H_1, H_2, ..., H_n\}$. Ентропията е нещо като усреднена неопределеност за дадено разделяне на пространството от елементарни събития.

Определение 8.3 *Ентропия на измеримото разделяне* $\gamma = \{H_1, H_2, \ldots, H_n\}$ *наричаме числото:*

$$H(\gamma) = \sum_{i=1}^{n} \Lambda(H_i)\,\mathsf{P}(H_i).$$

Така получаваме окончателната формула:

$$H(\gamma) = -\sum_{i=1}^{n} \mathsf{P}(H_i)\log_2 \mathsf{P}(H_i).$$

От тук се вижда, че максимална ентропия сред експериментите с k изхода ще има този с равновероятни изходи и тя е равна на $\log_2 k$.

Между крайните измерими разделяния и крайните булови подалгебри във вероятностното пространство съществува естествено съответствие — елементите на разделянето стават атоми на подалгебрата. Особено добре това се вижда в схемата на Бернули. На нарастващата редица от подалгебри W_n съответствуват все по - издребняващи разделяния w_n. Разделянето w_n, например, се състои от 2^n несъвместими събития. При преминаване към следващото разделяне всяко от тези събития се разделя на две и т.н. Това съответствие може да се продължи и за произволни разделяния и σ-подалгебри, но при определени условия.

Така ентропията може да се разглежда като измерител на "издребняването" на дадено разделяне и с нея могат да се сравняват и разделяния с различен брой елементи. По-подробно с математическата теория на информацията човек може да се запознае в (Мартин и Ингленд 1988).

Пример 8.1 *Пресметнете ентропията на схема на Бернули с* n *опита и произволно* p *и покажете, че тя е максимална и равна на* n*, тогава и само тогава, когато* $p = q$.

Тема 9

Непрекъснати разпределения

В тази лекция

- ще въведем формално интеграла на Лебег-Стилтес;
- ще разгледаме най-често срещаните непрекъснати разпределения;
- ще въведем преобразование на Лаплас и характеристични функции;

9.1 Интеграл на Лебег-Стилтес

Както видяхме вече с помощта на ф.р. могат да се записват всички сметки с дискретни и непрекъснати разпределения. Сега ще разгледаме и общия случай. Всички интеграли в тази секция са в граници от $-\infty$ до ∞.

Определение 9.1 *Множеството от линейни комбинации на ф.р. наричаме функции с ограничена вариация (ф.о.в.).*

Лема 9.1 *Всяка ф.о.в. $F(x)$ се представя еднозначно във формата*

$$F(x) = \alpha F_+(x) - \beta F_-(x), \quad 0 \le \alpha, \beta,$$

където F_+ и F_- са функции на разпределение.

Определение 9.2 *Нека $g(x)$ е произволна непрекъсната функция на R^1 и $F(x)$ е ф.о.в. Казваме, че е зададен интеграла на Лебег - Стилтес $\int g(x)dF(x)$, ако са крайни интегралите $\int |g(x)|dF_+(x) < \infty$ и $\int |g(x)|dF_-(x) < \infty$.*

Теорема 9.1 *В частност за всяка ф.р. $F(x)$ и всяка ограничена непрекъсната $g(x)$ е краен и определен $\int |g(x)|dF(x) < \infty$. Лесно се проверяват следните твърдения:*

1. *всяка ф.р. (и ф.о.в.) $F(x)$ може да има най-много изброим брой точки на прекъсване;*

2. *всяка ф.р. (и ф.о.в.) $F(x)$ във всяка точка на непрекъснатост x, може да се представи като граница на редица от чисто скокообразни ф.р. $F_n(x)$ с краен брой скокове;*

3. *монотонност (интегралите трябва да съществуват):*

$$\int g(x)dF_+(x) \leq \int f(x)dF_+(x), \quad g(x) < f(x) \tag{9.1}$$

4. *линейност (интегралите отдясно трябва да съществуват):*

$$\int (\alpha g(x) + \beta f(x))dF(x) = \alpha \int g(x)dF(x) + \beta \int g(x)dF(x),$$
$$\int g(x)d(\alpha F(x) + \beta G(x)) = \alpha \int g(x)dF(x) + \beta \int g(x)dG(x);$$

5. *връзка с интеграла на Риман (съществува $F'(x)$ и всички интеграли отдясно са обсолютно сходящи):*

$$\int g(x)dF(x) = \int g(x)F'(x)dx;$$
$$\int g(x)dF(x) = g(+\infty)F(+\infty) - \int F(x)dg(x);$$

9.2 Преобразование на Лаплас

Определение 9.3 *Преобразованието на Лаплас на неотрицателна сл.в. ξ се задава с формулата:*

$$L(s) = \mathbf{E}\, e^{-s\xi} = \int\limits_0^\infty e^{-sx} dF(x). \tag{9.2}$$

Теорема 9.2 *Преобразованието на Лаплас притежава следните свойства:*

- $L(0) = 1$*;*
- $L'(0) = -\mathbf{E}\,\xi$*, когато съществува;*
- $L''(0) = \mathbf{E}\,\xi^2$*, когато съществува;*
- *когато* $\xi \perp\!\!\!\perp \eta$*,* $L_{\xi+\eta}(s) = L_\xi(s)L_\eta(s)$*;*
- *разпределението се възстановява еднозначно.*

Покажете ги.

Пример 9.1 *Експоненциално разпределение*

Експоненциалното разпределение има плътност: $f(x) = \lambda e^{-\lambda x}$. Следователно, неговото преобразование на Лаплас ще бъде:

$$L(s) = \lambda \int_0^\infty e^{-(s+\lambda)x} dx = \frac{\lambda}{s+\lambda}.$$

Лесно се пресмятат и моментите на това разпределение:

$$\mathbf{E}\,\xi = \frac{1}{\lambda}, \qquad \mathbf{D}\,\xi = \frac{1}{\lambda^2}.$$

Пример 9.2 *Гама-разпределение $\Gamma(a, \lambda)$ (виж формула (11.6) и фиг.11.2).*

Напомняме плътността на това разпределение $f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}$. Следователно, неговото преобразование на Лаплас ще бъде:

$$L(s) = \frac{\lambda^a}{\Gamma(a)} \int_0^\infty x^{a-1} e^{-(s+\lambda)x} dx = \frac{\lambda^a}{(s+\lambda)^a}.$$

Моментите на Гама-разпределението са:

$$\mathbf{E}\,\xi = \frac{a}{\lambda}, \qquad \mathbf{D}\,\xi = L''(0) - L'(0)^2 = \frac{a}{\lambda^2}.$$

9.3 Характеристични функции

Определение 9.4 *Характеристичната функция на произволна сл.в. ξ се задава с формулата:*

$$f(t) = \mathbf{E}\, e^{it\xi} = \int_{-\infty}^{\infty} e^{itx} dF(x). \tag{9.3}$$

Теорема 9.3 *Характеристичните функции притежават следните свойства:*

- $f(0) = 1$;
- *равномерна непрекъснатост*;
- *положителна определеност*: $\forall t_i \in R, a_i \in C$
$$\sum a_i \bar{a}_j f(t_i - t_j) \geq 0;$$
- $f'(0) = i\mathbf{E}\,\xi$, *когато съществува*;
- $f''(0) = -\mathbf{E}\,\xi^2$, *когато съществува.*
- *Когато* $\xi \perp\!\!\!\perp \eta$, $f_{\xi+\eta}(s) = f_\xi(s) f_\eta(s)$.

- *разпределението се възстановява еднозначно по характеристичната си функция.*

Доказателство: Ще докажем само някои от свойствата.
1.Равномерна непрекъснатост:

$$|f(t+h)-f(t)| \leq \int_{-\infty}^{\infty} |e^{ihx}-1| dF(x) \leq$$
$$\int_{|x|<N} |e^{ihx}-1| dF(x) + \int_{|x|\geq N} |e^{ihx}-1| dF(x) \leq$$
$$\int_{|x|<N} \varepsilon dF(x) + \int_{|x|\geq N} 2 dF(x) \leq 3\varepsilon.$$

Тук избрахме N така, че $\int_{|x|\geq N} dF(x) \leq \varepsilon$. След това зафиксирахме $|h| < \delta$ така, че $|e^{ihx}-1| \leq \varepsilon$ за $|x| < N$.
2. Положителната определеност:

$$\sum_{i=1}^{n}\sum_{k=1}^{n} a_i \bar{a}_k f(t_i - t_k) = \sum_{i=1}^{n}\sum_{k=1}^{n} a_i \bar{a}_k \mathbf{E}\, e^{i(t_i-t_k)\xi} =$$
$$\mathbf{E} \sum_{i=1}^{n}\sum_{k=1}^{n} a_i \bar{a}_k e^{it_i\xi} e^{-it_k\xi} = \mathbf{E}\,(Z\bar{Z}) = \mathbf{E}\,|Z|^2. \square$$

Първите три свойства са достатъчни за една функция да бъде характеристична - това е знаменитата теорема на Бохнер.

Пример 9.3 *Характеристичната функция на стандартното нормално разпределение $N(0,1)$ има вида: $f(t) = e^{-t^2/2}$.*

Доказателство: Характеристичната функция на нормалното разпределение може да се представи във вида:

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx) e^{-x^2/2} dx$$

Да означим с $I(t)$ последния интеграл и го диференцираме по t. Тогава

$$I'(t) = -\int_{-\infty}^{\infty} x \sin(tx) e^{-x^2/2} dx = \int_{-\infty}^{\infty} \sin(tx) d e^{-x^2/2} = -tI(t).$$

Значи той удовлетворява диференциалното уравнение:

$$I'(t)/I(t) = -t, \qquad \frac{d \log I(t)}{dt} = -t.$$

Следователно, $I(t) = C.e^{-t^2/2}$. Константата C определяме от равенството $f(0) = 1$.
$\square$

9.4 *Формула за обръщане и сходимости

Може да бъдат доказана и формула за обръщане, т.е. възстановяване на функцията на разпределение (или плътността) от характеристичната функция. За всеки две точки $x < y$ на непрекъснатост на F е изпълнено

$$F(y) - F(x) = \frac{1}{2\pi} \lim_{\sigma \to 0} \int_{-\infty}^{\infty} \frac{e^{ity} - e^{itx}}{it} f(t) e^{-\sigma^2 t^2} dt \tag{9.4}$$

Сходимостта на функциите на разпределение влече сходимост на съответните характеристични функции и обратно. За по-подробно запознаване със свойствата на х.ф. (виж, например, (Обретенов 1974))

Тема 10

Многомерни сл.в.

В тази лекция ще определим многомерни функция на разпределение и плътности. Ще разгледаме и условни разпределения.

10.1 Многомерни функции на разпределение

Многомерната функция на разпределение на сл.в. $\vec{\xi} \in R^n$ се определя просто:

$$F(\vec{x}) = F(x_1, x_2, \dots, x_n) = \mathsf{P}(\bigcap_{i=1}^{n}\{\xi_i < x_i\}), \quad \vec{x} = (\xi_1, \xi_2, \dots, \xi_n)'. \tag{10.1}$$

Тя притежава следните очевидни свойства:

1. $F(-\infty, x_2, \dots, x_n) = 0$;
2. нормираност - $F(\infty, \infty, \dots, \infty) = 1$;
3. монотонност - ако $x_1' < x_2''$, то $F(x_1', x_2, \dots, x_n) \leq F(x_1'', x_2, \dots, x_n)$;
4. ако $\xi_1 \perp\!\!\!\perp \xi_2 \perp\!\!\!\perp \dots \perp\!\!\!\perp \xi_n$, то $F_{\vec{\xi}}(\vec{x}) = \prod_{i=1}^{n} F_\xi(x_i)$;
5. маргиналното разпределение на сл.в. ξ_1 се възстановява лесно:
$$\mathsf{P}(\xi_1 < x) = F_{\xi_1}(x) = F(x, \infty, \dots, \infty).$$

Казваме, че съществува плътност на разпределение $f(x)$ на сл.в. $\vec{x}$, ако:

$$F(\vec{x}) = \int_{\vec{y}<\vec{x}} f(y)dy = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\dots\int_{-\infty}^{x_n} f(y_1, y_2, \dots, y_n) \quad dy_1 dy_2 \dots dy_n.$$

При това плътността се възстановява от функцията на разпределение:

$$f(\vec{x}) = \frac{\partial^n F(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n} \tag{10.2}$$

Плътността притежава следните очевидни свойства:

1. $f(\vec{x}) \geq 0$;
2. $\int_{R^n} f(y)dy = 1$;
3. ако сл.в. ξ_i са независими $f_{\vec{\xi}}(\vec{x}) = \prod_{i=1}^{n} f_\xi(x_i)$.
4. маргиналната плътност на сл.в. ξ_1 се възстановява лесно от многомерната плътност:

$$f_{\xi_1}(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x, y_2, \dots, y_n) dy_2, dy_3 \dots dy_n.$$

10.2 Условни разпределения

Да разгледаме първо двете целочислени сл.в. ξ, η. Тяхното съвместно разпределение се задава с таблицата: $p_{i,j} = \mathsf{P}(\xi = i, \eta = j)$.

Определение 10.1 *Ще наричаме условно разпределение на сл.в. ξ при условие η разпределението:*

$$\mathsf{P}(\xi = i/\eta = j) = \frac{p_{i,j}}{\sum\limits_k p_{k,j}}.$$

За разлика от маргиналните, условните разпределения могат да се определят само за "действителните" стойности на сл.в. η, т.е. тези с ненулева вероятност.

Нека сега разгледаме две непрекъснати сл.в. със съвместна плътност $f(x, y) > 0$. Тук също се оказва възможно определянето на условни разпределения във формата на плътности

$$f_\xi(x/y) = \frac{f(x, y)}{\int f(x, y)dx} = \frac{f(x, y)}{f_\eta(y)}.$$

В горните формули границите на сумите и интегриралите трябва да се избират така, че знаменателите да са положителни. В съответните граници е определено и условното разпределение

10.3 Многомерни моменти

Единствената разлика в случая, когато разглеждаме многомерна случайна величина е в границите на интегралите - това са определени интеграли по цялата област на стойности на непрекъснатата сл. в. или съответните мултииндексни суми за дискретни сл.в.

10.3.1 Коефициент на корелация

Често се използува следното неравенство за смесените моменти на две сл.в:

$$E\xi\eta \leq \sqrt{\mathbf{E}\,\xi^2 \mathbf{E}\,\eta^2} \tag{10.3}$$

Доказателство: Имаме следното неравенство:

$$0 \le E(\frac{\xi}{(\mathbf{E}\,\xi^2)^{1/2}} - \frac{\eta}{(\mathbf{E}\,\eta^2)^{1/2}})^2 =$$
$$= 1 - 2\mathbf{E}\,(\frac{\xi}{(\mathbf{E}\,\xi^2)^{1/2}}\frac{\eta}{(\mathbf{E}\,\eta^2)^{1/2}}) + 1 =$$
$$= 2(1 - \frac{\mathbf{E}\,\xi\eta}{(\mathbf{E}\,\xi^2\mathbf{E}\,\eta^2)^{1/2}}),$$

от което тривиално следва неравенството (10.3). □

Определение 10.2 *Коефициент на корелация на сл.в. ξ и η с краен втори момент наричаме числото*

$$r(\xi,\eta) = \frac{\mathbf{E}\,(\xi - \mathbf{E}\,\xi)(\eta - \mathbf{E}\,\eta)}{\sigma(\xi)\sigma(\eta)} = \mathbf{E}\,\tilde{\xi}\tilde{\eta}. \tag{10.4}$$

Ясно е от това определение, че при независими сл.в. коефициентът на корелация е нула. Когато вместо ξ и η в неравенството (10.3) поставим центрираните и нормирани сл.в. $\tilde{\xi}$ и $\tilde{\eta}$ ще получим, че $|r(\xi,\eta)| \le 1$. Вярна е, обаче следната теорема:

Теорема 10.1 *Ако коефициентът на корелация $|r(\xi,\eta)| = 1$, то между сл.в. съществува линейна връзка: $\eta = a\xi + b$.*

Доказателство: Нека, например, $r(\xi,\eta) = 1$. Да разгледаме равенството:

$$\mathbf{E}\,(\tilde{\xi} - \tilde{\eta})^2 = \mathbf{E}\,\tilde{\xi}^2 - 2\mathbf{E}\,\tilde{\xi}\tilde{\eta} + \mathbf{E}\,\tilde{\eta}^2 = 1 - 2 + 1 = 0 \tag{10.5}$$

От него и (5.7) следва, че $\tilde{\eta} = \tilde{\xi}$ с вероятност 1. Следователно, $\eta = a\xi + b$, където $a = \sigma(\eta)/\sigma(\xi)$, $\quad b = \mathbf{E}\,\eta - a\mathbf{E}\,\xi$. Аналогично се разглежда случая с $r(\xi,\eta) = -1$. Тогава $a = -\sigma(\eta)/\sigma(\xi)$. □

Коефициентът на корелация може да се разглежда като мярка за зависимост между сл.в., което и често се прави на практика.

Пример 10.1 *Зависими сл.в. с нулева корелация.*

Нека разгледаме сл.в. ξ, η приемащи едновременно следните стойности: (-1,1), (0,-1), (1,1), съответно с вероятности 0.25, 0.5, 0.25. Те са зависими, защото $\eta = 2 * \xi^2 - 1$. Пресметнете коефициентът им на корелация.

10.3.2 Ковариационна матрица

В многомерния случай м.о. на сл. вектор е вектор:

$$\mathbf{E}\,\vec{\xi} = \mathbf{E}\begin{pmatrix}\xi_1\\ \vdots \\ \xi_n\end{pmatrix} = \begin{pmatrix}\mathbf{E}\,\xi_1\\ \vdots \\ \mathbf{E}\,\xi_n\end{pmatrix}$$

Особен интерес представляват моментите от втори ред.

Определение 10.3 *Ковариационна матрица на векторната случайна величина се нарича матрицата:*

$$V(\xi) = \mathbf{E}\,(\xi - \mathbf{E}\,\xi)(\xi - \mathbf{E}\,\xi)'.$$

V е квадратна и симетрична матрица. Диагоналните елементи на матрицата V са дисперсиите на съответните сл.в. - координати, а извъндиагоналните елементи — се наричат коефициенти на ковариация. Ако с $\tilde{\xi}$ означим сл.в. съставена от центрираните и нормирани координати на ξ, то $V(\tilde{\xi})$ се нарича корелационна матрица. По диагонала тя съдържа единици, а извън диагоналните елементи са коефициентите на корелация.

Теорема 10.2 *Ковариационната матрица $V(\xi)$ е неотрицателно определена ($\forall y : y'Vy \geq 0$).*

Доказателство: Да означим за краткост $x = \xi - \mathbf{E}\,\xi$ и нека y е произволен вектор.

$$y'Vy = y'(\mathbf{E}\,xx')y = \mathbf{E}\,(y'x)(x'y) = \mathbf{E}\,|y'x|^2 \geq 0.\square$$

Естествено, че същото твърдение е верно и за корелационната матрица

Тема 11

Трансформации на случайните величини

Ще изведем формулата за пресмятане на плътността при аналитична трансформация на непрекъсната сл.в. и ще я приложим за редица разпределения.

11.1 Смяна на променливите

Теорема 11.1 *Нека $U : A \longrightarrow B$, където $A, B \in R^n$ са отворени множества, U е взаимноеднозначно съответствие и $V = U^{-1}$. Нека функцията $V(x)$ притежава непрекъснати производни в B. Нека ξ е сл.в. със стойности в A и тя има плътност $f_\xi(x)$ за $x \in A$. Тогава сл.в. $\eta = U(\xi)$ има плътност $f_\eta(y)$, която се задава по формулата:*

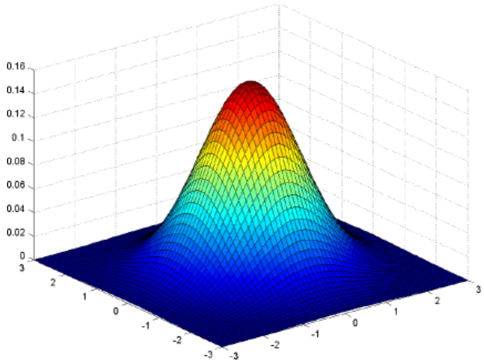
$$f_\eta(x) = |J(V)(x)| f_\xi(V(x)) \quad \text{за} \quad x \in B, \tag{11.1}$$

където с $J(V)$ е якобианът на трансформацията V, т.е. детерминантата на матрицата:

$$\left\{ \begin{array}{cccc} \frac{\partial V_1}{\partial x_1} & \frac{\partial V_1}{\partial x_2} & \dots & \frac{\partial V_1}{\partial x_n} \\ \frac{\partial V_2}{\partial x_1} & \frac{\partial V_2}{\partial x_2} & \dots & \frac{\partial V_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial V_n}{\partial x_1} & \frac{\partial V_n}{\partial x_2} & \dots & \frac{\partial V_n}{\partial x_n} \end{array} \right\}.$$

Тази теорема няма да доказваме – тя е следствие от стандартните теореми на анализа за смяна на променливите под знака на интеграла.

Пример 11.1 *Многомерно нормално разпределение*



Плътността на стандартното нормално разпределение $N(0, I)$ в R^n има вида:

$$\phi(\vec{x}) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}} e^{-x_i^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-||\vec{x}||^2/2},$$

където $\vec{x} \in R^n$.

Фиг. 11.1: Нормално $N(0, I)$ в R^2

Нека сл.в. $\xi \in N(0, I)$. Ще разгледаме линейната трансформация $\eta = A\xi + b$. Тук A е неизродена $n \times n$ матрица, а $b \in R^n$. Тогава плътността на η ще се изчисли по формулата (11.1).

$$f_\eta(x) = |J(V)| f_\xi(V(x)) = \frac{|A^{-1}|}{(2\pi)^{n/2}} e^{-\frac{1}{2}(x-b)'(AA')^{-1}(x-b)}.$$

Като означим матрицата $C = AA'$, получаваме стандартния вид на многомерното нормално разпределение $N(b, C)$ с параметри $\mathbf{E}\,\eta = b$ и $cov(\eta) = C$:

$$\phi(x, b, C) = \frac{1}{|C|^{1/2}(2\pi)^{n/2}} e^{-\frac{1}{2}(x-b)'C^{-1}(x-b)}. \qquad (11.2)$$

Да проверим тези равенства за параметрите:

$$\mathbf{E}\,\eta = A\mathbf{E}\,\xi + b = b, \quad cov(\eta) = \mathbf{E}\,(\eta - b)(\eta - b)' = A(\mathbf{E}\,\xi\xi')A' = AA'.$$

11.2 Конволюция на плътности

Ще приложим формулата 11.1 към следната задача:

Теорема 11.2 *Нека са дадени две независими сл.в. ξ и η с положителни плътности на разпределение. Тогава са изпълнени следните формули:*

$$f_{\xi+\eta}(x) = \int_{-\infty}^{\infty} f_\xi(x - y) f_\eta(y) dy \qquad (11.3)$$

$$\text{Ако} \quad \xi, \eta > 0, \text{то } f_{\xi*\eta}(x) = \int_{-\infty}^{\infty} \frac{1}{y} f_\xi(x/*y) f_\eta(y) dy \qquad (11.4)$$

$$\text{Ако} \quad \xi, \eta > 0, \text{то } f_{\xi/\eta}(x) = \int_{-\infty}^{\infty} y f_\xi(x * y) f_\eta(y) dy \qquad (11.5)$$

Доказателство: Да докажем формула (11.3). Разглеждаме двумерната сл.в. $\{\xi, \eta\}$. Тя има плътност $f(x,y) = f_\xi(y) f_\eta(y)$ защото двете сл.в. са независими. Нека разгледаме сега трансформациите:

$$U = \begin{cases} u = & x + y, \\ v = & y \end{cases} \quad \text{и} \quad V = U^{-1} = \begin{cases} x = & u - v, \\ y = & v \end{cases}$$

и приложим формула (11.1). Тъй като якобианът на V е равен на 1, получаваме за двумерната плътност на $U(\{\xi, \eta\})$ формулата:

$$f(u,v) = f_\xi(u - v) f_\eta(v).$$

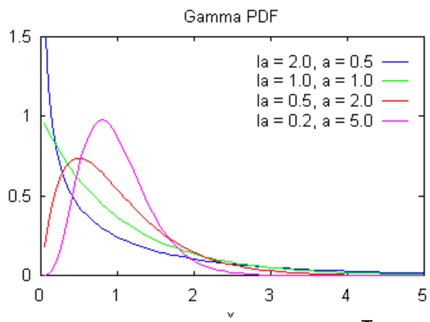
За да получим плътността на първата сл.в. $\xi + \eta$, трябва да интегрираме по втората променлива y. Формули (11.4) и (11.5) се доказват аналогично. $\square$

11.3 Гама и Бета разпределения

Тук ще се запознаем накратко с две много популярни семейства разпределения.

Определение 11.1 *Наричаме Гама-разпределение $\Gamma(a, \lambda)$ разпределение с плътност:*

$$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, \quad x > 0. \tag{11.6}$$



Фиг. 11.2: Гама разпределение

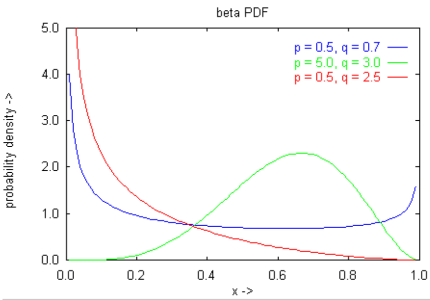
Това семейство е популярно в статистиката, защото е тясно свързано с нормалното. При стойности на a кратни на $1/2$ е известно като Хи-квадрат разпределение и описва разпределението на сума от квадрати на центрирани независими еднакво нормално разпределени сл.в.

Параметърът a, който определя формата му, има смисъла на степени на свобода - колкото по-голям е, толкова по-неопределени са стойностите на сл.в. Гама-разпределението има винаги положителна асиметрия, но тя клони към нула при нарастване на a. Вторият параметър λ е мащабен – той не оказва влияние на ексцеса и асиметрията. При $a \to \infty$ центрираното и нормирано Гама-разпределение клони към нормалното. Математическото му очакване е $\mu = a/\lambda$ а стандартното отклонение - $\sigma = \sqrt{a}/\lambda$.

Определение 11.2 *Наричаме Бета разпределение с плътност:*

$$f(x) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}, \quad 0 < x < 1. \tag{11.7}$$

Тук с $B(a,b)$ сме означили Бета-функцията.



Фиг. 11.3: Различни Бета разпределения

На фиг.11.3 са показани три различни плътности от семейството на Бета разпределенията. Вижда се, че те могат да имат различна по знак асиметрия. С нарастването на параметрите a и b, разпределението се изражда (дисперсията му клони към 0). Ако скоростта на нарастване е еднаква и то е правилно нормирано, Бета разпределението също клони към нормалното.

Ще приложим формулата (11.1) за да опишем връзката между Гама и Бета разпределенията.

Теорема 11.3 *Нека $\xi \in \Gamma(a, \lambda)$ и $\eta \in \Gamma(b, \lambda)$ са независими Гама - разпределени сл.в. Тогава*

1. сл.в. $\zeta = \xi + \eta \in \Gamma(a + b, \lambda)$;
2. сл.в. $\theta = \frac{\xi}{\xi+\eta} \in B(a, b)$;
3. сл.в. $\theta \perp\!\!\!\perp \zeta$.

Доказателство: Разпределението на двумерната сл.в. $\{\xi, \eta\}$ е

$$f(x, y) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \frac{\lambda^b}{\Gamma(b)} y^{b-1} e^{-\lambda y}.$$

Да разгледаме сега трансформациите:

$$U = \begin{cases} u = & x + y, \\ v = & \frac{x}{x+y} \end{cases} \quad \text{и} \quad V = U^{-1} = \begin{cases} x = & uv, \\ y = & u * (1 - v) \end{cases}$$

и приложим формула (11.1). Тъй като якобианът на V е равен на u, получаваме за двумерната плътност на $\{\zeta, \theta\}$ формулата:

$$f(u, v) = (\frac{\lambda^{a+b}}{\Gamma(a + b)} u^{a+b-1} e^{-\lambda u})(\frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} v^{a-1} (1 - v)^{b-1}),$$

откъдето следват всички твърдения на теоремата. □

Тема 12

Видове сходимост на редици сл.в.

Всъщност в по - голямата си част това е тема от математическия анализ. Видовете сходимост се разделят на две големи групи. Първата група касае само разпределенията на сл.в. и следователно, породените от тях мерки върху бореловата σ-алгебра. По традиция от тази група сходимости в теория на вероятностите се изучава само една — тази по разпределение.

Втората група сходимости е по - богата и значително по - използвана. Тук влизат всички сходимости на сл.в., или измерими функции върху абстрактно пространство с мярка.

12.1 Сходимост на разпределения на сл.в.

Определение 12.1 *Казваме, че редицата от функции на разпределение F_n е сходяща към функцията на разпределение F, ако редицата от числа $F_n(x)$ клони към числото $F(x)$ за всяко x, което е точка на непрекъснатост на F.*

Определение 12.2 *Казваме, че редицата от характеристични функции f_n е сходяща към характеристичната функция f, ако редицата от комплексни числа $f_n(t)$ клони към числото $f(t)$ за всяко t.*

Определение 12.3 *Казваме, че редицата от функции на разпределение F_n клони слабо към функцията на разпределение F, ако за всяка ограничена непрекъсната функция f е в сила сходимостта:*

$$\int_{-\infty}^{\infty} f(x)dF_n(x) \longrightarrow \int_{-\infty}^{\infty} f(x)dF(x).$$

Теорема 12.1 *Определенията 12.1, 12.2 и 12.3 са еквивалентни.*

Доказателство: Ще докажем еквивалентността на горните дефиниции в следния ред:

$$12.1 \to 12.3 \to 12.2 \to 12.1$$

Първата стрелка. Нека $F_n(x) \longrightarrow F(x)$ във всяка точка на непрекъснатост на F. Нека $f(x)$ е ограничена и непрекъсната - $|f(x)| < C$.
1. Съществува $L > 0$, такова, че $F(L) - F(-L) > 1 - \epsilon/C$. Тогава $|\int_{|x|<L} f(x)dF(x) - \int_{-\infty}^{\infty} f(x)dF(x)| < 2\epsilon$ и за $n > N$ $|\int_{|x|<L} f(x)dF_n(x) - \int_{-\infty}^{\infty} f(x)dF_n(x)| < 4\epsilon$.
2. В интервала $[-L, L]$ съществуват краен брой k точки на непрекъснатост $x_0 = -L < x_1 < \ldots < x_k = L$ на F, такива, че $\sup_{x_{i-1} \le x \le x_i} |f(x) - f(x_i)| < \epsilon$. Тогава $|\int_{|x|<L} f(x)dF(x) - \sum_{i=1}^{k} f(x_i)(F(x_i) - F(x_{i-1}))| < \epsilon$.
3. Остана да се използува, че $\sup_{i=1}^{k} |F_n(x_i) - F(x_i)| \to 0$.

Втората стрелка е тривиална. Функцията e^{ixt} е ограничена и непрекъсната по x при всяко фиксирано t.

Третата стрелка не е тривиална. Тя следва от формулата за обръщане (9.4) на х.ф. □

12.2 Сходимости на сл.в.

Определение 12.4 *Казваме, че редицата от сл.в. ξ_n е сходяща към сл.в. ξ по разпределение, ако F_n клони към F във всяка точка на непрекъснатост на F . Ще означим тази сходимост по следния начин:*

$$\xi_n \xrightarrow{d} \xi$$

.

Определение 12.5 *Казваме, че редицата от сл.в. ξ_n е сходяща към сл.в. ξ по вероятност, ако за всяко $\epsilon > 0$*

$$\mathsf{P}(|\xi_n - \xi| > \epsilon) \longrightarrow 0.$$

Ще означим тази сходимост по следния начин:

$$\xi_n \xrightarrow{p} \xi.$$

Определение 12.6 *Казваме, че редицата от сл.в. ξ_n е сходяща към сл.в. ξ в средно от степен r, ако*

$$\mathbf{E}\,|\xi_n - \xi|^r \longrightarrow 0.$$

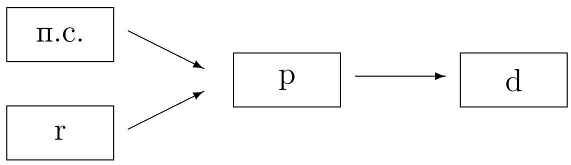
Ще означим тази сходимост по следния начин:

$$\xi_n \xrightarrow{r} \xi.$$

Определение 12.7 *Казваме, че редицата от сл.в. ξ_n е сходяща към сл.в. ξ почти сигурно (или с вероятност 1), ако*

$$\mathsf{P}(|\xi_n - \xi| \longrightarrow 0) = 1.$$

Ще означим тази сходимост по следния начин:



Фиг. 12.1: Сходимости на редици сл.в.

$$\xi_n \xrightarrow{\text{п.с.}} \xi.$$

Между различните видове сходимост съществуват естествени връзки отразени на фиг. 12.1. Стрелките показват от коя от сходимостите следва друга.

Теорема 12.2 *Диаграмата на фиг.12.1 е вярна.*

Доказателство:
Сходимост п.с. влече сходимост по вероятност. Да означим с A_n^r събитията $A_n^r = \{\omega : |\xi_n - \xi| > \frac{1}{r}\}$. Достатъчно е да покажем, че за всяко цяло $r > 0$ имаме $\mathsf{P}(A_n^r) \to 0$ при $n \to \infty$. От друга страна имаме представянето:

$$\{\omega : \xi_n \to \xi\} = \bigcap_{r=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \bar{A}_m^r.$$

Тъй като вероятността на това събитие е 1, то неговото допълнение ще има нулева вероятност.

$$0 = \mathsf{P}(\bigcup_{r=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m^r) = \mathsf{P}(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m^r) = \mathsf{P}(\bigcap_{n=1}^{\infty} B_n^r).$$

Но събитията $B_n^r = \bigcup_{m=n}^{\infty} A_m^r$ образуват намаляваща редица. Тогава от аксиомата за непрекъснатост следва, че $\mathsf{P}(B_n^r) \to 0$ и следователно $\mathsf{P}(A_n^r) \to 0$.
Сходимост в средно r *влече сходимост по вероятност.* Това твърдение следва директно от неравенството на Чебишов:

$$\mathsf{P}(|\xi_n - \xi|^r > \epsilon) \le \frac{\mathbf{E}\,|\xi_n - \xi|^r}{\epsilon^r}.$$

Сходимост по вероятност влече сходимост по разпределение. Да означим с $A_n = \{\xi_n - \epsilon < \xi < \xi_n + \epsilon\}$. При това $\mathsf{P}(A_n) \to 1$. Функцията на разпределение на ξ_n може да се запише по формулата за пълната вероятност в вида:

$$F_n(x) = \mathsf{P}(\xi_n < x) = \mathsf{P}(A_n)\,\mathsf{P}(\xi_n < x | A_n) + \mathsf{P}(\bar{A}_n)\,\mathsf{P}(\xi_n < x | \bar{A}_n).$$

Ще оценим отгоре и отдолу $F_n(x)$:

$$\begin{aligned}
F_n(x) \le\ & \mathsf{P}(\{\xi_n < x\} \textstyle\bigcap A_n) + \mathsf{P}(\bar{A}_n) \\
=\ & \mathsf{P}(\{\xi_n < x\} \textstyle\bigcap \{\xi - \epsilon < \xi_n\} \bigcap \{\xi_n < \xi + \epsilon\}) + \mathsf{P}(\bar{A}_n) \\
\le\ & \mathsf{P}(\{\xi < x + \epsilon\} \textstyle\bigcap \{\xi_n < \xi + \epsilon\}) + \mathsf{P}(\bar{A}_n) \\
\le\ & \mathsf{P}(\{\xi < x + \epsilon\}) + \mathsf{P}(\bar{A}_n).
\end{aligned}$$

При преминаването втория към третия ред използуваме съотношението:

$$\{\xi - \epsilon < \xi_n\} \bigcap \{\xi_n < x\} \subset \{\xi - \epsilon < x\} = \{\xi < x + \epsilon\}.$$

От доказаното неравенство следва, че $\limsup F_n(x) \le F(x + \epsilon)$.

Да разгледаме сега обратното неравенство.

$$\begin{aligned} F_n(x) \ge & \quad \mathsf{P}(\{\xi_n < x\} \bigcap A_n) \\ = & \quad \mathsf{P}(\{\xi_n < x\} \bigcap \{\xi - \epsilon < \xi_n\} \bigcap \{\xi_n < \xi + \epsilon\}) \\ \ge & \quad \mathsf{P}(\{\xi < x - \epsilon\} \bigcap \{\xi - \epsilon < \xi_n\} \bigcap \{\xi_n < \xi + \epsilon\}) = \mathsf{P}(B \bigcap A_n), \end{aligned}$$

където $B = \{\xi < x - \epsilon\}$. Сега използувахме обратното включване:

$$\{\xi < x - \epsilon\} \bigcap \{\xi_n < \xi + \epsilon\} \subset \{\xi_n < x\}.$$

Тъй като $\mathsf{P}(A_n) \to 1$, то $\mathsf{P}(B \bigcap A_n) \to \mathsf{P}(B)$. От тук следва, че $\liminf F_n(x) \ge F(x - \epsilon)$. Тъй като и двете неравенства са изпълнени за всяко ϵ, от тях следва равенството $\lim F_n(x) = F(x)$ за всяка точка на непрекъснатост на $F(x)$. $\square$

12.3 Контрапримери

Пример 12.1 *Сходимост по разпределение не влече сходимост по вероятност.*

Да разгледаме сл.в. ξ приемаща стойностите 1 и -1 с равни вероятности и редицата: $\xi_n = (-1)^n \xi$. Всички разпределения са еднакви, но $\mathsf{P}(|\xi_n - \xi| > \epsilon)$ не клони към 0.

Има и едно изключение, когато сходимостта по разпределение влече сходимост по вероятност — когато граничната сл.в. е константа. Докажете го.

Пример 12.2 *Сходимост по вероятност не влече сходимост п.с.*

Ще конструираме редицата от последователни групи сл.в. n-тата група ще се състои от 2^n сл.в. Да си представим, че вероятностното пространство е интервала $[0, 1)$ с мярка на Лебег. За да конструираме n-тата група ще го разделим на 2^n равни подинтервала (затворени от ляво). Сл.в. от групата приемат стойности 1 за точно един подинтервал и нула навсякъде другаде. Така $\mathsf{P}(|\xi_n| > \epsilon) \to 0$ и редицата клони по вероятност към константата 0. За всяко конкретно ω, обаче, съществуват безбройно много членове на редицата със стойности по - големи от ϵ, т.е. тя не е сходяща п.с.

От всяка редица сходяща по вероятност, може да се извлече подредица сходяща п.с. към същата гранична сл.в. Докажете това.

Пример 12.3 *Сходимост по вероятност не влече сходимост в средно* r.

Сходимостта на редицата по вероятност към 0 в пример 12.2 не зависи от стойностите на величините, когато са различни от 0. Значи, ако умножим сл.в. от n-тата група с $2^{n/r}$ ще получим, че $\mathbf{E}\left(|\xi_n|^r\right) = 1$, което означава, че редицата не е сходяща към 0 в средно, но остава сходяща по вероятност.

Пример 12.4 *Сходимост п.с. не влече сходимост в средно* r*.*

Достатъчно е да разгледаме подредица на редицата от пример 12.3, която е сходяща п.с. Това могат да бъдат, например, последните (най - десни) сл.в. от всяка група. Тук, както и в предния пример, сходимостта п.с. не зависи от стойностите им. Сходимостта в средно, обаче е нарушена.

Когато една редица сл.в. е ограничена и сходяща п.с., то тя е сходяща и в средно. Докажете го.

Пример 12.5 *Сходимост в средно* r *не влече сходимост п.с.*

Редицата от пример 12.2 има такова поведение. Проверете го.

Тема 13

Закони за големите числа

Изложението на тази лекция следва стандартните учебници по теория на вероятностите за математици. Целите са

- да се докаже най-простата форма на слабия закон за големите числа;
- да се опишат някои основни свойства на редици от независими сл.в.;
- да се изведат необходими и достатъчни условия за силния закон на големите числа.

13.1 Слаб закон за големите числа

Нека е дадена редица от независими сл.в.$\xi_1, \xi_2, \ldots$. Да означим с $S_n = \sum_{i=1}^n \xi_i$ редицата от парциални суми. Пита се при какви условия редицата $\eta_n = \frac{1}{n}(S_n - \mathbf{E}\, S_n)$ клони към 0. В този параграф ще разгледаме най-слабата възможна сходимост — тази по вероятност.

$$\mathsf{P}(\frac{1}{n}|S_n - \mathbf{E}\, S_n| > \epsilon) \to 0 \qquad \text{при} \qquad n \to \infty. \tag{13.1}$$

Определение 13.1 *Когато съотношението 13.1 е изпълнено за дадена редица сл.в. $\xi_1, \xi_2, \ldots$, казваме че за тази редица е в сила слаб закон за големите числа СЗГЧ.*

Теорема 13.1 *(Марков) Нека $\xi_1, \xi_2, \ldots$ е редица от сл.в. Ако*

$$\frac{1}{n^2}\mathbf{D} \sum_{i=1}^n \xi_i \to 0 \qquad \textit{при} \qquad n \to \infty. \tag{13.2}$$

Тогава за тази редица е в сила СЗГЧ.

Доказателство: Теоремата е директно следствие от неравенството на Чебишов (неравенство (5.9)). Наистина имаме:

$$\mathsf{P}(\frac{1}{n}(S_n - \mathbf{E}\, S_n) > \epsilon) < \frac{\mathbf{D}\, S_n}{\epsilon^2 n^2}.\square$$

Теорема 13.2 *(Хинчин) Нека $\xi_1, \xi_2, \dots$ е редица от независими и еднакво разпределени сл.в. с крайно математическо очакване $a = \mathbf{E}\xi_1$. Тогава за тази редица е в сила СЗГЧ.*

Доказателство: Да означим с $a = \mathbf{E}\,\xi_1$. Ясно е, че можем да разглеждаме центрирани сл.в. $a = 0$. Нека означим с $f(t) = \mathbf{E}\,e^{it\xi_1}$ х.ф. на сл.в. ξ_i (те всичките са еднакви). Имаме

$$\mathbf{E}\,e^{it\eta_k} = (1 - o(\frac{t}{k}))^k \to 1.$$

Използувахме развитието на Тейлор на $f(t)$ около 0 — първата производна съществува и е 0. Х.ф. $f(t) = 1$ съответствува на константата 0. $\square$

13.2 Редици независими сл.в.

Лема 13.1 *Нека е дадена редицата от събития A_n. Ако е сходящ реда*

$$\sum_{i=1}^{\infty} \mathbf{P}(A_n) < \infty, \tag{13.3}$$

то

$$\mathsf{P}(\limsup_{n\to\infty} A_n) = 0. \tag{13.4}$$

Доказателство: Да означим с $B_n = \bigcup_{k\geq n} A_n$. Тогава е очевидно неравенството:

$$\mathsf{P}(B_n) \leq \mathsf{P}(\bigcup_{k=n}^{\infty} A_k) \leq \sum_{k=n}^{\infty} \mathsf{P}(A_k) \to 0.$$

Условието (13.4) е еквивалентно на условието $\lim_n \mathsf{P}(B_n) = 0$. $\square$

Лема 13.2 *Лема на Борел-Кантели. Нека е дадена редицата от независими събития A_n. Ако е изпълнено (13.4), то е в сила (13.3) и обратно.*

Доказателство: Условието (13.4) означава, че $\mathsf{P}(B_n) \to 0$.

$$1 - \mathsf{P}(B_n) = \mathsf{P}(\bar{B}_n) = \mathsf{P}(\bigcap_{k=n}^{\infty} \bar{A}_k) = \prod_{k=n}^{\infty}(1 - \mathsf{P}(A_k)) \to 1.$$

Сходимостта на безкрайното произведение е еквивалентна на търсената сходимост. Така за независими събития условията (13.3) и (13.4) стават еквивалентни. $\square$

13.3 Неравенство на Колмогоров

Теорема 13.3 *(Колмогоров) Нека $\xi_1, \xi_2, \dots$ е редица от центрирани ($\mathbf{E}\,\xi_n = 0$) независими сл.в. Тогава е в сила следното неравенство:*

$$\mathsf{P}(\sup_{1\le k\le n} |S_k| > \epsilon) \le \frac{\mathbf{D}\,S_n}{\epsilon^2} \tag{13.5}$$

Доказателство: Да разгледаме пълната група събития:

$$H_0 = \{\sup_{1\le k\le n} |S_k| \le \epsilon.\}, \quad H_j = \{j = \text{първото } k \le n : |S_k| > \epsilon\}.$$

За тази група е лесно да напишем веригата неравенства:

$$\mathbf{E}\,S_n^2 \ge \sum_{j=1}^n \mathbf{E}\,(S_n^2 I_{H_j}) = \sum_{j=1}^n \mathbf{E}\,(S_j^2 + 2S_j \sum_{k=j+1}^n \xi_i + (\sum_{k=j+1}^n \xi_k)^2) I_{H_j} =$$
$$\sum_{j=1}^n \mathbf{E}\,(S_j^2 I_{H_j}) + 2\sum_{j=1}^n \mathbf{E}\,(S_j I_{H_j} \sum_{k=j+1}^n \xi_i) + \sum_{j=1}^n \mathbf{E}\,((\sum_{k=j+1}^n \xi_k)^2 I_{H_j}) \ge$$
$$\epsilon^2 \sum_{j=1}^n \mathsf{P}(H_j) = \epsilon^2\,\mathsf{P}(\sup_{1\le k\le n} |S_k| > \epsilon). \tag{13.6}$$

Тук използуваме, че събитието H_j се определя изцяло от стойностите на първите j сл.в. и, следователно, неговият индикатор I_{H_j} е функция на тези сл.в. Значи е сл.в. независима от сл.в. $\xi_{j+1}, \xi_{j+2}, \dots, \xi_n$. Така вторият член на получената сума е равен на 0. □

13.4 Силен закон за големите числа

Нека разгледаме сходимостта:

$$\mathsf{P}(\sup_{n\le k} \frac{1}{k}(S_k - \mathbf{E}\,S_k) > \epsilon) \to 0 \qquad \text{при} \qquad n \to \infty. \tag{13.7}$$

Определение 13.2 *Когато съотношението 13.7 е изпълнено за дадена редица $\xi_1, \xi_2, \dots,$ казваме че за тази редица е в сила усилен закон за големите числа (УЗГЧ).*

Ясно е, че при проверка на сходимостта (13.7) е достатъчно да се разглеждат центрирани величини. Тогава тя може да се запише във формата:

$$\mathsf{P}(\sup_{k>n} |\frac{S_k}{k}| > \epsilon) \longrightarrow 0, \text{ при } n \to \infty \text{ за всяко } \epsilon > 0. \tag{13.8}$$

Теорема 13.4 *(Колмогоров) УЗГЧ е в сила за редица от нееднакво разпределени независими сл.в., за която е сходящ реда:*

$$\sum_{n=1}^{\infty} \frac{\mathbf{D}\,\xi_n}{n^2} < \infty$$

Доказателство: Теоремата е директно следствие от неравенството на Колмогоров (13.5). Нека разгледаме събитията:

$$A_j = \{\max_{2^{j-1} \le k < 2^j} |\frac{S_k}{k}| > \epsilon\}, \qquad j = 1, 2, \dots .$$

Тогава 13.8 е еквивалентно на условието: $\mathsf{P}(\bigcup_{j=m}^{\infty} A_j) \longrightarrow 0$ при $m \to \infty$, което е изпълнено ако е сходящ редът $\sum \mathsf{P}(A_j) < \infty$. Това се проверява със следната верига неравенства (вторият ред следва от неравенството (13.5) на Колмогоров):

$$\mathsf{P}(A_j) \le \mathsf{P}(\max_{2^{j-1} \le k < 2^j} |S_k| > \epsilon 2^{j-1}) \le \mathsf{P}(\max_{k<2^j} |S_k| > \epsilon 2^{j-1}) \le$$
$$\frac{1}{\epsilon^2 2^{2(j-1)}} \sum_{k<2^j} \mathbf{D}\,\xi_k = \frac{4}{\epsilon^2 2^{2j}} \sum_{k<2^j} \mathbf{D}\,\xi_k.$$

Тогава получаваме

$$\sum_{j=1}^{\infty} \mathsf{P}(A_j) \le \sum_{j=1}^{\infty} \frac{4}{\epsilon^2 2^{2j}} \sum_{k<2^j} \mathbf{D}\,\xi_k =$$
$$\frac{4}{\epsilon^2} \sum_{k=1}^{\infty} \mathbf{D}\,\xi_k \sum_{j>\log_2 k} 2^{-2j} \le \frac{16}{3\epsilon^2} \sum_{k=1}^{\infty} \frac{\mathbf{D}\,\xi_k}{k^2}. \square$$

Теорема 13.5 *(Колмогоров) Нека $\xi_1, \xi_2, \dots$ е редица от еднакво разпределени и независими сл.в. Тогава за тази редица е в сила УЗГЧ, ако и само ако е ограничен първият абсолютен момент $\mathbf{E}\,|\xi_1| < \infty$.*

Доказателство: Доказателството ще проведем в три последователни стъпки:

1. Ще разгледаме редица от подходящо урязани сл.в., за които е изпълнена теорема 13.4.
2. Ще покажем, че за двете редици е изпълнен едновременно УЗГЧ.
3. Накрая ще покажем и необходимостта на условието за краен първи момент.

1. Нека разгледаме урязаните величини:

$$\xi_n^* = \begin{cases} \xi & \xi_n < n \\ 0, & \xi_n \ge n \end{cases}$$

Да покажем сега, че за редицата ξ_n^* са изпълнени условията на теорема 13.4. Това е следствие от следната лема:

Лема 13.3 $\mathbf{E}\,|\xi| < \infty$ *тогава и само тогава, когато* $\sum \mathsf{P}(|\xi| > n) < \infty$.

Доказателство: Да означим с $H_n = \{\omega : n-1 \le |\xi| < n\}$.

$$\mathbf{E}\,|\xi| = \mathbf{E}\,|\xi|(\sum_{i=1}^{\infty} I_{H_i}) = \sum_{i=1}^{\infty} \mathbf{E}\,|\xi| I_{H_i} \le \sum_{i=1}^{\infty} i\,\mathsf{P}(H_i) =$$

$$\sum_{i=1}^{\infty} \mathsf{P}(H_i) \sum_{k=1}^{i} 1 = \sum_{k=1}^{\infty}\sum_{i=k}^{\infty} \mathsf{P}(H_i) = \sum_{k=1}^{\infty} \mathsf{P}(|\xi| > k).$$

$$\mathbf{E}\,|\xi| = \sum_{i=1}^{\infty} \mathbf{E}\,|\xi| I_{H_i} \ge \sum_{i=1}^{\infty} (i-1)\,\mathsf{P}(H_i) = \sum_{k=1}^{\infty} \mathsf{P}(|\xi| > k) - 1.\square$$

Сега да покажем условията на теорема 13.4.

$$\sum_{n=1}^{N} \frac{\mathbf{D}\,\xi_n^*}{n^2} \le \sum_{n=1}^{N} \frac{\mathbf{E}\,|\xi|^2 (\sum_{i=1}^{n} I_{H_i})}{n^2} = \mathbf{E}\,|\xi|^2 \sum_{i=1}^{N} I_{H_i} \sum_{n=i}^{N} \frac{1}{n^2} \le$$

$$\mathbf{E}\,|\xi|^2 \sum_{i=1}^{N} I_{H_i} (\frac{1}{i} + \frac{1}{i^2}) \le \mathbf{E}\,|\xi| + 1.$$

2. Ако означим с $A_n = \{\xi_n \ne \xi_n^*\}$, то по лемата на Борел - Кантели имаме, че $\limsup A_n = 0$, защото

$$\sum \mathsf{P}(A_n) = \sum \mathsf{P}(|\xi| > n) < \infty.$$

Следователно, двете редици $\frac{1}{n}\sum_{i=1}^{n} \xi_n$ и $\frac{1}{n}\sum_{i=1}^{n} \xi_n^*$ са сходящи едновременно и към една и съща граница (събитията A_n се случват само за краен брой индекси). Остава да намерим границата на редицата

$$\lim_n \frac{1}{n} \sum_{i=1}^{n} \mathbf{E}\,\xi_n^* = \lim \mathbf{E}\,\xi_n^* = \lim \mathbf{E}\,\xi I_{\bar{A}_n} = \mathbf{E}\,\xi,$$

защото $\mathsf{P}(A_n) \to 0$.

3. Необходимостта следва лесно от съотношението:

$$\frac{1}{n}\xi_n = \frac{1}{n} S_n - \frac{n-1}{n}\frac{1}{n-1} S_{n-1} \to 0$$

и лемата 13.2 на Борел - Кантели:

$\mathsf{P}(\lim_n \frac{1}{n}\xi_n = 0) = 1$, следователно, $\mathsf{P}(\limsup_n \{\frac{1}{n}|\xi_n| > \epsilon\}) = 0$, следователно, $\sum \mathsf{P}(|\xi| > n\epsilon) = \sum \mathsf{P}(|\xi_n|/\epsilon > n) < \infty$ следователно, по лема 13.3, $\mathbf{E}\,|\xi| < \infty$. $\square$

Тема 14

Централна гранична теорема

Изложението на тази лекция следва стандартните учебници по теория на вероятностите за математици. Предмет на централната гранична теорема (ЦГТ) е следната задача. Нека е дадена редица от сл.в. $\xi_1, \xi_2, \ldots$. Да означим с $S_n = \sum_{i=1}^n \xi_i$ редицата от парциални суми. Пита се при какви условия съответно нормирана тази редица клони по разпределение към гаусова сл.в.:

$$\mathbf{P}\left(\frac{S_n - ES_n}{(DS_n)^{1/2}} < x\right) \underset{n\to\infty}{\longrightarrow} \Phi(x). \tag{14.1}$$

Определение 14.1 *Когато съотношението 14.1 е изпълнено за дадена редица сл.в. $\xi_1, \xi_2, \ldots$, казваме че за тази редица е в сила ЦГТ.*

Така основната задача на всички разработки в тази област е да се установят както необходимите условия за изпълнение на ЦГТ, така и скоростта, с която се достига граничното разпределение.

В тази лекция ще разглеждаме само редици от независими сл.в.

14.1 Еднакво разпределени събираеми

Теорема 14.1 *Нека $\xi_1, \xi_2, \ldots$ е редица от еднакво разпределени независими сл.в. с крайна дисперсия σ^2. Тогава за тази редица е в сила ЦГТ.*

Доказателство: Да означим с $a = \mathbf{E}\,\xi_1$. Да разгледаме редицата от центрирани и нормирани сл.в. $\eta_n = (\xi_n - a)/\sigma$ и означим с $f(t) = \mathbf{E}\,e^{it\eta}$ характеристичната им функция. Тъй като $\mathbf{E}\,\eta = 0$ и $\mathbf{D}\eta = 1$, то $f(t)$ притежава първа и втора производни в т.0 и може да се развие в ред на Тейлор около тази точка: $f(t) = 1 - \frac{t^2}{2} + O(t^3)$.

Сега да препишем вероятността в условие (14.1):

$$\mathbf{P}\left(\frac{S_n - ES_n}{(DS_n)^{1/2}} < x\right) = \mathbf{P}\left(\frac{1}{n^{1/2}}\sum_{i=1}^n \eta_i < x\right)$$

Да разгледаме сега характеристичната функция на сл.в. $s_n = \frac{1}{n^{1/2}}\sum_{i=1}^n \eta_i$:

$$\mathbf{E}\,e^{its_n} = (\mathbf{E}\,e^{itn^{-1/2}\eta_1})^n = (f(tn^{-1/2}))^n = \left(1 - \frac{t^2}{2n} + O\left(\frac{t^3}{n^{\frac{3}{2}}}\right)\right)^n \to e^{-\frac{t^2}{2}}.\Box$$

14.2 Условие на Линдеберг

Тук ще разгледаме общия случай на независими случайни величини с произволни разпределения. Ще въведем следните означения. $F_k(x) = \mathbf{P}(\xi_k < x), a_k = \mathbf{E}\,\xi_k, \sigma_k^2 = \mathbf{D}\xi_k, s_n^2 = \sum_{k=1}^n \sigma_k^2$.

Това, което първо е забелязано, е следното необходимо условие за пренебрежимост:

$$M_n \stackrel{def}{=} \max_{1\le k\le n} \frac{\sigma_k^2}{s_n^2} \underset{n\to\infty}{\longrightarrow} 0. \tag{14.2}$$

В случая с еднакво разпределени сл.в. условието за пренебрежимост е естествено изпълнено: $M_n = 1/n$.

Следното условие носи името на Линдеберг. За всяко $\epsilon > 0$

$$L_n(\epsilon) \stackrel{def}{=} \frac{1}{s_n^2}\sum_{k=1}^n \int_{|x-a_k|>\epsilon s_n} (x-a_k)^2 dF_k(x) \underset{n\to\infty}{\longrightarrow} 0. \tag{14.3}$$

Теорема 14.2 *(Линдеберг – Фелер) Нека $\xi_1, \xi_2, \dots$ е редица от независими сл.в. с крайни дисперсии, за които е изпълнено условието за пренебрежимост (14.2). Тогава, за да е в сила ЦГТ (14.1) необходимо и достатъчно е условието на Линдеберг (14.3).*

Доказателство: Да означим с $F_{nk}(x) = P(\xi_k - a_k < x s_n) = F_k(a_k + x s_n)$. Тогава можем да запишем двете условия (14.2) и (14.3) по - лесно:

$$\begin{aligned} M_n &= \max_{1\le k\le n} \tfrac{\sigma_k^2}{s_n^2} = \max_{1\le k\le n} \int\limits_{-\infty}^{\infty} (\tfrac{x-a_k}{s_n})^2 dF_k(x) = \max_{1\le k\le n} \int_{-\infty}^{\infty} x^2 dF_{nk}(x) \\ L_n(\epsilon) &= \sum_{k=1}^n \int\limits_{|x-a_k|>\epsilon s_n} \tfrac{(x-a_k)^2}{s_n^2} dF_k(x) = \sum_{k=1}^n \int_{|x|>\epsilon} x^2 dF_{nk}(x) \end{aligned}$$

Случайните величини $\eta_{nk} = (\xi_k - a_k)/s_n$ са с функции на разпределение F_{nk}. Да означим техните характеристични функции с f_{nk}. Тогава твърдението на ЦГТ (14.1) може да се запише по следния начин:

$$\prod_{k=1}^n f_{nk}(t) \to e^{-\frac{t^2}{2}} \qquad \text{или} \qquad \sum_{k=1}^n \ln f_{nk}(t) \to -\frac{t^2}{2}. \tag{14.4}$$

Доказателството ще проведем в няколко стъпки. При това първите три стъпки имат отношение както към необходимостта, така и към достатъчността.

А. Първо ще покажем,че условието за пренебрежимост (14.2) следва от условието на Линдеберг (14.3). Това твърдение всъщност не е необходимо за доказателството на теоремата, но е поучително.

$$M_n = \max_{1\le k\le n} \Big(\int_{|x|\le\epsilon} x^2 dF_{nk}(x) + \int_{\epsilon<|x|} x^2 dF_{nk}(x)\Big) \le \epsilon^2 + L_n(\epsilon).$$

Тъй като ϵ е произволно от тук следва условието за пренебрежимост.

Б. От условието за пренебрежимост (14.2) следват следните твърдения:

$$H_n(t) \stackrel{def}{=} \max_{1\le k\le n} |f_{nk}(t)-1| \to 0 \tag{14.5}$$

$$\sum_{k=1}^{n} |f_{nk}(t)-1| \le t^2/2. \tag{14.6}$$

Достатъчно е да се използува следното неравенство:

$$|f_{nk}(t)-1| = |\int\limits_{-\infty}^{\infty}(e^{itx}-1-itx)dF_{nk}(x)| \le \frac{t^2}{2}\int\limits_{-\infty}^{\infty} x^2 dF_{nk}(x) = \frac{t^2}{2}\frac{\sigma_k^2}{s_n^2}.$$

В. Ще докажем пак като следствие от пренебрежимостта (14.2) следното твърдение:

$$R_n(t) \stackrel{def}{=} \sum_{k=1}^{n}(\ln f_{nk}(t) - (f_{nk}(t)-1)) \to 0. \tag{14.7}$$

Тъй като е изпълнено (14.5) можем да изберем такова голямо n, че $|f_{nk}(t)-1| < 1/2$. Тогава имаме:

$$\ln f_{nk}(t) = \ln(1+(f_{nk}(t)-1)) = (f_{nk}(t)-1) + \sum_{j=2}^{\infty}(f_{nk}(t)-1)^j(-1)^j/j$$

.

$$|\ln f_{nk}(t) - (f_{nk}(t)-1)| \le (f_{nk}(t)-1)^2 \sum_{j=0}^{\infty}\frac{1}{(j+2)2^j} \le \frac{1}{2}(f_{nk}(t)-1)^2$$

Като сумираме тези неравенства и използуваме (14.6) получаваме:

$$R_n(t) \le \frac{1}{2}\sum_{k=1}^{n}(f_{nk}(t)-1))^2 \le \frac{1}{2}H_n(t)\sum_{k=1}^{n}|f_{nk}(t)-1| \le H_n(t)\frac{t^2}{4} \to 0.$$

Достатъчност (Линдеберг) Накрая оценяваме израза:

$$|I_n(t)| \stackrel{def}{=} |t^2/2 + \sum_{k=1}^{n}(f_{nk}(t)-1)|$$

$$= |\sum_{k=1}^{n}\int\limits_{-\infty}^{\infty}(e^{itx}-1-itx+\frac{t^2x^2}{2})dF_{nk}(x)| \le$$

$$\le \frac{|t|^3\epsilon}{6}\sum_{k=1}^{n}\int\limits_{|x|\le\epsilon}|x|^2dF_{nk}(x) + |t|^2\sum_{k=1}^{n}\int\limits_{\epsilon<|x|}|x|^2dF_{nk}(x)$$

$$\le \frac{|t|^3\epsilon}{6} + |t|^2L_n(\epsilon)$$

С това доказателството на достатъчността е завършено.

Необходимост (Фелер) Тъй като по условие пренебрежимостта (14.2) е изпълнена, то в сила са (14.5) и (14.6). Можем да предполагаме, че $I_n(t) \to 0$ за всяко t. Оценяваме отдолу:

$$Re(I_n(t)) = \sum_{k=1}^{n} \int_{-\infty}^{\infty} (\cos(tx) - 1 + \frac{t^2x^2}{2}) dF_{nk}(x) \geq$$

$$\sum_{k=1}^{n} (- \int_{|x|\leq\epsilon} \frac{(tx)^2}{2} dF_{nk}(x) + \int_{\epsilon<|x|} (-2) dF_{nk}(x) + \int_{-\infty}^{\infty} \frac{(tx)^2}{2} dF_{nk}(x)) =$$

$$\sum_{k=1}^{n} (\int_{\epsilon<|x|} \frac{(tx)^2}{2} dF_{nk}(x) - 2 \int_{\epsilon<|x|} dF_{nk}(x)) \geq$$

$$\frac{t^2}{2} L_n(\epsilon) - \sum_{k=1}^{n} \frac{\sigma_k^2}{\epsilon^2 s_n^2}$$

В последния ред използувахме неравенството на Чебишов. Като прехвърлим последния член и умножим на $2/t^2$ получаваме неравенството:

$$\frac{2}{t^2}(\frac{2}{\epsilon^2} + I_n(t)) \geq L_n(\epsilon).$$

Тъй като дясната страна не зависи от t, това е достатъчно да твърдим, че тя клони към нула. $\square$

Забележка В доказателството използувахме следните неравенства:

$$1 - \cos(\alpha) \leq \alpha^2/2, \quad |e^{it} - 1 - it| \leq \frac{t^2}{2},$$

$$|e^{it} - 1 - it + t^2/2| \leq \frac{t^3}{6}, \quad |e^{it} - 1 - it + t^2/2| \leq t^2,$$

както и неравенството на Чебишов.
Доказателство: Предполагаме t реално.

$$|e^{it} - 1| = |i \int_0^t e^{iz} dz| \leq \int_0^t |e^{iz}| dz = |t|$$

$$|e^{it} - 1 - it| = |i \int_0^t (e^{iz} - 1) dz| \leq \int_0^t |z| dz = |t|^2/2$$

Аналогично получаваме:

$$|e^{it} - 1 - it + t^2/2| \leq |t|^3/3!$$

$$|e^{it} - 1 - it + t^2/2| \leq t^2/2 + t^2/2 = t^2.\square$$

14.3 Следствия

Теорема 14.3 *(Теорема на Ляпунов) Нека $\xi_1, \xi_2, \dots$ е редица от независими сл.в. с крайни моменти от ред $2+\delta, \delta > 0$. Ако е изпълнено следното условие (14.8) на Ляпунов, то е в сила ЦГТ, т.е. условие (14.1).*

$$\Lambda_n(r) \stackrel{def}{=} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \int_{-\infty}^{\infty} (x-a_k)^{2+\delta} dF_k(x) = \int_{-\infty}^{\infty} |x|^{2+\delta} dF_{nk} \underset{n\to\infty}{\longrightarrow} 0, \qquad (14.8)$$

Доказателство: Доказателството следва веднага от неравенството:

$$L_n(\epsilon) = \sum_{k=1}^{n} \int_{|x|>\epsilon} x^2 dF_{nk} \le \frac{1}{\epsilon^\delta} \sum_{k=1}^{n} \int_{|x|>\epsilon} |x|^{2+\delta} dF_{nk} \le \frac{1}{\epsilon^\delta} \Lambda_n(r). \square$$

Тема 15

Процеси с независими нараствания

15.1 Въведение

Нека разгледаме семейството от случайни величини $\{\xi_\tau, \tau \geq 0\}$. За нас параметърът τ ще има ролята на време и затова такива семейства ще наричаме случаен процес. Процесите с дискретно време всъщност представляват редици случайни величини. Теория на случайните процеси представлява специален раздел в теория на вероятностите с голямо значение. (виж. (Гихман и Скороход 1977))

За всяко τ е определена функцията на разпределение $F(x,\tau) = P(\xi_\tau < x)$ и моментите на процеса (когато те съществуват) $m(\tau) = \mathbf{E}\,\xi_\tau, \sigma^2(\tau) = \mathbf{E}\,(\xi_\tau - m(\tau))^2$ и т.н. Ще предположим, че сл.в. $\xi_0 = 0$ и семейството $F(x,\tau)$ е непрекъснато в нулата, т.е. $F(x,+0) = F(x,0)$. Нека означим с $\theta_{\tau_1,\tau_2} = \xi_{\tau_2} - \xi_{\tau_1}$ нарастванията на процеса.

Определение 15.1 *При тези предположения ще казваме че процесът е с независими нараствания, ако сл.в. θ_{τ_1,τ_2} и ξ_{τ_2} са независими за всички τ_1, τ_2.*

Това определение не е най-точното, но за нашите цели е напълно достатъчно. Всъщност би трябвало да поискаме независимост в съвкупност на всеки краен брой нараствания.

Определение 15.2 *Процес с независими нараствания ще наричаме еднороден (е.п.н.н.), ако разпределението на сл.в. θ_{τ_1,τ_2} зависи само от разликата на двата параметъра $\tau_2 - \tau_1$.*

От това определение е ясно, че за е.п.н.н. сл.в. θ_{τ_1,τ_2} и $\xi_{\tau_2-\tau_1}$ имат еднакво разпределение. Също така от тук следва, че моментите на процеса (когато съществуват) удовлетворяват съотношенията: $m(\tau_1+\tau_2) = m(\tau_1)+m(\tau_2)$, $\sigma^2(\tau_1+\tau_2) = \sigma^2(\tau_1)+\sigma^2(\tau_2)$.

15.2 Характеристична функция

Нека разгледаме сега семейството от характеристични функции определени от еднороден процес с независими нараствания.

$$f(t,\tau) = Ee^{it\xi_\tau} = \int_{-\infty}^{\infty} e^{itx} dF(x,\tau) \tag{15.1}$$

Първо ще докажем една проста лема.

Лема 15.1 *Ако комплексната функция от реален аргумент удовлетворява условията:* $\Phi(0) = 1$*, непрекъснатост в нулата и функционалното уравнение:* $\Phi(x_1+x_2) = \Phi(x_1)\Phi(x_2)$ *за всички реални* x_1, x_2*, то:*
а. $\Phi(1) \neq 0$*;*
б. $\Phi(x) = \Phi(1)^x$.

Доказателство: а. Допускаме противното. Тогава от равенството $0 = \Phi(1) = \Phi(\frac{1}{n})^n$ следва $0 = \Phi(\frac{1}{n})$ за всяко n, което противоречи на непрекъснатостта в нулата.

б. За рационални $x = m/n$ имаме $\Phi(x) = \Phi(1)^x$. Но от непрекъснатостта в нулата следва непрекъснатост за всяко x:

$$\Phi(x + \Delta x) = \Phi(x)\Phi(\Delta x) \to \Phi(x).\square$$

Следствие Ако реална функция от реален аргумент удовлетворява условията: $\Phi(0) = 0$, непрекъснатост в нулата и функционалното уравнение: $\Phi(x_1+x_2) = \Phi(x_1)+\Phi(x_2)$ за всички реални x_1, x_2, то $\Phi(x) = \Phi(1)x$.
Доказателство: Достатъчно е да разгледаме функцията $e^{\Phi(x)}$ и приложим лемата. $\square$

Теорема 15.1 *За всеки еднороден процес с независими нараствания е изпълнена следната формула представяне на характеристичните му функции:*

$$\ln f(t,\tau) = \tau \lim_{\Delta\tau\to 0} \frac{1}{\Delta\tau} \int_{-\infty}^{\infty} (e^{itx} - 1) dF(x,\Delta\tau) \tag{15.2}$$

Доказателство: Нека фиксираме t. Тогава $f(t,0) = 1$, функцията f е непрекъсната по втория си аргумент в нулата (от условието за непрекъснатост на семейството от функции на разпределение) и удовлетворява функционалното уравнение. Прилагаме лемата и получаваме:

$$f(t,\tau) = (f(t,1))^\tau = e^{\tau \ln f(t,1)}$$

Тъй като логаритъмът е многозначна функция, при логаритмуването получаваме:

$$\ln f(t,1) = \ln|f(t,1)| + i(arg f(t,1) + 2k\pi).$$

Но $f(t,1) \neq 0$ и можем да изберем k така, че $f(0,1) = 1$ - само тогава $f(.,1)$ е характеристична функция.

Да разгледаме сега диференчното частно (t - фиксирано):

$$\frac{f(t,\Delta\tau) - 1}{\Delta\tau} = \frac{e^{\Delta\tau \ln f(t,1)} - 1}{\Delta\tau} = \ln f(t,1) + 0(\Delta\tau) = \frac{1}{\tau}\ln f(t,\tau) + 0(\Delta\tau).$$

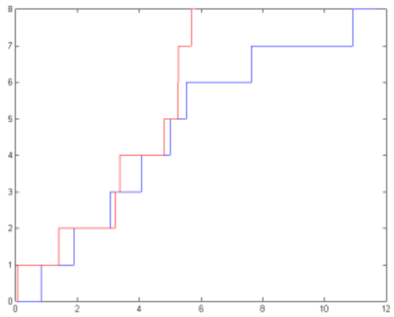
Тук функцията $0(\Delta\tau)$ клони към нула, когато $\Delta\tau$ клони към нула.

От друга страна, същото частно може да се запише така

$$\frac{f(t,\Delta\tau) - 1}{\Delta\tau} = \frac{1}{\Delta\tau}\int_{-\infty}^{\infty} (e^{itx} - 1) dF(x,\Delta\tau).\square$$

15.3 Поасонов процес

Тук като един от примерите за е.п.н.н. ще разгледаме процес с дискретно пространство на състоянията.



Тук сме показали две траектории на такъв процес. Изискването случайните величини да приемат за стойности само натуралните числа и свойството за независимост на нарастванията автоматически налага такава форма на траекториите.

Фиг. 15.1: Поасонов процес

Да означим съответните вероятности с $P_k(\tau) = \mathsf{P}(\xi_\tau = k), k = 0, 1, 2, \ldots$. Ясно е, че от непрекъснатостта на семейството от функции на разпределение следва, че $P_k(\tau) \to 0$, когато $\tau \to 0$ за всяко $k \neq 0$ и $P_0(\tau) \to 1$.

Теорема 15.2 *Нека допълнително поискаме условието (условие за ординарност)*

$$\sum_{k \geq 2} P_k(\tau) = o(P_1(\tau)). \tag{15.3}$$

Тогава

$$F(x, \tau) = \sum_{k<x} \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau},$$

т.е. всички разпределения са Поасонови.

Тук $o(\epsilon)$ означава произволна функция такава, че $o(\epsilon)/\epsilon \to 0$, когато $\epsilon \to 0$.
Доказателство: Ясно е, че $P_0(0) = 1$, $P_0(\tau_1 + \tau_2) = P_0(\tau_1)P_0(\tau_2)$ и $P_0(.)$ е непрекъсната в 0. Тогава от лемата следва, че може да запишем $P_0(.)$ във формата $P_0(\tau) = P_0(1)^\tau = e^{-\lambda\tau}$. Тук $P_0(1) = e^{-\lambda}, \lambda > 0$. Тогава $P_1(\tau) \leq \sum_{k \geq 1} P_k(\tau) = 1 - e^{-\lambda\tau} \to 0$.

За да изразим останалите вероятности, вече ще трябва да използуваме условието за ординарност. Имаме, че за всяко τ е изпълнено равенството:

$$P_0(\tau) + P_1(\tau) + \sum_{k=2}^{\infty} P_k(\tau) = 1.$$

Следователно

$$\begin{aligned} P_0(\Delta\tau) = e^{-\lambda\Delta\tau} &= 1 - \lambda\Delta\tau + o(\Delta\tau) \\ P_1(\lambda\Delta\tau) &= \lambda\Delta\tau + o(\Delta\tau) \\ \sum_{k=2}^{\infty} P_k(\Delta\tau) &= o(\Delta\tau), \end{aligned}$$

Сега използуваме представянето на характеристичната функция на е.п.н.н.:

$$\ln f(t,\tau) = \tau \lim_{\Delta\tau\to 0} \frac{1}{\Delta\tau}\int_{-\infty}^{\infty}(e^{itx}-1)dF(x,\Delta\tau) =$$

$$\tau \lim_{\Delta\tau\to 0}\frac{1}{\Delta\tau}((e^{-it0}-1)P_0(\Delta\tau)+(e^{-it1}-1)(\lambda\Delta\tau+o(\Delta\tau))+o(\Delta\tau)) =$$

$$= \tau\lambda(e^{-it}-1).$$

Следователно получаваме

$$f(t,\tau) = e^{\tau\lambda(e^{-it}-1)}, \tag{15.4}$$

което е характеристичната функция на Поасоново разпределение. □

Пример 15.1 *В телефонна станция постъпват средно по 60 повиквания за един час. Каква е вероятността да не постъпи нито едно повикване за 30 сек.*

Решение: $\mathsf{P}(\xi_{\frac{1}{2}}=0) = e^{-\frac{\lambda}{2}} = e^{-\frac{1}{2}} = 0.6065.$

Пример 15.2 *Колко стафидки трябва да сложим в тестото, така че с вероятност .99 във всяка кифличка да попадне поне една стафидка.*

Решение: Да означим с λ броят стафиди на единица обем и с v обема на една кифличка. Имаме

$$\mathsf{P}(\xi_v \geq 1) = 1 - e^{-\lambda v} \geq .99$$

$$e^{-\lambda v} < 0.01, \lambda v > \ln 100 = 4.6$$

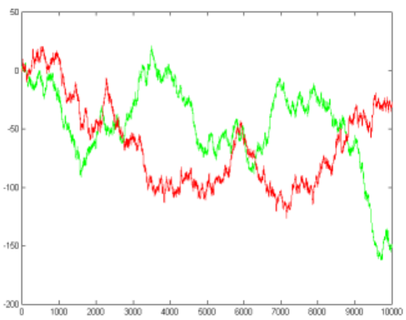
Следва да предвидим в тестото по 4.6 стафидки на кифличка.

15.4 Винеров процес

Ще разгледаме втори пример за е.п.н.н. Този път на случайните величини ще бъде разрешено да приемат произволни реални стойности.За сметка на това ще наложим ограничение върху съществуването и "пренебрежимостта" на третия момент:

$$\int_{-\infty}^{\infty}|x|^3 dF(x,\Delta\tau) = o(\Delta\tau) \tag{15.5}$$

Тук са нарисувани две симулирани траектории на стандартно Брауново движение. Те апроксимират при подходяща нормировка траекториите на стандартен Винеров процес.



Фиг. 15.2: Винеров процес

Теорема 15.3 *При указаните изисквания характеристичните функции на процеса имат представянето*

$$\ln f(t,\tau) = itm\tau - \frac{1}{2}\sigma^2 t^2 \tau,$$

където m и $\sigma > 0$ са подходящо подбрани константи.

Така всички разпределения на процеса стават гаусови. Когато $m = 0$ и $\sigma = 1$ процесът се нарича стандартен Винеров процес или Брауново движение с непрекъснато време.

Доказателство: Да отбележим, че съществуването и непрекъснатостта в нулата на третия абсолютен момент влече съществуването и непрекъснатостта на първия и втория моменти в нулата.

$$m(\tau) = \int_{-\infty}^{\infty} x dF(x,\tau), \qquad \sigma^2(\tau) = \int_{-\infty}^{\infty} x^2 dF(x,\Delta\tau) - m^2(\tau)$$

Да разгледаме функцията $m(\tau)$. Имаме $m(0) = 0, m(\tau_1+\tau_2) = m(\tau_1)+m(\tau_2)$ и непрекъснатост в нулата. От следствието на лемата следва, че съществува константа m такава, че $m(\tau) = m\tau$. Същото е верно и за функцията $\sigma^2(\tau) = \sigma^2\tau$.

Нека приложим сега нашата теорема за е.п.н.н.

$$\ln f(t,\tau) = \tau \lim_{\Delta\tau\to 0} \frac{1}{\Delta\tau} \int_{-\infty}^{\infty} (e^{itx} - 1) dF(x,\Delta\tau) =$$

$$\tau \lim_{\Delta\tau\to 0} \frac{1}{\Delta\tau} (\int_{-\infty}^{\infty} (e^{itx} - 1 - itx + \frac{t^2x^2}{2}) dF(x,\Delta\tau) +$$

$$+ itm\Delta\tau - \frac{t^2}{2}(\sigma^2\Delta\tau + m^2\Delta\tau^2)).$$

Първият член на това представяне клони към нула поради неравенството $|e^{itx} - 1 - itx + \frac{t^2x^2}{2}| \leq \frac{|t^3x^3|}{6}$. Последният член е квадратичен по $\Delta\tau$ и също клони към нула. Останалите два члена са търсените. $\square$

15.5 Гранична теорема

Ергодично свойство ще наричаме преход в някакво състояние на случайния процес независимо от началното състояние. Тук ще докажем една проста теорема за еднородните процеси с независими нараствания и крайна дисперсия - аналог на централната гранична теорема.

Да означим нормирания и центриран случаен процес с $\eta_\tau = \frac{\xi_\tau - m_\tau}{\sigma_\tau}$ и с $\Phi(x)$ ф.р. на стандартното гаусово разпределение.

Теорема 15.4 *На лице е сходимостта:*

$$\lim_{\tau\to\infty} F_\eta(x,\tau) = \Phi(x).$$

Доказателство: Характеристичната функция на модифицирания процес е:

$$\ln f_\eta(t,\tau) = \ln(e^{i\frac{tm}{\sigma\tau^{1/2}}} f(\frac{t}{\sigma\tau^{1/2}},1)^\tau) = i\frac{tm}{\sigma\tau^{1/2}} + \tau \ln f(\frac{t}{\sigma\tau^{1/2}},1),$$

където $m_\tau = m\tau$ и $\sigma_\tau^2 = \sigma^2\tau$. Нека сега развием характеристичната функция $\ln f(t,1)$ в ред на Тейлор около нулата $(\ln f(0,1))' = f'(0,1) = im$ и $(\ln f(0,1))' = f'(0,1) = -(\sigma^2 + m^2)$:

$$\ln f(\frac{t}{\sigma\tau^{1/2}},1) = im\frac{t}{\sigma\tau^{1/2}} - (\sigma^2+m^2)\frac{t^2}{\sigma^2\tau} + o(\frac{1}{\tau}).$$

Остава да извършим умножението и съкратим излишните членове. □

15.6 Заключителни бележки

И двата примера изложени в предходните секции се оказа, че притежават крайни дисперсии и даже моменти от произволен ред. Това се дължи на наложените ограничения, които водят до диференцируемост на характеристичните функции по първия аргумент. За съжаление това не винаги е така, както показва следният пример:

$$\ln f(t,\tau) = \alpha\tau|t|^\gamma.$$

При $\gamma = 1$ такава характеристична функция съответствува на разпределение на Коши, което не притежава даже първи момент.Проверете, че такова семейство характеристични функции може да бъде получено чрез теоремата и напишете в този случай съответното твърдение.

Тема 16

Марковски вериги

Марковските процеси са обобщение на процесите с независими нараствания. Нека е зададен процесът $\xi_\tau, \tau \geq 0$.

Определение 16.1 *Ще казваме, че процесът е Марковски, ако бъдещето и миналото са независими при фиксирано настояще.*

Същият смисъл имат думите: Бъдещето зависи само (се определя изцяло) от последния наблюдаван момент. Разгледания в миналата лекция Поасонов процес е очевиден пример за Марковски процес. Състоянието на процеса в даден бъдещ момент е проста сума на състоянието му в момента и съответната случайна величина определяща нарастването му. Така е очевидно и за всички процеси с независими нараствания — те са марковски.

За разлика от е.п.н.н., Марковските процеси могат да притежават и стойности, които не са числови. Наистина, стойностите на е.п.н.н. трябва да принадлежат на адитивна група — това изисква определението им.

Особеността при дефиниция на марковски процеси идва от това, че трябва да се разглежда и проверява условна независимост. Когато вероятността на дадено фиксирано състояние е нулева (както това е при процеси с непрекъснато пространство на състоянията), такава проверка е нетривиална — тя изисква наличието на многомерна вероятностна плътност.

16.1 Марковски вериги

Тук ще се спрем на процеси с дискретно време и крайно дискретно пространство на състоянията. Такива процеси наричаме марковски вериги. Подробно с тях може да се запознаете в книгите (Кемени и Снелл 1970) и (Димитров 1980). Ще предположим, че множеството от състояния на веригата са числата $1, 2, \ldots, k$.

Ако условно означим събитията в "миналото", "настоящето" и "бъдещето" със буквите A, B, C, то Марковското свойство означава просто $\mathsf{P}(A \cap C|B) = \mathsf{P}(A|B)\,\mathsf{P}(C|B)$. Сега дефиницията за марковост може да се препише по - точно:

Определение 16.2 *Ще казваме, че процесът е крайна Марковска верига, ако за всички възможни набори $\{i_0, i_1, \ldots, i_n\}$, $(1 \leq i_k \leq k)$ и всяко $n \geq 1$ е изпълнено:*

$$\begin{gathered}\mathsf{P}(\xi_0 = i_0, \xi_1 = i_1, \ldots, \xi_n = i_n, \xi_{n+1} = i_{n+1} | \xi_k = i_k) = \\ \mathsf{P}(\xi_0 = i_0, \xi_1 = i_1, \ldots, \xi_{k-1} = i_{k-1} | \xi_k = i_k) * \\ \mathsf{P}(\xi_{k+1} = i_{k+1}, \ldots, \xi_n = i_n | \xi_k = i_k).\end{gathered}$$

Марковското свойство дава възможност съвместните разпределения на множеството случайни величини да бъдат пресмятани просто, както показва следната теорема.

Теорема 16.1 *Разпределенията на крайна Марковска верига се определят от началното разпределение* $\{\pi_i = P(\xi_0 = i), i = 1, 2, \ldots, k\}$ *и матриците на преходни вероятности (във всеки един момент* $\{n = 0, 1, 2, \ldots\}$ *на времето)* $P_{i,j}(n) = \mathsf{P}(\xi_{n+1} = j | \xi_n = i)$*,* $(i, j = 1, 2, \ldots k,)$ *по формулата:*

$$\mathsf{P}(\xi_0 = i_0, \xi_1 = i_1, \ldots, \xi_n = i_n) = \pi_{i_0} P_{i_0,i_1}(0) \ldots P_{i_{n-1},i_n}(n-1). \tag{16.1}$$

Доказателство: Доказателството се прави по индукция и следва лесно от определението и формулата за умножение на вероятности. Да отбележим, че понеже $\mathsf{P}(A|B) = \mathsf{P}(A \cap B|B)$, Марковското свойство може да се запише и така: $\mathsf{P}(A \cap B \cap C|B) = \mathsf{P}(A \cap B|B)\,\mathsf{P}(C \cap B|B)$.
Нека да допуснем, че сме доказали формулата (16.1) за някое n (за $n = 0, 1$ тя е очевидна) и всички възможни набори $\{i_0, i_1, \ldots, i_n\}$. От определението 16.2 следва формулата:

$$\begin{gathered}\mathsf{P}(A \cap B \cap C|B) = \mathsf{P}(\xi_0 = i_0, \xi_1 = i_1, \ldots, \xi_n = i_n, \xi_{n+1} = i_{n+1} | \xi_n = i_n) = \\ \mathsf{P}(\xi_0 = i_0, \xi_1 = i_1, \ldots, \xi_n = i_n | \xi_n = i_n)\,\mathsf{P}(\xi_n = i_n, \xi_{n+1} = i_{n+1} | \xi_n = i_n) = \\ \mathsf{P}(\xi_0 = i_0, \xi_1 = i_1, \ldots, \xi_n = i_n | \xi_n = i_n) P_{i_n,i_{n+1}}(n) = \mathsf{P}(A \cap B|B)\,\mathsf{P}(C|B)\end{gathered}$$

Като умножим тази формула с вероятността $\mathsf{P}(B) = \mathsf{P}(\xi_n = i_n)$ получаваме търсената рекурсия. □

Определение 16.3 *Ще казваме, че Марковската верига е еднородна, ако преходните вероятности* $P_{i,j}(n)$ *не зависят от* n*.*

Да означим тази единствена матрица на преходните вероятности с P и вектора от началните вероятности с π_0. Матрицата P се нарича стохастична, защото сумата на елементите й по редове е 1. Една от основните задачи на теорията на Марковските вериги е изследването на граничното поведение на $P^n = P * P * \cdots * P$. Да означим елементите на тази матрица с $p^n_{i,j}, n = 2, 3, \ldots$. Очевидни са следните уравнения на А.Марков:

$$p^{n+1}_{i,j} = \sum_{l=1}^{k} p_{i,l} p^n_{l,j} \tag{16.2}$$

Сега лесно можем да изразим и вектора от вероятности в момента n: $\pi_n = \{\mathsf{P}(\xi_n = i)\}$, $(i = 1, 2, \ldots, k)$:

$$\pi'_n = \pi_0{}' P^n.$$

16.2 Гранично и стационарно разпределения

Следната теорема се нарича ергодична теорема на А.А.Марков

Теорема 16.2 *Ако всички елементи на стохастичната матрица са строго положителни, то съществуват числа $p_j, 0 \leq p_j \leq 1, \sum_j p_j = 1$, такива че $\lim_{n\to\infty} p^n_{i,j} = p_j$ независимо от индекса i.*

Доказателство: Снабдяваме линейното пространство R^k с норма: $||x|| = \sum_{i=1}^k |x_i|$. В тази норма то е пълно — всяка сходяща по Коши редица има граница. Да разгледаме подмножеството K на R^k определено от условията: $\sum_{i=1}^k x_i = 1, x_i \geq 0, i = 1, 2, \ldots, k$.То е компактно и трансформацията P го изобразява в себе си. Ще покажем, че P е свиваща трансформация и, следователно, има единствена неподвижна точка, към която се стремят неговите итерации.

А. Върху елементите на K разстоянието, породено от указаната норма, се пресмята по следния начин:

$$||x - y|| = 2\sum_{i\in J}(x_i - y_i). \tag{16.3}$$

Тук с J сме означили множеството от индекси на положителни елементи под знака на сумата. Да отбележим, че това множество е винаги строго по - малко от цялото множество от индекси. Ако не беше така, би било невъзможно $\sum_{i=1}^k (x_i - y_i) = 0$.

Б. Нека означим с $\alpha = \min_{i,j} p_{i,j}$. Тогава е изпълнено условието за свиване:

$$||x'P - y'P|| \leq (1-\alpha)||x - y||. \tag{16.4}$$

Това следва от следната верига неравенства:

$$\begin{array}{rcl} ||x'P - y'P|| = & 2\sum_{j\in J}\sum_i p_{i,j}(x_i - y_i) & = \\ = & 2\sum_i(\sum_{j\in J} p_{i,j})(x_i - y_i) & \leq \\ \leq & 2\sum_{i\in I}(\sum_{j\in J} p_{i,j})(x_i - y_i) & \leq \\ \leq & 2\sum_{i\in I}(1-\alpha)(x_i - y_i) & = (1-\alpha)||x - y||. \end{array}$$

В. Нека означим със z неподвижната точка на P - точката, удовлетворяваща уравнението: $z'P = z'$. Тъй като диаметърът на множеството K е 2 имаме:

$$||z' - x'P^n|| \leq ||(z - x)'P^n|| \leq 2(1-\alpha)^n.$$

Г. Тогава границата P^∞ съществува, тя е линеен оператор и е проектор върху тази точка и, следователно, може да се запише във формата: $P^\infty = ez'$. □

Коментар От доказателството е ясно, че граничното състояние на такава верига е единствено и се достига с експоненциална скорост. Ясно е също, че теоремата може да се приложи и за вериги, при които само някаква степен на преходната матрица притежава положителни членове.

Определение 16.4 *Ще наричаме разпределението $p = \{p_i, i = 1, 2, \ldots, k\}$ върху пространството на състояния на еднородна крайна Марковска верига стационарно, ако удовлетворява следното съотношение:*

$$p'P = p'. \tag{16.5}$$

Теорема 16.3 *Граничното разпределение (когато съществува) е стационарно.*

Доказателство: Удобно е да се използуват за целта т.н. уравнения на Колмогоров:

$$P^{n+1} = P^n P. \tag{16.6}$$

Те се извеждат точно както и уравненията на Марков (16.2). Не е трудно да се провери за граничното разпределение, че то трябва да ги удовлетворява, т.е. да бъде стационарно. Да означим вектора $e = \{1, 1, \ldots, 1\}'$ и с p - този от граничните вероятности. Тогава от съществуването на гранично разпределение получаваме $P^n \to ep'$, а преминаването в граница на уравненията (16.6) води до $ep' = ep'P$. Така се получава, че граничното разпределение p е ляв (нетривиален) собствен вектор на P. Това означава, че ако в даден момент веригата има такова разпределение, то ще се запази неизменно и за в бъдеще. □

16.3 Класификация на състоянията

Ще започнем тази тема с два тривиални примера.

Пример 16.1 *Да разгледаме следните матрици на преходни вероятности:*

$$P_1 = \begin{pmatrix} q & p \\ 0 & 1 \end{pmatrix}, \qquad P_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

При матрицата P_1 системата остава завинаги в състоянието 2, когато го достигне. Такова състояние се нарича *поглъщащо.* Понякога цял клас състояния може да бъде поглъщащ – веднъж достигнат вече не може да бъде напуснат. Поглъщащото състояние (или клас) може да се достига в случаен момент на времето – за тази конкретна верига разпределението му е геометрично.

Състоянието 1 в този пример пък е такова, че веднъж напуснато то не може да бъде достигнато никога. Състояние (или клас състояния) с това свойство наричаме *невъзвратни.*

При матрицата P_2 периодично се сменят три състояния в реда:

$$\begin{aligned} 1, 2, 3, 1, 2, 3, .. \text{ или} \\ 2, 3, 1, 2, 3, 1, .. \text{ или} \\ 3, 1, 2, 3, 1, 2, .. \end{aligned}$$

и всъщност цялата вероятностна мярка е съсредоточена върху три траектории, започващи от различни начални състояния в зависимост от вектора на началните вероятности π_0. Бъдещото поведение на такава верига става напълно предсказуемо — то детерминирано се определя от нейното настояще. Състояние от този тип се нарича периодично. Естествено е, че състоянията, през които преминава траекторията при своето периодично движение, образуват *периодичен* клас състояния. Такъв клас е винаги поглъщащ, но може, разбира се, да бъде достигнат в случаен момент на времето. Състояние, което не принадлежи на никой периодичен клас се нарича *апериодично*.

Изобщо казано състоянията на веригата могат да бъдат частично наредени: казваме че състоянието $i \prec j$, ако $p_{i,j}^n > 0$ за някое n. Ясно е, че еквивалентните състояния в тази наредба образуват клас. Когато от класа не може да се излезе, класът е максимален в указаната частична наредба. Една верига може да има няколко такива класа. Такъв клас ще наричаме *ергодичен*, а състоянията в него *ергодични*.

Класификацията на състояния представлява особен интерес в теорията на Марковските вериги поради следните теореми:

Теорема 16.4 *Възвратната апериодична Марковска верига е ергодична.*

Всъщност не е трудно да се съобрази, че теоремата на Марков следва от тази теорема — периодичността и невъзвратността винаги са свързани с наличието на нулеви вероятности. Така неговата матрица с положителни елементи е матрица на ергодична верига.

Теорема 16.5 *Граничното разпределение на ергодична Марковска верига е единствено и не зависи от началното състояние.*

16.4 Примери и задачи

Пример 16.2 *"Канал с шум" или Марковска верига с две симетрични състояния.*

Системата сменя състоянието си с вероятност $p < 1/2$ или остава в същото състояние с вероятност $q = 1 - p$. Тогава преходната матрица P е симетрична. Нейната n-та степен P^n е също симетрична и двете съответно имат вида:

$$P = \begin{pmatrix} q & p \\ p & q \end{pmatrix}, \qquad P^n = \begin{pmatrix} 1-p_n & p_n \\ p_n & 1-p_n \end{pmatrix}.$$

Уравненията на Марков в този случай се записват просто:

$$\begin{aligned} 1 - p_n &= (1-p)(1-p_{n-1}) + pp_{n-1} \\ p_n &= p(1-p_{n-1}) + (1-p)p_{n-1}. \end{aligned} \tag{16.7}$$

Чрез изваждане на второто уравнение от първото получаваме:

$$1 - 2p_n = (1-2p)(1-2p_{n-1})$$

$$1 - 2p_n = (1-2p)^n$$

Следователно получаваме, че $p_n \to \frac{1}{2}$. Така в този частен случай граничното разпределение е равномерното. $\square$

Пример 16.3 *Неергодична верига.*

Да разгледаме веригата със симетрична преходна матрица:

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & q & p \\ 0 & p & q \end{pmatrix}.$$

Да означим с π вектора от начални вероятности. Покажете, че граничното разпределение (зависещо от началното) е

$$\pi_1, (\pi_2 + \pi_3)/2, (\pi_2 + \pi_3)/2.$$

Приложение А

Допълнения

Мястото на тези допълнения е по-скоро в курса по Вероятности 2. Тук са дадени само за пълнота.

А.1 Теорема на Каратеодори

Тук ще докажем тази знаменита теорема (1.1).

Теорема А.1 *Ако една вероятност* P*, зададена върху буловата алгебра* $\mathfrak{F}$*, е непрекъсната в* $\emptyset$*, то тя е продължима еднозначно върху* $\sigma(\mathfrak{F})$.

Определение А.1 *Нека определим върху всички подмножества на* Ω *функцията горна мярка:*

$$\mu(A) = \inf\{\mathsf{P}(B) : B \in \mathfrak{F}, A \subset B\}.$$

Горната мярка μ притежава следните важни и почти очевидни свойства:

1. монотонност - $\forall A \subset B, \quad \mu(A) \leq \mu(B)$;
2. изчислимост - $\forall \epsilon > 0, \exists B \in \mathfrak{F} : A \subset B, \mu(B) - \mu(A) < \epsilon$;
3. полуадитивност (крайна или изброима)- $\mu(\bigcup_i A_i) \leq \sum_i \mu(A_i)$;
4. на $\mathfrak{F}$ горната мярка μ съвпада с P и е σ-адитивна (и непрекъсната в $\emptyset$).

Докажете сами тези свойства.

Определение А.2 *Да означим с* $\mathfrak{R}$ *класа от подмножества* $B \subset \Omega$*, за които е изпълнено равенството*

$$1 = \mu(B) + \mu(\overline{B}).$$

Ясно е, че $\mathfrak{R} \supset \mathfrak{F}$ и $\mu(B) = \mathsf{P}(B)$, когато $B \in \mathfrak{F}$.

Лема А.1 *За да принадлежи множеството* B *на* $\mathfrak{R}$ *е необходимо и достатъчно да бъде "апроксимируемо", т.е. да съществуват редиците* $\{B_n^+\}, \{B_n^-\} \in \mathfrak{F}$ *такива, че:*

$$B_n^- \subset B_{n+1}^- \subset B \subset B_{n+1}^+ \subset B_n^+, \qquad \mu(B_n^+) - \mu(B_n^-) = \mathsf{P}(B_n^+ \setminus B_n^-) \to 0.$$

Доказателство: Необходимост. За всяко B по определение (A.1) съществува редица $B_n \in \mathfrak{F}$ такава, че $B_n \supset B$ и $\mu(B) = \lim_n \mathsf{P}(B_n)$. Да означим с $B_n^+ = \cap_{i=1}^n B_i$. Тогава $\mu(B) = \lim_n \mathsf{P}(B_n^+)$. Същото е валидно и за $\overline{B}$ - $\exists B_n \in \mathfrak{F} : \mu(\overline{B}) = \lim_n \mathsf{P}(B_n), B_n \supset \overline{B}$. Нека означим с $B_n^- = \overline{\cap_{i=1}^n B_i}$. Така получаваме, че $B_n^- \subset B \subset B_n^+$. Тъй като $B \in \mathfrak{R}$, то $\mu(B) = 1 - \mu(\overline{B})$ и $\mathsf{P}(B_n^-) \uparrow \mu(B)$, $\mathsf{P}(B_n^+) \downarrow \mu(B)$.

Достатъчността е очевидна:

$$1 \le \mu(B) + \mu(\overline{B}) \le \mathsf{P}(B_n^+) + 1 - \mathsf{P}(B_n^-) \to 1.\square$$

Доказателство: (Теорема A.1.)
1.Да покажем, че $\mathfrak{R}$ е алгебра. $\mathfrak{R}$ очевидно съдържа $\Omega, \emptyset$, както и допълнението на всяко множество $B \in \mathfrak{R}$. Нека $A, B \in \mathfrak{R}$. Да означим с $A_n^+, B_n^+, A_n^-, B_n^-$ някои апроксимиращи редици на двете множества. Имаме очевидното включване: $A_n^- \cap B_n^- \subset A \cap B \subset A_n^+ \cap B_n^+$.

$$\begin{aligned}
\mathsf{P}(A_n^+ \cap B_n^+) - \mathsf{P}(A_n^- \cap B_n^-) &= \\
\mathsf{P}(A_n^+ \cap B_n^+) - \mathsf{P}(A_n^- \cap B_n^+) + \mathsf{P}(A_n^- \cap B_n^+) - \mathsf{P}(A_n^- \cap B_n^-) &= \\
\mathsf{P}((A_n^+ \setminus A_n^-) \cap B_n^+) + \mathsf{P}((B_n^+ \setminus B_n^-) \cap A_n^-) &\le \\
\mathsf{P}(A_n^+ \setminus A_n^-) + \mathsf{P}(B_n^+ \setminus B_n^-) &\to 0
\end{aligned}$$

Така множеството $A \cap B$ е "апроксимируемо" и съгласно лема A.1 $A \cap B \in \mathfrak{R}$, а следователно, и множествата $A \cup B, A + B, A \setminus B \in \mathfrak{R}$. $\mathfrak{R}$ е булова алгебра. От същите апроксимационни сметки следва, че μ е адитивна функция на $\mathfrak{R}$:

$$\mathsf{P}(A_n^-) + \mathsf{P}(B_n^-) \le \mu(A + B) \le \mu(A) + \mu(B) \le \mathsf{P}(A_n^+) + \mathsf{P}(B_n^+).$$

2. Да покажем сега, че $\mathfrak{R}$ е и σ-алгебра. Нека $\{B_n \in \mathfrak{R}, B_n \supset B_{n+1}\}$ е намаляваща редица. Да означим с $A_n = B_n \setminus B_{n+1}$. Да означим с $B = \cap B_n$. От монотонността следва, че съществува граница на намаляващата редица $\mu(B_n)$ и $\mu(B) \le \lim_n \mu(B_n)$. От полуадитивността следва, че $\mu(B) + \mu(\overline{B}) \ge 1$. От друга страна, от доказаната вече адитивност получаваме (за всяко n):

$$\sum_{i=1}^n \mu(A_i) = \mu(B_1) - \mu(B_{n+1}), \text{ т.е. } \sum_{i=1}^\infty \mu(A_i) = \mu(B_1) - \lim_n \mu(B_n).$$

Тъй като $\overline{B} = \overline{B_1} + \sum_{n=1}^\infty A_n$ ще получим: $\mu(\overline{B}) \le 1 - \mu(B_1) + \sum_{n=1}^\infty \mu(A_n) = 1 - \lim_n \mu(B_n)$. Като съберем двете неравенства получаваме: $\mu(B) + \mu(\overline{B}) \le 1$. Значи $B \in \mathfrak{R}$.

3. За да покажем, че μ е σ-адитивна, ще покажем, че е непрекъсната в $\emptyset$ и ще използваме лема 1.1. Нека $\{B_n \in \mathfrak{R}, B_n \supset B_{n+1}\}$ е намаляваща редица такава, че $\cap B_n = \emptyset$. Както и в предната част получаваме, че $\mu(B) \le \lim_n \mu(B_n^+) = \lim_n \mu(B_n^-)$. Но $\cap B_n^- \subset \cap B_n = \emptyset$. Тъй като $B_n^- \in \mathfrak{F}$ и $\mu = \mathsf{P}$, която е непрекъсната в $\emptyset$, получаваме $\mu(B) \le \lim_n \mathsf{P}(B_n^-) = 0$.

Тъй като $\mathfrak{F} \subset \mathfrak{R}$, то и $\sigma(\mathfrak{F}) \subset \mathfrak{R}$. Единствеността на така построената мярка върху $\sigma(\mathfrak{F})$ следва от нейната изчислимост от P. Лесно се получава и, че $\mathfrak{R}$ е пълна, т.е. съдържа подмножествата на множества с нулева горна мярка. $\square$

А.2 Проектори. Определения и свойства

В тази секция ще разгледаме проекторите – най-простите и най-използувани в математиката преобразования.

Определение А.3 *Нека P е преобразование на множеството X в себе си. Казваме, че P е проектор, ако $P^2 = P$. Понякога това свойство се нарича идемпотентност.*

От това следват следните свойства:

- Множеството X се разпада на две непресичащи се подмножества: $X = PX + \overline{PX}$;
- $PX \neq \emptyset$;
- ако $\overline{PX} = \emptyset$, т.е. $(PX = X)$, то $P = I$;
- траекториите на P съдържат най-много два различни елемента - единият е винаги в PX, другият - в $\overline{PX}$;

Теорема А.2 *Ако P и Q са комутиращи проектори ($PQ = QP$), то*

1. *PQ е проектор;*
2. *$(PQ)X = QX \bigcap PX$.*

Доказателство: $1.(PQ)(PQ) = PQ^2P = PQP = P^2Q = PQ$.
2. Следователно, $\forall x \in X$ е изпълнено: $QPx \in QX, PQx \in Px$. Т.е. $PQX \subset PX \bigcap QX$. Нека сега $p \in PX \bigcap QX$. Тогава $Pp = p, Qp = p$. Но тогава $PQp = p$, значи $p \in PQX$. Значи $QX \bigcap PX \subset PQX$. □

Теорема А.3 *Нека сега X е векторно пространство и P е линеен оператор ($P(ax) = aPx, P(x + y) = Px + Py$) и проектор. Тогава:*

1. *Образът PX е линейно подпространство;*
2. *ядрото $N_P = \{x : Px = 0\}$ е линейно подпространство;*
3. *$Q = I - P$ е проектор, $QP = PQ = 0$, $QX = N_P$;*
4. *за всеки комутиращ с P проектор Q имаме, че $P - QP$,$Q - QP$ са проектори.*

Доказателство: 1. Нека $x, y \in PX$. Тогава $P(ax + bx) = aPx + bPy = ax + by$.
2. Нека $x, y \in N_P$. Тогава $P(ax + bx) = aPx + bPy = 0$.
3. Нека $Qx = x - Px = (I - P)x$,$P(I - P) = P - P^2 = 0$. Тогава $Q^2 = (I - P)(I - P) = I - 2 * P + P = I - P$.
4. Ако $QP = PQ$, то $(I - P)Q = Q(I - P)$, $(I - Q)P = P(I - Q)$, $(I - P)(I - Q) = (I - Q)(I - P)$. □

Нека сега X е Хилбертово пространство и с (x, y) сме означили скаларното произведение в X.

Определение А.4 *Линеен самоспрегнат* ($(Px, y) = (x, Py)$) *проектор се нарича ортогонален.*

Теорема А.4 *За да бъде линеен проектор* P *в Хилбертовото пространство* X *самоспрегнат е необходимо и достатъчно* PX *и* $(I-P)X = N_P$ *да са ортогонални линейни подпространства:* $(Px, (I-P)y) = 0$;

Доказателство: Необходимост. $(Px, (I-P)y) = (x, P(I-P)y) = 0$;
Достатъчност. Следователно, $||P-H||^2 = sup_{||y||\leq 1}((P-H)x, y) = 0$. □

Пример А.1 *Не ортогонален проектор.*

Да разгледаме матрицата
$$Z = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$
като линеен оператор в $X = R^2$. Очевидно е, че $Z^2 = Z$. Но
$$ZX = \begin{pmatrix} x \\ 0 \end{pmatrix}, \quad N_P = \begin{pmatrix} x \\ -x \end{pmatrix}$$
и тези две едномерни подпространства не са ортогонални.

Теорема А.5 *Нека* Z *е линейно подпространство на* X. *За всяко* $x \in X$ *да означим с* $H : X \to Z$ *преобразованието*
$$x \longrightarrow \underset{z\in Z}{argmin}\, ||z - x||.$$
H *е ортогонален проектор.*

Доказателство: Очевидно преобразованието H е проектор. Трябва да покажем, че H е равен на своя спрегнат H^*, който се определя така: $\forall x, y \in X, (Hx, y) = (x, H^*y)$. Нека $y \in X$. Можем да го запишем във вида $y = (I-P)y + Py = y_1 + y_2$, където $y_1 =\in N_P, \quad y_2 \in PX$. Имаме:
$$\begin{aligned}((P-H)x, y) = ((P-H)x, y_1) + ((P-H)x, y_2) &= \\ (Px, y_1) - (Hx, y_1) + (Px, y_2) - (Hx, y_2) &= \\ (Px, y_1) - (x, Py_1) + (Px, y_2) - (x, Py_2) &= \\ (Px, (I-P)y) - 0 - (Px, Py) - (x, Py) &= \\ (Px, (I-P)y) + ((P-I)x, Py) &= 0.\end{aligned}$$

Следователно, $||P-H||^2 = sup_{||y||\leq 1}((P-H)x, y) = 0$. □

Литература

А.А.Боровков (1972). *Курс теории вероятностей.* Москва: Наука. 6.2

Гихман, И. и А. Скороход (1977). *Введение в теорию случайных процессов*. Москва: Наука. 15.1

Гнеденко, Б. (1965). *Курс теории вероятностей.* Москва: Наука. (document)

Димитров, Б. и Н. Янев (1990). *Теория на вероятностите и математическа статистика.* София: Наука и изкуство. (document)

Димитров, Б. (1980). *Марковски вериги.* София: Наука и изкуство. 16.1

Кемени, Д. и Д. Снелл (1970). *Конечные цепи Маркова.* Москва: Наука. 16.1

Мартин, Н. и Д. Инглenд (1988). *Математическая теория энтропии.* Москва: Мир. 8.3

Обретенов, А. (1974). *Теория на вероятностите.* София: Наука и изкуство. 9.4

Фелър, У. (1979). *Увод в теория на вероятностите и нейните приложения*, Том 1. София: Наука и изкуство. (document)

Индекс

Списък на илюстрациите

Означения

$(\Omega, \mathfrak{A}, \mathsf{P})$	-	Вероятностно пространство;
Ω	-	Множество от *елементарни събития*; достоверно събитие;
$\mathfrak{A}$	-	σ-алгебра от подмножества на Ω;
$\mathsf{P}(.)$	-	Вероятност определена на $\mathfrak{A}$;
$\mathsf{P}(.)$	-	Вероятност определена на $\mathfrak{A}$;
$A, B, \ldots, Z$	-	Множества,*събития* (елементи на $\mathfrak{A}$) или матрици;
$\overline{A}$	-	Допълнение на множеството; противоположно събитие;
$\xi, \eta, \ldots, \zeta$	-	*случайни величини (сл.в.)*;
γ, ν	-	*измерими разделяния*; *пълни групи от събития*;
$\emptyset$	-	Празно множество; невъзможно събитие;
$A \bigcap B$	-	сечение на множествата A и B; сбъдват се и двете събития;
$A \bigcup B$	-	обединение на множествата A и B; сбъдва се поне едното събитие;
$A + B$	-	обединение на несъвместими събития; сума на матрици;
AB	-	сечение на множествата A и B; произведение на матрици;
$A \perp\!\!\!\perp B, \xi \perp\!\!\!\perp \eta$	-	независими събития и сл.в.;
$\mathbf{E}\,, \mathbf{D}$	-	Математическо очакване и дисперсия;
$f(\cdot) : A \longrightarrow B$	-	Функция, дефинирана в множеството A със стойности в множеството B;
$R = R^1$	-	Реалната числова права;
R_+	-	Неотрицателните реални числа;
$A \times B$	-	Декартово произведение на множества;
$x = (x^1, \ldots, x^n)'$	-	n-мерен вектор (точка в) R^n;
$\lvert\lvert x \rvert\rvert$	-	Норма на $x \in R^n$;
$x'y$	-	Скаларно произведение на вектори;
$\exists$	-	Знак означаващ "съществува";
$\forall$	-	Знак означаващ "за всяко".
$\bar{x}$	-	$(1/n) \sum_{i=1}^n x_i$
$s^2(x)$	-	$(1/(n-1)) \sum_{i=1}^n (x_i - \bar{x})^2$
п.с.	-	почти сигурно, с вероятност 1
$f(x, \theta)$	-	функция на правдоподобие, плътност на наблюдавана сл.в.
$LL(x, \theta)$	-	$\log f(x, \theta)$.