Основе Вероватноће

Наход Вуковић
Факултет организационих наука
Универзитет у Београду

Београд, 2012.

Основе вероватноће
Проф. др Наход Вуковић

Рецензенти:
др Томислав Зечевић, професор
др Срђан Богосављевић, професор
Издаје:
Факултет организационих наука
Јове Илића 154, Београд
http://www.fon.bg.ac.rs/
e-mail: vukovic.nahod@fon.bg.ac.rs
Штампа:
Publish, Београд
Техничка припрема:
др Зоран Радојичић, Марина Доброта,
Факултет организационих наука

CIP – Каталогизација у публикацији
Народна библиотека Србије, Београд

51921(0758)

ВУКОВИЋ, Наход
Основе вероватноће / Наход Вуковић. -
Београд : Факултет организационих наука,
2012. (Београд : Publish) . – IV измењено издање,
316 стр., граф. прикази, табеле; 24 цм

Тираж 500. – Библиографија: стр. А-Ф
ISBN 978-86-7680-228-9

а) Теорија вероватноће
COBISS.SR-ID 133829132

Тираж: 500 примерака

ПРЕДГОВОР

Кад поседујемо довољно информација користимо **дедуктивни начин закључивања.** Међутим, у реалном свету смо најчешће у ситуацији поседовања некомплетних информација, а то значи да морамо користити неке методе **индуктивног закључивања** или **закључивања базираног на веродостојности**. Али које су то методе и шта ће бити њихова основа?

Неизвесност је израз који се вешто користи на раузличите начине у различитим областима укључујући филизофију, статистику, економију, финансије, осигуранје, психологију, инжињеринг и остале науке. Примењује се за предвиђање будућих догађаја, за објашњења већ обављених „мерња“ или за објашњење непознатог. **Вероватноћа** је главни инструмент за било какве закључке у овим условима.

Ризик је опште прихваћен термин дефинисан као мера неизвесност базирана на квантитативном изразу, који се назива **„вероватноћа“.** Формално, ризик је производ вероватноће да се неки догађај оствари и последица које би настале реализацијом тог догађаја.

Поставља се питање: Шта би требало научити и знати да би се у оваквим условима доносили ваљани и научно прихватљиви закључци?

Прво и највжније је усвојити Основе Вероватноће,

- ***као главни инструмент ваљаности залључивања,*** *затим*
- ***подручја активности у којима је неизвесност важна*** *и*
- ***веза између неизвесности, вероватноће и ризика.***

Ова књига је намењена свима који желе да буду успешни у својим активностима, било да су **менаџери**, **предузетници**, **инжињери** или **истраживачи** у областима **друштвених**, **природних** или **технуичких наука**. Стечено и усвојено знање биће им од велике помоћи за постизање успеха у свакодневним активностима. Овладаће довољним знањем о условима сталне присутности неизвесности и инструментима који ће им помоћи у доношењу правих одлука.

Ради бољег увида у најширу област примене теорије вероватноће, у првом делу књиге дат је преглед тзв. ***Дескриптивне статистике*** на основу које ће се увидети јасна потреба за усвајањем основних еленате Теорије вероватноће.

У било каквим статистичким испитивањима потребно је обавити веома велик број рачунања. Развојем речунаске технологије створени су услови да се то обави брзо и ефикасно. Зато је у два поглавља дат преглед и упутства за примену и коришћење најбољег статистичког пакете **SPSS.** У првом од њих дата су упуства за дескриптивну статистику, а у другом је описан поступак за одређивање вероватноћа из пробабилистичких модела.

Београд, Септембар, 2006. године



САДРЖАЈ

1. Основни појмови 1

1.1. Предмет статистичких истраживања 1

Статистички скуп 2

Статистичко обележје 3

Матрица података 4

Подела статистике 5

2. Дескриптивна статистика 7

2.1. Анализа релативним бројевима 7

Пропорција 8

Релативни показатељи 9

Индекси 10

2.2. Расподела фреквенција 14

Релативне фреквенције 17

Кумулативне фреквенције 19

2.3. Графичко приказивање расподеле фреквенција 21

2.4. Средње вредности 26

Аритметичка средина 27

Геометријска средина 36

Хармонијска средина 39

Модус и медијана 41

2.5. Мере варијабилитета..... 44
Размак варијације..... 45
Квартилна девијација..... 46
Средња девијација..... 49
Варијанса и стандардна девијација..... 53
Коефицијент варијације..... 58
2.6. Коефицијенти асиметрије и спљоштености..... 60

3. Случајни догађаји и вероватноћа 67
3.1. Сигурни, немогући и случајни догађаји..... 68
Скуп елементарних догађаја..... 69
Операције са случајним догађајима..... 73
3.2. Вероватноћа 80
Класична дефиниција вероватноће..... 80
Емпиријска вероватноћа..... 89
Аксиоматска дефиниција вероватноће..... 92
3.3. Услвоне вероватноће и независност..... 96
Формула потпуне (тоталне) вероватноће..... 98
Bayes-ова формула..... 102
Независни догађаји..... 106

4. Закон великих бројева 115
4.1. Бернулијева шема независних понављања експеримената 115
4.2. Гаусова крива 122

4.3. Бернулијев и Борел - Кантелијев закон великих бројева 127

4.4. Моавр-Лапласово решење бернулијевог проблема . 132

4.5. Пуасоново решење за ретке догађаје..... 136

5. Једнодимензионална случајна променљива.....139

5.1. Расподела случајних променљивих..... 140

Закон вероватноћа прекидне случајне променљиве .. 142

Закон вероватноћа непрекидне случајне променљиве149

Функција расподеле случајне променљиве 154

5.2. Анализа случајних променљивих..... 162

Очекивана вредност 163

Моменти случајне променљиве..... 170

Функција генератриса..... 174

Параметри случајних променљивих 177

Мере централне тенденције 177

Мере варијабилитета..... 179

Коефицијенти асиметрије и спљоштености ... 181

Чебишевљева теорема 181

5.3. Задаци 183

6. Вишедимензионалне случајне променљиве 189

6.1. Расподеле вишедимензионалних случајних променљивих 189

Закон вероватноћа дводимензионалних случајних променљивих 192

Закон вероватноћа к-димензионалне случајне променљиве 201
Маргинални закони вероватноћа 203
Условни закони вероватноћа 208
Независне случајне променљиве 213
6.2. Параметри расподела вишедимензионалних случајних променљивих 217
Очекивана вредност 218
Моменти случајних променљивих 220
Коефицијент корелације 224
Дисперсиона и корелациона матрица 233
Ранг дисперсионе и корелационе матрице 237
6.3. Регресиона анализа 241
Регресионе криве 241
Регресионе хиперповршине 248
Средње квадратне регресионе праве 250
Средње квадратне регресионе хиперравани 256
6.4. Задаци 259

7. Модели расподела случајних променљивих 267
7.1. Модели прекидних расподела 268
Биномна расподела 268
Пуасонова расподела 270
Геометријска расподела 272
7.2. Модели непрекидних расподела 274
Униформна расподела 274

Нормална расподела 276
Хи-квадрат расподела..... 280
Студентова расподела..... 283
Ф расподела 285
Дводимензионална Нормална расподела..... 287
Параметри дводимензионалне Нормалне расподеле. 291
Регресионе криве код Нормалне расподеле 294

8. Расподела неких функција случајних променљивих297

8.1. Линеарна функција..... 297
8.2. Квадратна функција..... 303
8.3. Студентов количник..... 306
8.4. Fisher – Snedekor-ов F – количник..... 309

ЛИТЕРАТУРА 311

1. ОСНОВНИ ПОЈМОВИ

1.1. ПРЕДМЕТ СТАТИСТИЧКИХ ИСПИТИВАЊА

Циљ научних испитивања је да се опишу појаве које нас окружују и да се утврде законитости које владају међу тим појавама. У целокупном нашем окружењу највећи број појава су појаве масовног карактера. То су појаве које могу бити описане једино преко великог броја информација нумеричког и атрибутивног својства које се односе на поједине елементе посматране појаве. На пример, ако се жели анализирати запосленост морају се посматрати сви запослени, ако се жели испитивати кретање становништва, мора се посматрати целокупно становништво, за испитивање производње у Србији морају се посматрати сви произвођачи и сл.

Статистичке методе се баве проучавањем оваквих појава, тј. проучавањем појава изражених преко великог броја нумеричких и ненумеричких података који садрже информације о посматраним појавама.

Ако желимо да испитамо продуктивност у Србији морамо знати колика је продуктивност сваког привредног учесника у Србији. Другим речима, морамо посматрати скуп привредних учесника и код сваког елемента тог скупа ,,измерити" продуктивност. Исто тако кад хоћемо да посматрамо кретање личних доходака запослених, морамо посматрати цео скуп запослених и за сваки његов елемент утврдити висину личног дохотка. На тај начин добијамо статистичке податке који постају предмет статистичких испитивања.

Статистичко испитивање вршимо на једном скупу елемената кога ћемо звати **статистички скуп** или **популација** и на сваком елементу тог скупа ,,меримо" одређено својство које ћемо звати **статистичко обележје**.

Статистички скуп

Елементи који имају једну заједничку особину чине један скуп. Али заједничка особина тих елемената мора бити тако дефинисана да потпуно и недвосмислено одређује да ли посматрани елемент припада датом скупу или му не припада. На елементима потпуно одређеног скупа приступа се испитивању или мерењу једног или више обележја. При том посматрана обележја могу имати различите вредности код елемената датог скупа.

На пример, на скупу пословних субјеката у Србији у датом тренутку можемо испитивати број запослених, просечна лична примања, остварени доходак и сл. Исто тако на скупу студената Београдског Универзитета можемо посматрати физичку издржљивост, успех на појединим предметима, број прочитаних књига итд.

Статистички скуп или популација је онај скуп на чијим елементима меримо једно или више обележја чије вредности варирају од елемента до елемента посматраног скупа.

Статистички скуп мора бити дефинисан просторно и временски.

Временско дефинисање статистичког скупа утврђује тренутак или раздобље којим ће се обухватити они елементи који ће бити предмет статистичких испитивања. Тако на пример, није довољно рећи: посматрамо скуп путничких аутомобила у Београду. Морамо додати и време. На пример: дана 31.6.2006. године.

Просторно дефинисање статистичког скупа подразумева означавање простора на коме се налазе елементи датог скупа. Ако посматрамо скуп саобраћајних удеса морамо рећи да ли је то скуп удеса на посматраном путу, у Србији или на простору Балкана и сл.

У статистичкој анализи се веома често испитује део статистичког скупа и на основу особина елемената тог дела доносе закључци о целом статистичком скупу или популацији. Такав део статистичког скупа назива се **узорак**. Основно подручје **Аналитичке статистике** представљају узроци и статистичка анализа базирана на њима.

Статистичко обележје

Под статистичким обележјем подразумева се оно својство елемената статистичког скупа које варира од елемента до елемента датог скупа. За сваки елемент статистичког скупа утврђује се ,,вредност" посматраног обележја.

Статистичка обележја се деле на ***квалитативна*** и ***квантитативна***.

Квалитативна обележја изражавају квалитативна својства елемената статистичког скупа. На пример: боја косе, школска спрема, пол, место рођења и сл. Свако квалитативно обележје има одређене категорије и елементи посматраног скупа сврставају се у одговарајуће категорије.

Квантитативна обележја изражавају квантитативна својства елемената статистичког скупа. Вредности таквих обележја су реални бројеви, а утврђивање тих вредности врши се „мерењем". На пример, код скупа домаћинстава можемо „измерити" број чланова домаћинстава, или код скупа становника Београда може се мерити висина или тежина сваког Београђанина.

Квантитативна обележја се деле на ***прекидна*** и ***непрекидна*** обележја.

Ако статистичко обележје може узети неку вредност из коначног скупа могућих вредности онда кажемо да је то **прекидно обележје.** На пример: број повреда на послу, реализована производња, број чланова домаћинства и сл.

Кад посматрано статистичко обележје може узети било коју вредност из коначног или бесконачног интервала, кажемо да је то **непрекидно обележје**. На пример: век трајања електронске цеви, висина студента, температура и сл.

Матрица података

Предмет изучавање СТАТИСТИКЕ су статистички скупови и обележја елемената статистичког скупа. У пракси су то велики скупови са великим бројем обележја. Например, скуп становника Београда, Србије, и сл. са обележјима: пол, старост, место рођења, школска спрема, економски статус, итд. Имамо најмање 2 милиона становника са најмање 20 обележја, а то је огромна маса података. Исто тако, скуп предузећа у Србији са свим подацима о њиховим карактеристикама и активностима, биће описан са огромном масом података.

Означимо са N број елемената статистичког скупа, а са $X_1, X_2, ..., X_k$ **k** обележја. Измерене вредности свих обележја не сваком елементу статистичког скупа можемо приказату у виду матрица реда Nxk, у којој ће врсте представљати елементе скупа, а колона одговарајуће вредности посматраних обележја. Тако ћемо добити **матрицу података**

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix}$$

са онолико врста колико има статистички скуп елемената и колона колико обележја меримо на елементима статистичког скупа. Ова матрица представља **СТАТИСТИЧКУ ГРАЂУ** и предмет је статистичких проучавања.

Основни проблем је непрегледност статистичких података, па је циљ да се прикупљени подаци **среде и учине прегледнијим**, а затим да се пређе на њихову **анализу**.

Подела статистике

Статистика или статистичке методе се могу поделити у две класе:

- **Дескриптивна статистика**
- **Аналитичка статистика.**

Под **дескриптивном статистиком** се подразумевају оне статистичке методе које се баве прикупљањем, сређивањем и анализом статистичких података. Њен циљ је да се опише измерена појава или процес **у домену мерења**.

Аналитичка статистика има за циљ да на основу мерења утврди општије законитости које важе и **у домену оног што није било мерено**. Наиме, за потпуно описивање појаве или процеса потребно је бесконачно много мерења. Пошто увек располажемо са резултатима коначног броја мерења морамо користити научне методе којима ћемо на егзактан начин описати комплетну појаву или процес.

На пример, ако се жели почети производња неког новог производа мора се утврдити време потребно за његову израду. Извршиће се одређен број мерења у предпроизводњи и методама дескриптивне статистике анализирати добијени резултати. Да би тачно одредили потребно време израде тог производа и у будућој производњи морамо користити методе аналитичке статистике којима ћемо на основу мерења у предпроизводњи „предвидети" време које ће бити потребно за тај производ у наредном периоду.

Исто тако, ако посматрамо неку економску категорију можемо методама дескриптивне статистике утврдити особине те категорије у домену мерења (на датом простору и у датом времену). Али за потпуно испитивање те категорије и њених законитости и изван домена мерења морамо се користити методама аналитичке статистике.

Основа аналитичке статистике је **теорија вероватноће** која математичким моделима описује случајне експерименте.

2. ДЕСКРИПТИВНА СТАТИСТИКА

2.1. АНАЛИЗА РЕЛАТИВНИМ БРОЈЕВИМА

Прикупљени статистички подаци представљају основу статистичких испитивања. Први задатак у тим испитивањима је припрема података за анализу и интерпретација добијених резултата. Другим речима, масу података морамо свести на мањи број нумеричких показатеља који ће бити погодни за анализу и интерпретацију, помоћу којих ћемо моћи да објаснимо посматране појаве или процесе везане за дати статистички скуп. Зато је наредно поглавље посвећено прегледу основних метода које се користе у овој фази укупних статистичких испитивања.

У многим истраживањима ова област је веома значајна и широко распрострањена. Због тога се појављује чак као и нова грана статистике позната под називом ***«Експлораторна анализа».***

Статистички подаци представљају апсолутне бројне износе или атрибутивне карактеристике појединих елемената статистичког скупа или на његове поједине делове. Међусобна поређења апсолутних бројева или карактеристика не дају нам праву слику о посматраној појави. У поређењима нас интересују релативни односи одговарајућих апсолутних износа који омогућавају поређења више различитих показатеља посматране појаве у статистичком скупу. Тако је на пример, далеко боље утврдити тзв. стопу незапослености него број незапослених. Исто тако пожељно је знати остварени доходак по запосленом, а не само укупан остварен доходак у предузећу.

У статистичкој пракси се користе три категорије релативних бројева: **пропорције, релативни показатељи** и **индекси**.

Пропорција

У анализи статистичких података који се односе на квалитативна својства статистичког скупа служимо се најчешће пропорцијама.Претпоставимо да својство X има k категорија, ($k \geq 2$) и да N_1 елемената статистичког скупа припада првој категорији, N_2 елемената припада другој категорији, ..., k-тој категорији припада N_k елемената посматраног скупа. Означићемо са N укупан број елемената статистичког скупа. Тада је

$$\sum_{i=1}^{k} N_i = N \ .$$

Количници

$$P_i = \frac{N_i}{N} \qquad (2.1)$$

i = 1,2, ..., k,

називају се **пропорције** *одговарајућих категорија својства X. Често се пропорције изражавају у процентима. Процентуално изражене пропорције се добијају ако се P_i помножи са 100.*

Податке о пропорцијама дајемо у виду табела. Тако је у Табели 2.1. дата структура запослених у Србији у 2006. години према сменама у којима раде запослени. Структура је дата преко одговарајућих пропорција.

Смена у којој је радио запослени	Број запослених	Структура	
		Пропорција	Пропорција у %
Прва смена	943.255	0,686	68,6
Друга смена	256.371	0,186	18,6
Трећа смена	176.267	0,128	12,8
У к у п н о	1.375.893	1,00	100,0

Табела 2.1. Запослено особље према сменама

Пропорције су саставни део статистичких табела публикованих података у скоро свим статистичким публикацијама.

Касније ће се видети како се пропорције могу користити и у анализи квантитативних обележја. Веома су погодне за поређења вредности обележја X на статистичким скуповима са великим бројем елемената.

Релативни показатељи

Многе се појаве могу изразити преко релативних односа два различита обележја. Тако се на пример, посматра густина насељености као количник броја становника и величине територије, доходак *per capita* (по главни становника) као показатељ оствареног дохотка земље, чист доходак по запосленом итд.

Релативни показатељ појаве је количник два обележја на посматраном статистичком скупу.

Изабрана обележја у релативним показатељима треба да су најважнија за објашњавање посматране појаве. Вредности тих обележја могу бити директно утврђена „мерењем" на елементима статистичког скупа. Али је често потребно те вредности кориговати и на тај начин добити боље показатеље дате појаве. На пример, у посматрању проблема заштите на раду, основни показатељ је број повреда. Бољи показатељ је број изгубљених радних сати због повреда. Ово обележје треба поредити са бројем запослених у организацији. Наравно, боље је поређење извршити према броју запослених у производњи (без администрације). Ради упоредивости ова два обележја треба израчунати укупан број радних сати свих запослених у производњи и са њим поредити број изгубљених сати због повреда. Тако ће се добити најбољи показатељи организованости заштите на раду који ће бити упоредив са таквим показатељима у различитим предузећима и у различитим временима.

За релативне показатеље је од изузетне важности база показатеља, а то је обележје са којим се дели друго обележје. Тако се избором базе може доћи до сасвим супротних закључака о посматраној

појави на основу истих вредности једног обележја X. У Табели 2.2. дати су подаци о саобраћајним несрећама у SAD за 1950. годину и 1964. годину.

Број погинулих	1950.	1964.	Стопе промена
Укупан број	34.763	47.700	+37
Према броју становника	23,0	24,9	+ 8
Према броју аутомобила	7,1	5,5	-23
Према броју пређених километара	7,6	5,7	-25

Извор: National Safety Council Accident Facts, 1965.

Табела 2.2. Број погинулих у аутомобилским несрећама

Из четврте колоне се види да је у 1964. години дошло до повећања броја погинулих за 37% у односу на 1950. годину. Такође је број погинулих према броју становника порастао за 8%. Међутим, код броја погинулих према броју аутомобила дошло је до смањења за 23%. Слично је и са бројем погинулих према укупном броју пређених километара.

Индекси

Један од важнијих задатака статистичких испитивања је праћење и анализа динамичких кретања друштвених, економских и природних појава. За овакве анализе подаци се прикупљају тако што се у датим временским интервалима врше мерења посматране појаве преко статистичких обележја и формирају се тзв. **временске серије** посматраних обележја.

За најједноставнију анализу таквих серија служимо се индексима који представљају показатеље промена посматраних обележја у датим временским интервалима (дневне промене, месечне, годишње и сл.).

Индекс *је број који представља количник вредности посматраног обележја X у тренутку t,* (X_t)*, и вредности тог обележја у неком другом тренутку t',* $(X_{t'})$*,*

$$I(t,t') = \frac{X_t}{X_{t'}} \tag{2.2}$$

Постоје две врсте индекса: ***ланчани*** и ***базни*** индекси.

Ланчани индекси *су количници вредности обележја X у тренутку t,* (X_t)*, и његове вредности у претходном тренутку мерења* $t-1, (X_{t-1})$*.*

Ако временска серија садржи податке о вредностима обележја X у тренуцима мерења t, при чему је

$$t \in [1,2,...,T],$$

тада су ланчани индекси дати низом бројева

$$\frac{X_2}{X_1}, \frac{X_3}{X_2}, ..., \frac{X_T}{X_{T-1}} \tag{2.3}$$

Најчешће се ови бројеви множе са 100 па се ланчани индекси изражавају процентуално. Низ ланчаних индекса (2.3) изражава узастопне промене обележја X. Означаваћемо их са

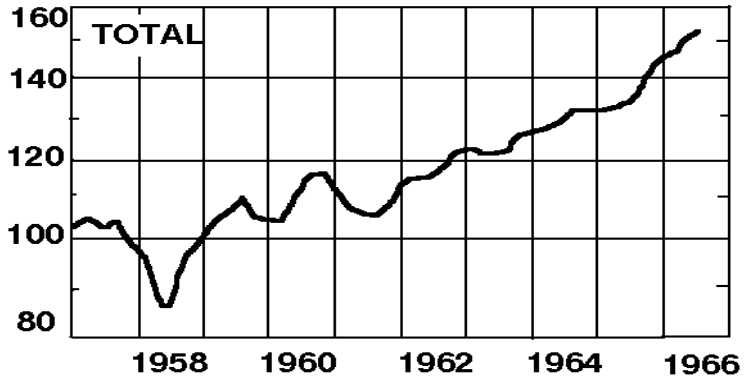
$$I_2, I_3, ...I_T.$$

За временску серију од T података добија се (T-1) ланчаних индекса јер се први податак у серији не може поредити са претходним.

Ако од процентуално израженог индекса у тренутку t одузмемо 100 добићемо процентуално изражену промену обележја X од момента (t—1) до момента t коју називамо ***стопа промене***. На пример, ако је I_t= 141,5 то значи да је обележје X порасло за 41,5%, а ако је I_t=92,3 онда се обележје X смањило за 7,7%.

Низ ланчаних индекса се може представити графички. На хоризонталној оси се наноси време (дани, месеци, године и сл.), а на вертикалној оси одговарајуће вредности ланчаних индекса. Добијене тачке се споје полигоналном линијом. Тако се добије графикон који на очигледан начин показује кретање посматраног обележја временске серије. На Слици 2.1. приказано је кретање индустријске производње преко месечних индекса у периоду 1958-1966. године.

Ланчани индекси се користе за анализу временских серија различитих обележја. На пример, ако се посматра кретање броја запослених и вредност производње онда се било каква анализа ова два обележја базира на поређењима одговарајућих ланчаних индекса датих временских серија.



Слика 2.1.Графикон кретања индустријске производње

Базни индекси *служе за исказивање промене посматраног обележја у временској серији у односу на један тренутак мерења са којим желимо да поредимо настале промене.*

Базни индекси су количници вредности обележја X у тренуцима мерења t, при чему је

$$t \in [1,2,...,T],$$

и вредности обележја X у тренутку t_o. Базни индекси изражени у процентима добијају се ако се количници помноже са 100.

Ако су вредности обележја X у посматраним тренуцима

$$X_1, X_2, ..., X_T \ ,$$

а X_0 његова вредност у тренутку t_0, тада низ количника

$$\frac{X_1}{X_0}, \frac{X_2}{X_0}, \ \cdots\cdots \frac{X_T}{X_0}$$

представља базне индексе посматране временске серије.

2.2. РАСПОДЕЛА ФРЕКВЕНЦИЈА

Посматраћемо статистички скуп од N елемената на којима ћемо испитивати квантитативно обележје X. Претпоставимо да је обележје X таквог карактера да на једном елементу статистичког скупа може узети саму једну вредност из скупа бројева

$$\{x_1, x_2, \ldots, x_k\}$$

при чему је k коначан али не сувише велики број. k представља број могућих различитих вредности посматраног обележја.

На пример, ако се посматра број чланова домаћинства на скупу домаћинстава у једном региону онда је скуп могућих вредности овог обележја скуп бројева

$$\{1,2,\ldots,10\}$$

Исто тако број смена у предузећима може узети једну од вредности

$$\{1,2,3\}$$

Код таквог обележја све његове вредности на посматраном скупу можемо изразити ***расподелом апсолутних фреквенција*** која се добија тако што се за сваку могућу вредност обележја x_i утврди колико елемената статистичког скупа узима ту вредност.

Означимо са f_i број елемената статистичког скупа код којих је $X = x_i$.

Тада се скуп парова вредности

$$(x_1, f_1), (x_2, f_2), \ldots, (x_k, f_k)$$

назива **расподела апсолутних фреквенција** *посматраног обележја X на датом статистичком скупу.*

Најчешће се, ради прегледности, расподела даје у виду табеле (Табела 2.2.).

Вредност обележја X (X)	Апсолутне фреквенције (*f*)
x_1	f_1
	f_2
x_2	
.	.
.	.
.	.
	f_k
x_k	
Укупно	**N**

Табела 2.2. Расподела обележја X

По правилу је статистички скуп велики, а број могућих вредности обележја X мали у односу на величину статистичког скупа.

Зато нам расподела фреквенција чини масу података прегледнијом, а исто тако и погоднијом за даљу анализу. Сад се уместо N вредности обележје X посматра k парова вредности (x_i, f_i) и при томе се не губи ништа у укупној количини информација која је садржана у тим подацима.

Претпоставићемо да је обележје X непрекидног типа и да може узети било који реалан број из интервала (a; b). Интервал (a; b) може бити коначан или бесконачан.

Поделићемо интервал (a; b) на k подинтервала

$$a_0\text{-}a_1,\ a_1\text{-}a_2,\ \ldots,\ a_{k-1}\text{-}a_k,$$

при чему је $a_0 = a,\ a_k = b$, *које ћемо звати* **групни интервали**.

На основу измерених вредности обележја X на елементима статистичког скупа одређује се расподела фреквенција тако што се за сваки групни интервал a_{i-1}-a_i. утврди број елемената статистичког скупа код којих вредност X упада у тај интервал, тј. код којих је

$$\mathbf{a_{i-1} < X \leq a_i;}$$

за свако $i \in \{1,2, ..., k\}$. Означимо број тих елемената са f_i. Добијена расподела фреквенција приказана је табеларно (Табела 2.3.).

Групни интервали	Апсолутне фреквенције
$a_0 - a_1$	f_1
$a_1 - a_2$	f_2
.	.
.	.
.	
$a_{k-1} - a_k$	f_k
Укупно	**N**

Табела 2.3. Расподела обележја X

Групни интервали у Табели 2.3. не морају бити једнаке дужине. Најчешће су први и последњи интервал највећи и са најмањим фреквенцијама. Дужине осталих (k-2) интервала зависе од њиховог броја и од појаве коју изражава обележје X. Одређивање броја групних интервала се врши веома различито.

Води се рачуна да превелик број интервала отежава анализу и чини податке непрегледним. Са друге стране, превише мали број интервала не садржи довољну количину информација о статистичком скупу.

При одређивању групних интервала треба се придржавати следећег принципа: ***одредити број и дужине групних интервала тако да они садрже максималну количину информација уз њихов минималан број.***

У циљу одређивања оптималног броја k групних интервала за статистички скуп од N елемената, често се користи **Sturges**-ово правило:

$$k = 1 + 3.3 \log N$$

Релативне фреквенције

Расподела апсолутних фреквенција је подесна за анализу и посматрање појединих обележја одвојено.

Међутим, ако треба поредити обележје X на два или више статистичких скупова, онда апсолутне фреквенције не одражавају упоредиве реалне износе.

Исто тако, ако треба поредити обележје X са неким другим обележјем Y тада нам апсолутне фреквенције не омогућавају таква поређења. Зато је потребно апсолутне бројеве претворити у релативне и анализу вршити на основу тзв. релативних фреквенција**.**

Претпоставимо да обележје X има неку од расподела апсолутних фреквенција датих табелама: Табела 2.2. и Табела 2.3. За сваки групни интервал и за сваку вредност обележја X одредимо релативну фреквенцију као количник.

$$p_i = \frac{f_i}{N}$$

при чему је $i \in \{1, 2, ..., k\}$, а са p_i смо означили вредности релативне фреквенције. На тај начин се добија расподела релативних фреквенција као у Табели 2.4.

Вредност обележја X (X)	Апсолутне фреквенције (f)	Релативне фреквенције (p)	Релативне фреквенције у процентима (%)
$a_0 - a_1$	f_1	$p_1 = \frac{f_1}{N}$	$p_1 \cdot 100$
$a_1 - a_2$	f_2	$p_2 = \frac{f_2}{N}$	$p_2 \cdot 100$
.	.	.	.
.	.	.	.
.		.	.
$a_{k-1} - a_k$	f_k	$p_k = \frac{f_k}{N}$	$p_k \cdot 100$
Укупно	**N**	**1,00**	**100,00**

Табела 2.4. Расподела релативних фреквенција[1)]

Релативне фреквенције су бројеви који задовољавају следећа два услова:

1) $p_i \geq 0$, за свако $i \in \{1, 2, ..., k\}$ и

2) $p_1 + p_2 + ... + p_k, = 1$, tj. $\sum_{i=1}^{k} p_i = 1$.

Релативне и апсолутне фреквенције се најчешће дају у једној истој табели као што је то у Табели 2.4 приказано. Процентуално изражавање релативних фреквенција чини их погоднијим за поређења и за графичко представљање.

[1)] Ако је обележје X прекидног типа онда у првој колони уместо групних интервала стоје одговарајуће вредности тог обележја.

Расподеле фреквенција су саставни део публикација Статистичког система и често чине полазну основу истраживања друштвених, економских или природних процеса.

Кумулативне фреквенције

У анализама неког обележја X на статистичком скупу често нас интересује одговор на следеће питање: Колико има елемената у статистичком скупу код којих обележје X има вредност мању од неког унапред датог броја x?

На пример, интересује нас колико има радника са личним дохотком мањим од 4000 дин, или колико има општина са националним дохотком по глави становника мањим од просечног националног дохотка у Републици Србији и сл.

Да би лакше дошли до одговора на оваква питања служимо се тзв. кумулативним фреквенцијама које се одређује на основу апсолутних или релативних фреквенција.

Из расподеле апсолутних фреквенција вредност кумулативне функције за i-ти групни интервал добија се кад се све фреквенције за претходне интервале саберу. Тако ће се добити да је

$$c_i=f_1+f_2+\ldots+f_i,$$

за $i\in\{1, 2, \ldots, k\}$, *а* c_i *означава* **кумулативну фреквенцију** *i-тог групног интервала.*

Кад се кумулативне фреквенције c_i поделе са бројем елемената статистичког скупа добиће се релативне кумулативне фреквенције које означавамо са F_i и које су дате са

$$F_i = \frac{c_i}{N}$$

за свако $i\in\{1, 2, \ldots, k\}$.

Расподела кумулативних фреквенција приказана је табеларно (Табела 2.5.)

Групни интервали (X)	Апсолутне фреквенције (f)	Кумулативне фреквенције (c)	Релативне кумулативне фреквенције (F)
$a_0 - a_1$	f_1	$c_1 = f_1$	$F_1 = \frac{c_1}{N}$
$a_1 - a_2$	f_2	$c_2 = f_1 + f_2$	$F_2 = \frac{c_2}{N}$
.	.	.	.
.	.	.	.
.	.	.	.
$a_{k-1} - a_k$	f_k	$c_k = f_1 + f_2 + ... f_k$	$F_k = \frac{c_k}{N} = 1$
Тотал	**N**		

Табела 2.5. Расподела кумулативних фреквенција

Релативне кумулативне фреквенције се могу одредити и из релативних фреквенција. Наиме, из израза за C_j и F_i види се да је

$$F_i = \frac{f_1}{N} + \frac{f_2}{N} + ... + \frac{f_i}{N},$$

односно

$$F_i = \sum_{j=1}^{i} p_j,$$

за свако $i \in \{1, 2, ..., k\}$.

Релативне кумулативне фреквенције

$$F_1, F_2, ..., F_k$$

чине један неопадајући низ реалних бројева из интервала [0,1].

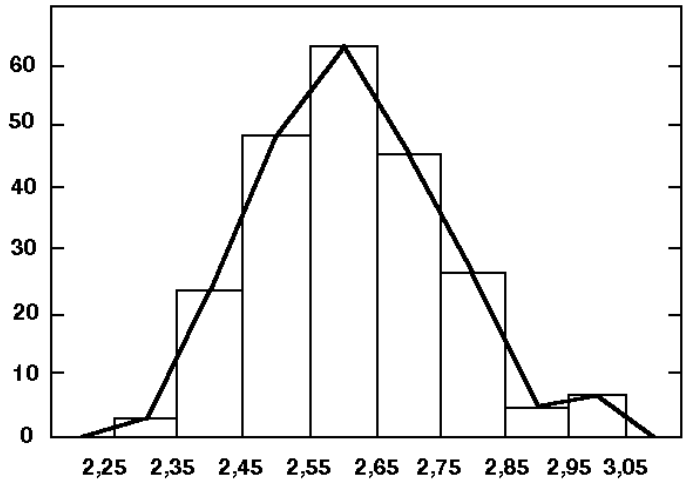
2.3. ГРАФИЧКО ПРИКАЗИВАЊЕ РАСПОДЕЛА ФРЕКВЕНЦИЈА

Да би лакше уочили и истакли главне карактеристике расподеле фреквенција неког обележја X на посматраном статистичком скупу служимо се графиконима.

Графиконе расподела фреквенција конструишемо тако што на хоризонталној оси наносимо вредности обележја X (ако је X прекидног типа) или границе групних интервала (ако је X непрекидног типа). Уместо граница на оси X се могу наносити средине групних интервала. Код графикона апсолутних фреквенција на оси Y наносимо вредности одговарајућих фреквенција за групни интервал или за вредности X и конструишемо правоугаонике изнад групних интервала са висинама једнаким фреквенцијама. Ако тачке које одговарају срединама групних интервала и фреквенција спојимо дужима добићемо тзв. ***полигон апсолутних фреквенција.*** Пример: Табела 2.6. и Слика 2.2.

(X)	(f)	(P)	(C)
2,25 – 2,35	2	0,01	2
2,35 – 2,45	23	0,11	25
2,45 – 2,55	49	0,23	74
2,55 – 2,65	63	0,29	137
2,65 – 2,75	45	0,21	182
2,75 – 2,85	25	0,12	207
2,85 – 2,95	3	0,01	210
2,95 – 3,05	4	0,02	214
Укупно	**214**	**1,00**	

Табела 2.6. Расподела обележја X

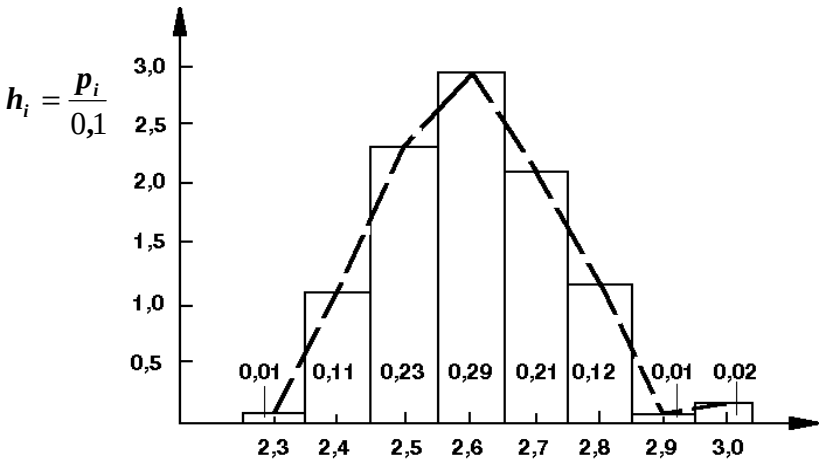


Слика 2.2. Полигон и расподела фреквенција обележја X

За нас су најзанимљивији тзв. ***хистограми*** који служе за приказивање расподела релативних фреквенција. Конструишемо их тако што на оси X наносимо средине групних интервала, а затим око тих тачака конструишемо правоугаонике чије ће површине бити једнаке вредностима релативних фреквенција.

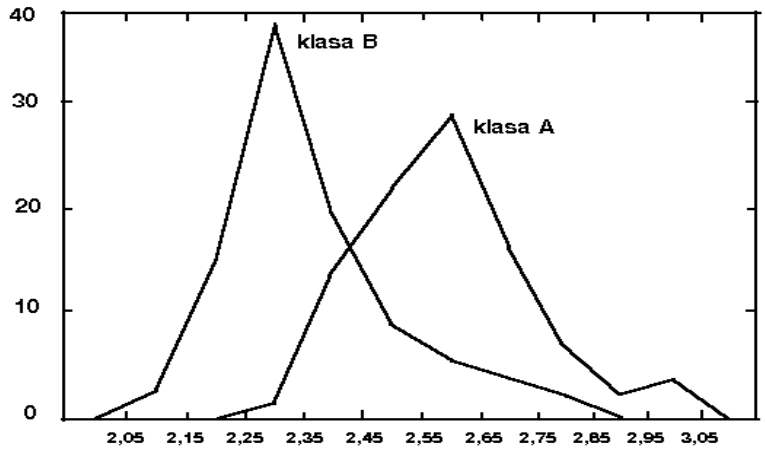
Ако је дужина i-тог интервала $d_i = a_i - a_{i-1}$, онда висина правоугаоника изнад тачке $x_i = \frac{a_i + a_{i-1}}{2}$, треба да је $h_i = \frac{p_i}{d_i}$, при чему i = 1,2, ...,k, а p одговарајућа релативна фреквенција.

Тако се на пример за податке из табеле 2.6. добија хистограм дат на слици 2.3.



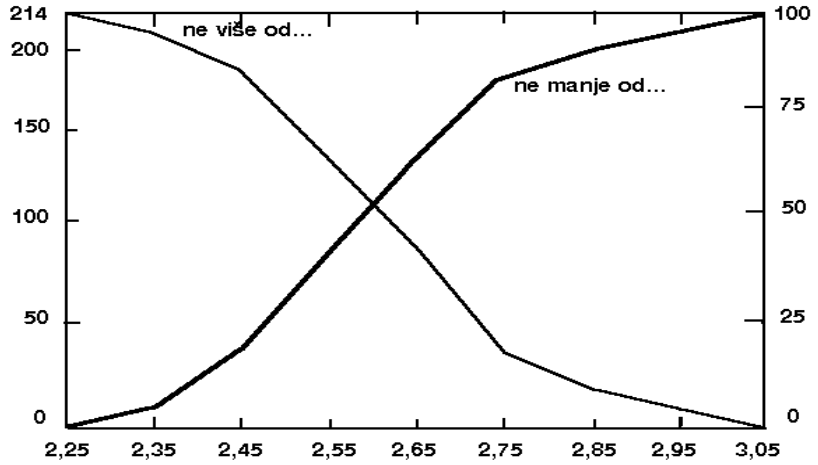
Слика 2.3. Хистограм релативних фреквенција

Хистограм је веома погодан за поређење расподела фреквенција једног обележја на више статистичких скупова. Тако се на пример са Слике 2.4, на којој су дати хистограми обележја X – две класе A и B, види да постоје разлике у расподелама. У класи A веће вредности обележја X су код великог броја елемената посматраног скупа, а у класи B је обрнуто.



Слика 2.4. Расподела обележја X на два скупа

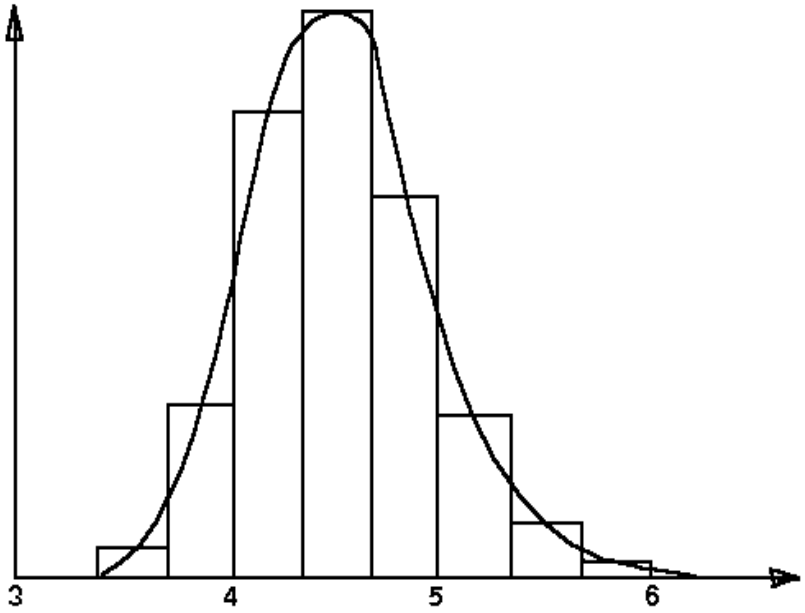
Кумулативне фреквенције се такође могу приказати и графички као што је дато на Слици 2.5. на којој су приказани подаци из Табеле 2.6.



Слика 2.5. Графикон кумулативних фреквенција

На основу хистограма који одговара расподели релативних фреквенција непрекидног обележја X, може се изабрати крива линија која ће апроксимирати посматрану расподелу фреквенција. Та крива линија се зове **закон вероватноће** *или* **функција густине** *и представља гранични облик хистограма ако се број елемената у статистичком скупу неограничено повећава, а дужине групних интервала на основу којих правимо хистограм неограничено смањују.*

Такав један закон вероватноћа дат је на Слици 2.6. Одређивање закона вероватноћа на основу података о вредностима у коначном статистичком скупу представља један од основних проблема аналитичке статистике. Кад се утврди тачно закон вероватноћа који одговара датим мерењима може се сматрати да су успешно решени скоро сви проблеми везани за одлучивања на основу посматраног обележја.



Слика 2.6. Закон вероватноћа

2.4. СРЕДЊЕ ВРЕДНОСТИ

Користећи релативне бројеве и расподелу фреквенција може се стећи известан глобални утисак о посматраној појави и посматраном статистичком скупу. Ипак за даљу и сврсисходнију анализу потребне су нам прецизније методе којима ћемо масу статистичких података обрадити тако да постане употребљива у процесу доношења одлука.

Анализу статистичких података можемо вршити тако што ћемо дефинисати извесне показатеље или параметре, чије ће нам вредности изражавати одређене сумарне карактеристике датих података. Вредности утврђених параметара омогућиће доношење закључака о одређеној појави или процесу који су изражени посматраним подацима.

Прва група таквих параметара су тзв. ***средње вредности*** или ***просеци*.** Веома често се користе и у свакодневном животу. Наиме, сваком је познато шта значи просечан лични доходак, просечна стопа раста, или просечна продуктивност и сл. Ови параметри показују неку централну вредност посматраног обележја X на елементима статистичког скупа. Пожељно је да средње вредности имају следеће особине:

- Ако су све вредности посматраног обележја X на статистичком скупу међусобно једнаке, онда и њихова средња вредност треба да је једнака тој вредности;
- У датом статистичком скупу постоји најмања и највећа вредност посматраног обележја X. Средња вредност треба да је већа од најмање, а мања од највеће вредности обележја X, и
- Средња вредност треба да зависи од свих вредности обележја X на целом статистичком скупу.

У ову групу параметара спадају: аритметичка, хармонијска и геометријска средина, затим модус и медијана.

Аритметичка средина

У свакодневном животу највише се користи аритметичка средина као средња вредност. Зато се често под појмом „просек" мисли на аритметичку средину.

Аритметичка средина *низа бројева је број који се добије кад се њихов збир подели са укупним бројем чланова тог низа.*

Основни метод израчунавања аритметичке средине је исти за негруписане и груписане податке, а формуле се мало разликују.

Посматрајмо статистички скуп од N елемената на којима испитујемо квантитативно обележје X. Нека су утврђене следеће вредности обележја X код елемената статистичког скупа:

$$X_1, X_2, ..., X_N.$$

Тада је аритметичка средина обележја X једнака

$$\overline{x} = \frac{1}{N}\left(X_1 + X_2 + \ldots + X_N\right)$$

или краће написано

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (2.4)$$

при чему је $\overline{x}$ („икс бар") ознака за аритметичку средину обележја X.

Ако је обележје X прекидног типа са k могућих различитих вредности

$$x_1, x_2, ..., x_k$$

тада су вредности обележја X на статистичком скупу дате расподелом фреквенција,

X	x_1 x_2 … x_k
f	f_1 f_2 … f_k

при чему је

$$\sum_{i=1}^{k} f_i = N$$

Аритметичка средина овог обележја је

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{k} x_i \cdot f_i \qquad (2.5)$$

јер у изразу (2.4) међу бројевима X_1, X_2, ..., X_N који одговарају вредностима X на елементима статистичког скупа, f_1 ових бројева имаће вредност x_1, f_2 ће имати вредност x_2, итд., f_k ће имати вредност x_k, па у њиховом укупном збиру треба свако x_i помножити са f_i. Тако се добија формула дата изразом (2.5).

И коначно претпоставићемо да је обележје X непрекидног типа и да су његове вредности на статистичком скупу дате расподелом фреквенција, при чему је

X	$a_0 - a_1$	$a_1 - a_2$	...	$a_{k-1} - a_k$
f	f_1	f_2	...	f_k

$$\sum_{i=1}^{k} f_i = N \qquad (2.6)$$

Да би тачно одредили аритметичку средину за ово обележје X сматраћемо да сви елементи статистичког скупа код којих обележје X упада у један групни интервал, имају исту вредност. Најприхватљивије је да за ту вредност узмемо средину одговарајућег интервала.

Зато ћемо у случају кад је дужина првог и последњег групног интервала приближно једнака дужини осталих интервала, i-том интервалу придружити вредност обележја x_i која је једнака

$$x_i = \frac{a_{i-1} + a_i}{2} \qquad (2.7)$$

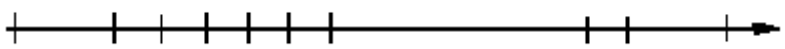
за i = 1,2, ..., k.

Ако се први и последњи групни интервал знатно разликују по дужини од осталих интервала, или ако је неки од њих бесконачан, тада ћемо за x_1 и x_k узетих вредности

$$x_1 = a_1 - \frac{\overline{d}}{2}, \quad x_k = a_{k-1} + \frac{\overline{d}}{2}, \tag{2.8}$$

где је $\overline{d}$ просечна дужина преосталих (k-2) интервала. Тако ће се добити следећи низ тачака на реалној оси.

$$x_1 = a_1 - \frac{\overline{d}}{2} \qquad x_2 \qquad x_3 \qquad x_k = a_{k-1} + \frac{\overline{d}}{2}$$



$$a_0 = a \;\; a_1 \;\; a_2 \;\; a_3 \qquad a_{k-1} \;\; a_k = b$$

Аритметичка средина обележја X датог расподелом (2.6) једнака је

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{k} x_i \cdot f_i \tag{2.9}$$

при чему су вредности x_i одређене преко (2.7) и (2.8).

Аритметичку средину можемо одредити и из расподеле релативних фреквенција. С обзиром да је

$$p_i = \frac{f_i}{N}$$

за свако $i \in (1,2, ..., k)$, то се израз (2.5), односно (2.6) за аритметичку средину може написати и у облику

$$\overline{x} = \sum_{i=1}^{k} x_i \cdot p_i \tag{2.10}$$

Аритметичка средина има све три особине које су пожељне за једну средњу вредност. Заиста

1. Ako су све вредности obeležja X на посматраном скупу једнаке, тј. ако је за свако i=1,2,...,N $X_i = c$, тада је

$$\bar{x} = \frac{1}{N} N \cdot c = c$$

   што значи да $\bar{x}$ има особину 1;

2. Нека је x најмања вредност обележја X у посматраном статистичком скупу. Лако је проверити да је $\bar{x} > x$. Ако је x' највећа вредност обележја X у посматраном скупу, тада је $\bar{x} < x'$, а то значи да $\bar{x}$ има особину 2, тј. већа је од најмање, а мања од највеће вредности посматраног обележја у датом скупу;

3. Из израза (2.4) којим је дефинисана аритметичка средина види се да постоји директна функционална зависност од свих вредности обележја X.

Поред ових особина аритметичка средина има и неке друге особине које ће нам помоћи да лакше интерпретирамо добијене вредности овог параметра. Исто тако неке од њих ће нам послужити за лакша рачунања аритметичке средине.

Особина 1. *Збир одступања вредности обележја X на посматраном скупу од њихове аритметичке средине је једнако нули, тј.*

$$\sum_{i=1}^{N} (X_i - \bar{x}) = 0 \qquad (2.11)$$

односно

$$\sum_{i=1}^{k} (x_i - \bar{x}) f_i = 0$$

кад је X дато расподелом фреквенција.

Доказ. Растављањем на два збира леве стране израза (2.11) добиће се

$$\sum_{i=1}^{N}(X_i-\overline{x})=\sum_{i=1}^{N}X_i-N\cdot\overline{x}.$$

Са друге стране из израза којим је дефинисана аритметичка средина добија се

$$\sum_{i=1}^{N}X_i=N\cdot\overline{x}$$

тако да је

$$\sum_{i=1}^{N}(X_i-\overline{x})=N\cdot\overline{x}-N\cdot\overline{x}=0 \qquad (2.12)$$

а то је и требало проверити.

Особина 2. *Збир квадрата одступања вредности обележја X на елементима статистичког скупа од било ког броја* a *је најмањи ако је* $a=\overline{x}$ *, mj.*

$$\sum_{i=1}^{N}(X_i-a)^2=\text{minimum} \qquad (2.13)$$

за $a=\overline{x}$.

Доказ. Означимо са $F(a)$ функцију која представља леву страну израза (2.13). Да би одредили минимум функције $F(a)$ можемо користити изводе. Али ћемо овде то урадити једноставније. Функција $F(a)$ је

$$F(a)=\sum_{i=1}^{N}(X_i-a)^2 \qquad (2.14)$$

Додавањем и одузимањем $\overline{x}$ у малој загради у функцији F(a), квадрирањем и растављањем на три збира добиће се

$$F(a)=\sum_{i=1}^{N}(X_i-\overline{x})^2-2\sum_{i=1}^{N}(X_i-\overline{x})(\overline{x}-a)+\sum_{i=1}^{N}(\overline{x}-a)^2$$

односно

$$F(a)=\sum_{i=1}^{N}(X_i-\overline{x})^2-2(\overline{x}-a)\sum_{i=1}^{N}(X_i-\overline{x})+N(\overline{x}-a)^2 \qquad (2.15)$$

Из особине 1. следи да је други члан у (2.15) једнак нули тако да је

$$F(a)=\sum_{i=1}^{N}(X_i-\overline{x})^2+N(\overline{x}-a)^2 \qquad (2.16)$$

Први члан на левој страни израза (2.16) не зависи од a и има ненегативну вредност. Други члан је такође ненегативан и зависи од a. Минималну вредност функција $F(a)$ ће имати за оно a за које други члан има најмању вредност. То ће бити у случају кад је $\overline{x}-a=0$, односно $\overline{x}=a$.

Особина 3. *Ако између обележја X и Y која се испитују на једном статистичком скупу, постоји линеарна зависност, тада постоји иста таква зависност и између њихових аритметичких средина, тј. ако је*

$$Y=aX+b \qquad (2.17)$$

тада је

$$\overline{y}=a\overline{x}+b \qquad (2.18)$$

при чему су a и b реални бројеви ($a\neq 0$), $\overline{x}$ и $\overline{y}$ аритметичка средина обележја X, односно Y.

Доказ. Из линеарне зависности (2.17) следи да је за сваку вредност обележја X_i, одговарајућа вредност обележја Y_i једнака

$$Y_i=aX_i+b \qquad (2.19)$$

Аритметичка средина обележја Y је

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

одакле се после замене (2.19) добија

$$\bar{y} = \frac{1}{N}\left\{\sum_{i=1}^{N} aX_i + N \cdot b\right\}$$

односно

$$\bar{y} = a \cdot \frac{1}{N}\sum_{i=1}^{N} X_i + b \tag{2.20}$$

Пошто је

$$\frac{1}{N}\sum_{i=1}^{N} X_i = \bar{x}$$

то значи да је из израза (2.20)

$$\bar{y} = a\bar{x} + b$$

а то је и требало проверити.

Особина 4. *Претпоставимо да на једном статистичком скупу од N_1 елемената обележје X има аритметичку средину $\bar{x}_1$, а на другом скупу од N_2 елемената има аритметичку средину $\bar{x}_2$. Тада је аритметичка средина за X на оба статистичка скупа заједно једнака тзв.* ***пондерисаној аритметичкој средини***

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2} \tag{2.21}$$

Доказ. Ако са $X_1, X_2, ..., X_{N_1}$, односно $X_1', X_2', ..., X_{N_2}'$ означимо вредности обележја X на елементима првог, односно другог скупа, аритметичка средина оба скупа биће једнака

$$X = \frac{1}{N_1 + N_2}\left\{\sum_{i=1}^{N_1} X_i + \sum_{i=1}^{N_2} X_i^{'}\right\} \tag{2.22}$$

Пошто је

$$\sum_{i=1}^{N_1} X_i = N_1 \cdot \overline{x}_1 \text{ i } \sum_{i=1}^{N_2} X_i^{'} = N_2 \cdot \overline{x}_2 \tag{2.23}$$

заменом у израз (2.22) добиће се (2.21), односно особина 4.

Особина 5. *Посматрајмо s статистичких скупова са N_1, N_2,...N_S, елемената респективно и са аритметичким срединама $\overline{x}_1$, $\overline{x}_2$, ... $\overline{x}_s$. Тада је аритметичка средина обележје X на свим скуповима заједно једнака пондерисаној аритметичкој средини*

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{s} N_i \cdot \overline{x}_i \tag{2.24}$$

при чему је N укупан број елемената свих статистичких скупова.

Доказ. Тачност израза (2.24) може се проверити на основу особине 4. користећи принцип математичке индукције.

Наведене особине аритметичке средине омогућавају у извесним случајевима знатно лакша и тачнија рачунања. Ако је обележје X непрекидног типа и дато расподелом фреквенција

X	$a_0 - a_1$	$a_1 - a_2$	…	$a_{k-1} - a_k$
f	f_1	f_2		f_k

тада се погодним избором два броја A и d може увести смена

$$Y = \frac{X - A}{d} \tag{2.25}$$

тако да се уместо бројева $x_1, x_2, ..., x_k$ добију мањи бројеви $y_1, y_2, ..., y_k$ који ће бити погодни за рачунање.

Кад су групни интервали једнаки, онда за d треба узети дужину групних интервала, а за A узети средину оног интервала који има највећу фреквенцију. У том случају вредности за Y ће бити цели бројеви 0, ±1, ±2,... па је лако за Y срачунати његову аритметичку средину. Из релације (2.25) и Особине 3. следи да је аритметичка средина X једнака

$$\bar{x} = d\bar{y} + A$$

У пракси су чести случајеви кад нам је позната средина једног обележја, а желимо да одредимо средину неког другог обележја.

На пример, ако је просечни нето лични доходак у једном предузећу $\bar{x} = 5000$ динара, какав би био просечан лични доходак после повећања? Нека је по завршном рачуну прихваћен предлог за 8% повећања и плус 300 динара за сваког радника за топли оброк. Колики ће бити просечан лични доходак? Ако је X садашњи лични доходак, а Y нови лични доходак, тада је

$$Y = 1{,}05X + 300$$

те се из *Особине 3*. добија да је

$$\bar{y} = 1{,}05\bar{x} + 300$$

односно

$$\bar{y} = 5350 \text{ динара}$$

То је просечни лични доходак после повећања. Као што се види, за његово одређивање није нам било потребно рачунање личних доходака свих радника, па на основу тих рачунања утврђивање просека.

Често се статистички скупови састоје из тзв. ***стратума.*** *Ако је цео скуп подељен на подскупове који се међусобно искључују, онда кажемо да је то стратификовани статистички скуп, а одговарајући подскупови су стратуми.*

За стратификовани скуп статистички подаци се прикупљају одвојено по стратумима и врши се њихова обрада такође по

стратумима. Анализа целог статистичког скупа изискује додатна рачунања на основу свих података, а то је некад и немогуће урадити. Зато се морају знати методе за обраду података које ће омогућити да се на основу резултата анализе по стратумима дође до резултата за цео скуп.

Геометријска средина

У анализама временских серија најпогоднија средња вредност је геометријска средина.

Геометријска средина *низа бројева је N-ти корен из производа његових чланова. Нека су* X_1, X_2, ..., X_N *вредности посматраног обележја X на елементима статистичког скупа. Геометријска средина обележја X је једнака*

$$G=\sqrt[N]{X_1 \cdot X_2 \cdot ... \cdot X_N} \tag{2.26}$$

Кад је обележје X дато расподелом фреквенција,

X	x_1	x_2	…	x_k
f	f_1	f_2	…	f_k

геометријска средина је

$$G=\sqrt[N]{x_1^{f_1} \cdot x_2^{f_2} \cdot ... \cdot x_k^{f_k}} \tag{2.27}$$

јер се свака вредност обележја x_i појављује f_i пута.

Да би одредили геометријску средину за свако N вредности обележја X морају бити позитивне. Зато је и употреба геометријске средине ограничена само на она обележја која су позитивна.

Логаритмовањем израза (2.26) добиће се

$$\log G = \frac{1}{N} \sum_{i=1}^{N} \log X_i \tag{2.28}$$

што значи да је логаритам геометријске средине једнак аритметичкој средини логаритама вредности обележја X на елементима статистичког скупа.

На сличан начин се и за обележје X које је дато расподелом фреквенција из (2.27) добија

$$\log G = \frac{1}{N} \sum_{i=1}^{k} f_i \cdot \log x_i \qquad (2.29)$$

Из расподеле релативних фреквенција добиће се да је логаритам геометријске средине једнак

$$\log G = \sum_{i=1}^{k} p_i \cdot \log x_i \qquad (2.30)$$

при чему је

$$p_i = \frac{f_i}{N}$$

за i∈ {1,2, ...,k}.

Све особине аритметичке средине поседује и логаритам геометријске средине јер је то такође једна аритметичка средина.

За временску серију обележја X

$$X_1, X_2,...,X_T$$

ланчани индекси чине низ бројева

$$\frac{X_2}{X_1}, \frac{X_3}{X_2}, ..., \frac{X_T}{X_{T-1}}$$

који показују узастопне промене обележја X. Да би добили средњу вредност тих промена за цео период T, одредићемо геометријску средину овог низа

$$G = \sqrt[T-1]{\frac{X_2}{X_1} \cdot \frac{X_3}{X_2} \cdot \ldots \cdot \frac{X_T}{X_{T-1}}} \qquad (2.31)$$

После скраћивања под кореном у изразу (2.31) добија се

$$G = \sqrt[T-1]{\frac{X_T}{X_1}} \qquad (2.32)$$

одакле се види да геометријска средина зависи од тзв. ***стопе раста*** у посматраном интервалу. Њено рачунање је доста једноставно. Довољно је познавати вредност обележја X на почетку и на крају временске серије.

Геометријска средина је погодна за одређивање просечних промена са непотпуним подацима, тј. за серије у којима не знамо вредности обележја X за све године. Тако се на пример на основу пописа становништва добијају вредности у годинама пописа, али се на основу тих вредности морају извршити процене вредности и за године између пописа. Прва анализа таквих података укључује у себе одређивање просечне стопе раста.

У Табели 2.7. дати су резултати пописа становништва бивше СФРЈ и одговарајуће геометријске средине.

Година пописа	Број становника (X) (у мил.)	(log X)	Логаритми ланчаних индекса	Геометријска средина
1948	15842	4,199814	-	-
1953	16991	4,230216	0,030402	1,014
1961	11549	4,268318	0,038102	1,011
1971	20523	4,312243	0,043925	1,010

Табела 2.7. Рачунање просечних стопа раста старовништва бивше СФРЈ

Из последње колоне Табеле 2.7 види се да је просечна стопа раста у периоду 1948-1953. године била 1,4%, у периоду 1953-1961. године 1,1% а у периоду 1961-1971. године 1,0%.

Хармонијска средина

За анализу оних појава чији је интензитет обрнуто пропорционалан вредностима посматраног обележја, најпогоднија средња вредност је хармонијска средина. На пример, при испитивању продуктивности мери се време потребно за израду неког производа, али је продуктивност обрнуто пропорционална том времену, па зато у овом случају треба користити хармонијску средину као средњу вредност.

Хармонијска средина *низа бројева је реципрочна вредност аритметичке средине реципрочних вредности чланова тог низа.*

Ако су

$$X_1, X_2,...X_N$$

вредности обележја X на N елемената статистичког скупа тада је хармонијска средина једнака

$$H = \frac{N}{\frac{1}{X_1} + \frac{1}{X_2} + ... + \frac{1}{X_N}} \tag{2.33}$$

односно

$$H = \frac{N}{\sum_{i=1}^{N} \frac{1}{X_i}}$$

Из (2.33) добија се реципрочна вредност хармонијске средине

$$\frac{1}{H} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{X_i} \tag{2.34}$$

што значи да је реципрочна вредност хармонијске средине једнака аритметичкој средини реципрочних вредности посматраног обележја X.

За обележје X дато расподелом фреквенција добиће се следећи израз за хармонијску средину

$$\frac{1}{H}=\frac{1}{N}\sum_{i=1}^{k}\frac{f_i}{x_i} \tag{2.35}$$

а преко релативних фреквенција хармонијска средина ће се добити из формуле

$$\frac{1}{H}=\frac{1}{N}\sum_{i=1}^{k}\frac{1}{x_i}\cdot p_i \tag{2.36}$$

На пример, да би се одредило потребно време за вожњу аутобуса од места A до места B извршено је 40 пробних вожњи и добијени су следећи подаци:

Утрошено време	4	5	6	10
Број вожњи	5	5	20	10

За даљу анализу потребно је одредити средњу вредност за ова мерења. Како је време обрнуто пропорционално жељеном ефекту вожње, то је хармонијска средина најпогоднија средња вредност. Тако ћемо добити

$$\frac{1}{H}=\frac{1}{40}\left(\frac{1}{4}\cdot 5+\frac{1}{5}\cdot 5+\frac{1}{6}\cdot 20+\frac{1}{10}\cdot 10\right)$$

односно H = 6,07 па у даљој анализи као показатељ „просечно" потребног времена треба узети ову вредност. Интересантно је поредити аритметичку, геометријску и хармонијску средину.

То поређење омогућава тзв.

Cauchy-јева теорема: *Хармонијска средина је мања од геометријске, а геометријска је мања од аритметичке средине, тј.*

$$H<G<\overline{x} \tag{2.29}$$

Модус и медијана

У многим испитивањима интересују нас вредности обележја X које се најчешће појављују у посматраном скупу. На пример, ако се посматра једна телефонска централа онда нас интересује који је најчешћи број позива у посматраном временском интервалу. Или нас може интересовати величина личног дохотка коју има највећи број запослених и сл. Параметар обележја X који изражава ову особину зове се *модус*.

Модус *је она вредност обележја X која има највећу фреквенцију у посматраном статистичком скупу или она вредност у чијој се околини најчешће појављују „измерене" вредности X на статистичком скупу***.**

Модус је добар показатељ средње вредности за оне статистичке скупове који су хомогени. Највише се користи у процесу доношења разних пословних одлука.

Кад је обележје X груписано и дато расподелом фреквенција, за модус се узима средина оног групног интервала који има највећу фреквенцију.

При испитивањима једног статистичког скупа можемо се интересовати за ону вредност обележја X која дели цео скуп на два једнака дела. Та вредност обележја X биће једна средња вредност и зове се *медијана*.

Медијана *је она вредност обележја X која дели статистичку скуп на два једнака дела.*

Медијану ћемо означавати са M_e и то је она вредност обележја X за коју пола елемената статистичког скупа има обележје X не веће од M_e, а друга половина има вредности које нису мање од M_e.

Нека су

$$X_1, X_2, ..., X_N \qquad (2.38)$$

измерене вредности обележја X на N елемената статистичког скупа. Да бисмо одредили медијану прво ћемо вредности (2.38) поређати у неопадајући низ

$$X'_1 \le X'_2 \le \ldots \le X'_N$$

Медијана је једнака

$$M_e = \begin{cases} X'_{\frac{N+1}{2}} & \text{ако је N непаран број.} \\ \frac{1}{2}\left(X'_{\frac{N}{2}} + X'_{\frac{N}{2}+1} \right) & \text{ако је N паран број.} \end{cases} \qquad (2.39)$$

За обележје X које је груписано и дато расподелом фреквенција медијану одређујемо тако што прво одредимо групни интервал у коме се налази медијана M_e, а затим ћемо по договору изабрати једну вредност из тог интервала за медијану.

Нека је обележје X дато расподелом.

X	$a_0 - a_1$	$a_1 - a_2$	…	$a_{k-1} - a_k$
(f)	f_1	f_2		f_k

За одређивање интервала медијане, одредићемо онај индекс s за који ће бити задовољене следеће две једначине

$$\sum_{i=1}^{s} f_i \le \frac{N}{2}$$

$$\sum_{i=1}^{s+1} f_i > \frac{N}{2}$$

За медијану ћемо узети вредност обележја X у интервалу $a_s - a_{s+1}$ тако што ћемо на леву границу тог интервала додати једну вредност која ће зависити од дужине тог интервала, фреквенције и

разлике у збиру фреквенција до тог интервала и половине елемената статистичког скупа.

Тако ћемо добити

$$M_e = a_s + \frac{a_{s+1} - a_s}{f_{s+1}} \left(\frac{N}{2} - \sum_{i=1}^{s} f_i \right) \tag{2.40}$$

На основу података за 429 предузећа у бившој СФРЈ у 1972. години, посматраћемо број запослених и одредити медијану. Подаци и резултати су дати у Табели 2.8.

Број запослених (X)	Број предузећа (*f*)	Кумулативне фреквенције (c)
до 15	22	22
15 - 30	10	32
30 - 60	27	59
60 - 125	68	127
125 - 250	77	204
250 - 500	100	304
500 - 1000	65	369
1000 - 2000	38	407
преко 2000	22	429
Укупно	**429**	

Табела 2.8. Број запослених у саобраћајним предузећима

Из колоне за кумулативне фреквенције види се да је

$$\sum_{i=1}^{5} f_i = 204 < 214{,}5 \qquad \sum_{i=1}^{6} f_i = 304 > 214{,}5$$

тако да је s = 5, а медијана је

$$M_e = 250 + \frac{500 - 250}{100} \left(\frac{429}{2} - 204 \right)$$

односно $M_e = 276{,}25$.

2.5. МЕРЕ ВАРИЈАБИЛИТЕТА

Другу групу параметара којима се описују резултати посматраног обележја X на датом статистичком скупу чине параметри који изражавају варијабилитет посматраног обележја. То су они показатељи који зависе од укупних разлика у вредностима обележја X на елементима статистичког скупа. Та група параметара су тзв. мере варијабилитета или мере дисперзије.

У мере варијабилитета спадају следећи параметри:

- Размак варијације,
- Квартилна девијација,
- Средња девијација,
- Варијанса и стандардна девијација, и
- Коефицијент варијације.

Вредности параметара који спадају у мере варијабилитета треба да укажу на две чињенице. Прво, њихов износ говори колико су средње вредности обележја X добри представници свих његових вредности на посматраном скупу, друго, ови параметри треба да покажу колико се сви елементи на посматраном скупу међусобно разликују у односу на дато обележје X.

Ако је варијабилитет посматраног обележја мали, онда то значи да средња вредност може бити добар представник посматраног обележја, а ако је његов варијабилитет велики то значи да ће средња вредност бити лош представник тог обележја.

На пример, ако се утврди да је варијабилитет у личним примањима запослених велики, онда то значи да ће просечна лична примања бити лош показатељ примања појединих радника.

Исто тако, ако се утврди за домаћинства у Београду да је варијабилитет њихових издатака за основне животне потребе (стан, храна и сл.) јако мали, то ће значити да ће просечни издаци

домаћинства бити доста добар показатељ издатака свих домаћинстава у Београду.

Са друге стране, ако се зна колики је варијабилитет неке појаве може се утицати на ту појаву у циљу њене стабилизације.

На пример, промене, односно варијабилитет у температури, пулсу, крвном притиску и сл. служе као основни показатељи у постављању дијагнозе и одређивању терапије код разних обољења.

Размак варијације

Размак варијације је најједноставнија мера варијабилитета самим тим и најлакше га је одредити. То је размак између највеће и најмање вредности обележја X на посматраном статистичком скупу.

Означимо са X_{MAX} највећу вредност обележја X, а са X_{MIN} његову најмању вредност на статистичком скупу. Тада је ***размак варијације*** *једнак*

$$R = X_{MAX} - X_{MIN} \tag{2.41}$$

где смо са R означили размак варијације.

Нека је обележје X груписано и дато расподелом фреквенција,

(X)	$a_0 - a_1$	$a_1 - a_2$	…	$a_{k-1} - a_k$
(f)	f_1	f_2		f_k

(2.42)

при чему је

$$\sum_{i=1}^{k} f_i = N$$

Да бисмо тачно одредили размак варијације морамо поред расподеле (2.42) знати најмању вредност обележја X у првом интервалу a_0 - a_1 и највећу вредност обележја X у последњем интервалу $a_{K-1} - a_K$. Ако нам ове вредности нису познате не можемо ни одредити

размак варијације. Размак варијације је погодна мера варијабилитета за она обележја код којих се највећа и најмања вредност не разликују много од осталих вредности обележја X. Исто тако, то је важан показатељ за оне појаве које су описане великим и малим вредностима обележја.

На пример у временским извештајима важна је разлика између највише и најниже температуре, или између највишег и најнижег водостаја.

Највећу примену размак варијације има у контроли квалитета у индустријској производњи.

Овај показатељ је лоша мера варијабилитета за оне статистичке скупове код којих постоји мали број елемената са малим или са великим вредностима обележја X, док се вредности X код осталих елемената не разликују много. Тако би на пример у посматрањима цена свих приватних кућа у граду, размак варијација био лош показатељ варијабилитета јер су поједине луксузне куће веома скупе у односу на најјефтиније.

Квартилна девијација

Да би елиминисали утицај екстремних вредности у посматраном статистичком скупу можемо податке редуковати, тако што ћемо из посматрања искључити оне елементе статистичког скупа код којих обележје има те вредности. Од укупног броја елемената статистичког скупа треба искључити један одређен проценат. Најчешће то чинимо преко квартила.

Означимо са $X_{0,25}$ ону вредност обележја X за коју 25% елемената посматраног скупа има X мање од $X_{0,25}$. Са $X_{0,75}$ означимо ону вредност обележја X за коју 75% елемената статистичког скупа има вредност X мању од $X_{0.75}$, или ону вредност за коју 25% елемената има вредност X већу од $X_{0.75}$.

Вредност $X_{0.25}$ назива се **доњи квартил** *обележја X, а $X_{0.75}$* **горњи квартил** *обележја X.*

Доњи и горњи квартил деле цео статистички скуп на три дела. Једна четвртина елемената су они елементи код којих је $X < X_{0.25}$, друга четвртина су елементи код којих је $X > X_{0.75}$ а преосталих 50% елемената има вредност обележја X која је $X_{0.25} < X < X_{0.75}$

Квартилна девијација је број одређен изразом

$$Q = \frac{X_{0,75} - X_{0,25}}{2} \tag{2.43}$$

где је Q ознака за квартилну девијацију.

Доњи и горњи квартили се рачунају слично као и медијана. Ако су вредности обележја X код N елемената статистичког скупа X_1, X_2, ..., X_N поређане по величини, треба одредити оне вредности у низу X_1, X_2, ..., X_N за које је четвртина низа мања од $X_{0.25}$, а четвртина већа од $X_{0.75}$.

Аналогно одређивању медијане за груписане податке одредићемо и квартиле. Нека је обележје X дато расподелом,

(X)	$a_0 - a_1$	$a_1 - a_2$	…	$a_{k-1} - a_k$
(f)	f_1	f_2		f_k

при чему је

$$\sum_{i=1}^{k} f_i = N$$

За одређивање доњег квартила одредићемо прво групни интервал у коме ће се налазити $X_{0.25}$. То ће бити интервал са редним бројем с за који су испуњене следеће неједначине

$$\sum_{i=1}^{s} f_i < \frac{N}{4} \qquad \sum_{i=1}^{s+1} f_i \geq \frac{N}{4} \tag{2.44}$$

Тада је доњи квартил

$$X_{0,25} = a_s + \frac{a_{s+1} - a_s}{f_{s+1}} \left(\frac{N}{4} - \sum_{i=1}^{s} f_i\right) \tag{2.45}$$

На сличан начин одређујемо и горњи квартил. Прво одредимо интервал са редним бројем р за који су испуњене неједначине

$$\sum_{i=1}^{p} f_i < \frac{3N}{4} \qquad \sum_{i=1}^{p+1} f_i \geq \frac{3N}{4}.$$

Горњи квартил је

$$X_{0,75} = a_p + \frac{a_{p+1} - a_p}{f_{p+1}} \left(\frac{3N}{4} - \sum_{i=1}^{p} f_i\right) \tag{2.46}$$

Квартиле ћемо најлакше израчунати ако се користимо табелом података и кумулативним фреквенцијама.

Посматрајмо пример расподеле броја запослених у угоститељским предузећима. Расподела је дата у Табели 2.9.

Број запослених (X)	Број предузећа (f)	Кумулативне фреквенције (c)
- 6	73	73
6- 15	51	124
15- 30	45	169
30- 60	63	232
60- 125	72	304
125- 250	54	358
250- 500	30	388
500-1000	14	402
1000-2000	2	404
Укупно	404	

Табела 2.9. Расподела броја запослених у угоститељству

Овде је N=404, па је N/4=101 тако да се из треће колоне Табеле 2.9. види да је доњи квартил у интервалу 6-15. На основу обрасца (2.45) добиће се

$$X_{0,25} = 6 + \frac{15-6}{51}(101-73) = 11$$

Горњи квартил се налази у интервалу 60-125, и на основу формуле (2.46) има вредност

$$X_{0,75} = 60 + \frac{125-60}{72}(303-232) = 62$$

Квартилна девијација за број запослених на основу (2.43) има вредност

$$Q=25,5$$

Средња девијација

Вредности размака варијације и квартилне девијације не зависе од свих вредности обележја X на статистичком скупу. Ипак, за бољу анализу података потребно је одредити параметре варијабилитета који би зависили од свих вредности обележја X. Такав параметар је ***средња девијација*** или ***средње апсолутно одступање***. Дефинише се на одступањима појединих вредности X од средње вредности. У пракси се најчешће посматрају одступања од аритметичке средине обележја X. Касније ћемо видети да би због извесних оптималних услова требало посматрати одступања од медијане. Разлике које се добију кад се посматрају одступања у односу на медијану и у односу на аритметичку средину нису велике, па се зато средња девијација рачуна на основу одступања од аритметичке средине.

Средња девијација *је аритметичка средина апсолутних вредности одступања X од његове аритметичке средине* $\bar{x}$ *на посматраном статистичком скупу.*

Нека су $X_1, X_2, \ldots, X_N$, *вредности обележја X на N елемената статистичког скупа, а* $\bar{x}$ *аритметичка средина обележја X на овом скупу. Тада је низ*

$$\left|X_1 - \bar{x}\right|, \left|X_2 - \bar{x}\right| \ldots, \left|X_n - \bar{x}\right| \qquad (2.47)$$

низ апсолутних одступања X од његове аритметичке средине. Аритметичка средина низа (2.47) је **средња девијација** *која је дата изразом*

$$e_m = \frac{1}{N}\sum_{i=1}^{N} | X_i - \bar{x} | \qquad (2.48)$$

За груписане податке дате расподелом фреквенција,

(X)	$a_0 - a_1$	$a_1 - a_2$	…	$a_{k-1} - a_k$
	x_1	x_2	…	x_k
(f)	f_1	f_2		f_k

при чему су x_1, x_2,...,x_k средине групних интервала, а f_1, f_2, ...,f_k одговарајуће фреквенције, образац (2.48) за средњу девијацију се своди на

$$e_m = \frac{1}{N}\sum_{i=1}^{k} | x_i - \bar{x} | f_i \qquad (2.49)$$

или на израз

$$e_m = \sum_{i=1}^{k} | x_i - \bar{x} | p_i \qquad (2.50)$$

при чему су p_1, p_2,..., p_k релативне фреквенције за посматрано обележје X.

На сличан начин смо могли дефинисати меру варијабилитета на основу низа апсолутних одступања у односу на било који број a, тј. на основу низа

$$|X_1 - a|,\ |X_2 - a|,\ \ldots,\ |X_N - a| \qquad (2.51)$$

Особина низа апсолутних одступања: *збир апсолутних одступања низа (2.51) има минималну вредност ако за a узмемо медијану, тј. ако је $a = M_e$.*

Доказ. Да би проверили ову особину претпоставићемо да су вредности обележја X поређане у неопадајући низ

$$X_1 \le X_2 \le \cdots \le X_N \tag{2.52}$$

Прво ћемо посматрати случај кад је $a \ge M_e$. Тада у низу (2.52) постоји највећи број X који није мањи од медијане. Нека је то X_s. Исто тако постоји највећи број X који није већи од a. Нека је то X_r.

Збир чланова низа може се раставити на три члана

$$\sum_{i=1}^{N} |X_i - a| = \sum_{i=1}^{s} |X_i - a| + \sum_{i=s+1}^{r} |X_i - a| + \sum_{i=r+1}^{N} |X_i - a| \tag{2.53}$$

Зна се да је апсолутна вредност неког броја z једнака

$$|z| = \begin{cases} z, & \text{ако је z} > 0. \\ -\text{z}, & \text{ако је z} < 0. \end{cases} \tag{2.54}$$

Користећи (2.54), збир (2.53) се може написати у облику

$$\sum_{i=1}^{N} |X_i - a| = \sum_{i=1}^{s} (X_i - a) - \sum_{i=s+1}^{r} (X_i - a) + \sum_{i=r+1}^{N} (X_i - a) \tag{2.55}$$

После додавања и одузимања медијане M_e у загради, први члан на десној страни израза (2.55) се може написати у облику

$$-\sum_{i=1}^{s} (X_i - a) = -\sum_{i=1}^{s} (X_i - M_e) - s(M_e - a)$$

односно

$$-\sum_{i=1}^{s} (X_i - a) = \sum_{i=1}^{s} |X_i - M_e| + s(a - M_e) \tag{2.56}$$

јер су разлике $(X_i - M_e)$ негативне.

На сличан начин се може изразити и други члан у изразу (2.55)

$$-\sum_{i=s+1}^{r}(X_i - a) = -\sum_{i=s+1}^{r}|X_i - M_e| - (r-s)(M_e - a) \tag{2.57}$$

Трећи члан израза (2.55) је једнак

$$\sum_{i=r+1}^{N}(X_i - a) = \sum_{i=r+1}^{N}|X_i - M_e| + (N-r)(M_e - a) \tag{2.58}$$

Из (2.56), (2.57) и (2.58) се види да ће збир апсолутних одступања зависити од a преко следећег израза

$$s(a - M_e) - (r-s)(M_e - a) + (N-r)(M_e - a)$$

тј. од вредности

$$(2r - N)(a - M_e) \tag{2.59}$$

Пошто је (2.59) позитиван број то ће он достићи минималну вредност кад је $a = M_e$, а за ту вредност ће и збир апсолутних одступања (2.53) имати минимум.

На сличан начин би могли проверити ову особину и за случај кад би претпоставили да је $a \le M_e$.

Варијанса и стандардна девијација

Видели смо да варијабилитет обележја X можемо изразити преко одступања појединих његових вредности од средње вредности. Та одступања могу бити позитивна или негативна, а за укупан варијабилитет није нам важан знак одступања. Зато можемо посматрати квадрате тих одступања и на основу њих дефинисати параметар варијабилитета. Тако се дефинишу *варијанса и стандардна девијација*.

Варијанса *је аритметичка средина квадрата одступања врцдности обележја X од њихове аритметичке средине. Позитивна вредност корена варијансе назива се* **стандардна девијација**.

Нека су измерене вредности обележја X на N елемената статистичког скупа

$$X_1, X_2, ...,X_N.$$

Тада је низ квадрата одступања вредности обележја X од његове аритметичке средине

$$(X_1-\bar{x})^2,(X_2-\bar{x})^2,...,(X_N-\bar{x})^2 \tag{2.60}$$

Аритметичка средина низа (2.60) је

$$S^2=\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{x})^2 \tag{2.61}$$

и назива се **варијанса** *обележја X, а*

$$S=+\sqrt{S^2} \tag{2.62}$$

је његова **стандардна девијација**.

Нека је обележје X груписано и дато расподелом фреквенција,

(X)	a_0-a_1 a_1-a_2 … $a_{k-1}-a_k$ x_1 x_2 … x_k
(f)	f_1 f_2 f_k

тада је варијанса обележја X дата изразом

$$S^2 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot f_i \tag{2.63}$$

Варијанса обележја X се може изразити и преко релативних фреквенција

$$S^2 = \sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot p_i \tag{2.64}$$

Ради тачности рачунања и прегледности варијансу и стандардну девијацију треба рачунати преко табеле (Табела 2.10).

X	f	$X \cdot f$	$(X - \bar{x})$	$(X - \bar{x})^2 \cdot f$
x_1	f_1	$x_1 \cdot f_1$	$x_1 - \bar{x}$	$(x_1 - \bar{x})^2 \cdot f_1$
x_2	f_2	$x_2 \cdot f_2$	$x_2 - \bar{x}$	$(x_2 - \bar{x})^2 \cdot f_2$
…	…	…	…	…
x_k	f_k	$x_k \cdot f_k$	$x_k - \bar{x}$	$(x_k - \bar{x})^2 \cdot f_k$
Укупно		$\sum_{i=1}^{k} x_i f_i$		$\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot f_i$

Табела 2.10. Рачунање варијансе

Из података се прво попуне прва и друга колона табеле 2.10. Трећа колона се добија множењем прве две колоне и из њеног збира се одређује $\bar{x}$. Затим се од прве колоне одузима $\bar{x}$ и добија се четврта колона. Квадрирањем четврте колоне и множењем са другом добија се пета колона чији збир треба поделити са N и на тај начин смо одредили варијансу S^2.

Видели смо да је аритметичка средина имала следећу особину: збир квадрата одступања вредности обележја X од било ког броја најмањи је за $a = \bar{x}$. Зато се аритметичка средина $\bar{x}$ користи за одређивање варијансе и стандардне девијације обележја X.

Навешћемо неке особине варијансе, које ће омогућити њену употребу у даљој статистичкој анализи и које ће олакшати рачунања.

Особина 1. *Варијанса обележја X је једнака нули, ако и само ако је обележје X константно на свим елементима статистичког скупа, тј.*

$$\left\{S^2 = 0\right\} \Leftrightarrow \left\{X_i = c, \forall i \in [1, 2, N]\right\}$$

Доказ. Пошто је варијанса збир позитивних бројева, онда ће она имати вредност нулу ако и само ако је сваки тај број једнак нули. То ће бити испуњено једино кад је $X = c$.

Особина 2. *Варијанса обележја X се може изразити формулом*

$$S^2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \bar{x}^2 \tag{2.65}$$

а ако је обележје X груписано, варијанса је

$$S^2 = \frac{1}{N}\sum_{i=1}^{k} x_i^2 f_i - \bar{x}^2 \tag{2.66}$$

Доказ. Пошто је варијанса дата изразом

$$S^2 = \frac{1}{N}\sum_{i=1}^{N} \left(X_i - \bar{x}\right)^2$$

после квадрирања и растављања десне стране на три сабирка добија se

$$S^2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \frac{2}{N}\sum_{i=1}^{N} X_i \cdot \bar{x} + \frac{1}{N}\sum_{i=1}^{N} \bar{x}^2 \tag{2.67}$$

У другом члану израза (2.67) $\bar{x}$ не зависи од индекса и па се може извући пред знак сигма, а у трећем члану збир је једнак $N \cdot \bar{x}$. Пошто је

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

то се после замене (2.67) своди на

$$S^2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 - 2 \cdot \bar{x} \cdot \bar{x} + \frac{1}{N} N\bar{x}^2$$

одакле се после скраћивања и сабирања добија образац (2.65).

Особина 3. *Нека обележје X има варијансу* S_x^2*, а обележје Y има варијансу* S_y^2 *и нека између X и Y постоји линеама веза облика*

$$Y = aX + b. \tag{2.68}$$

Тада је варијанса обележја Y једнака

$$S_y^2 = a^2 S_x^2 \tag{2.69}$$

Доказ. За сваку вредност обележја X = X_i, из релације (2.68) добиће се одговарајућа вредност обележја Y

$$Y_i = aX_i + b \tag{2.70}$$

Аритметичка средина обележја Y је

$$\bar{y} = a\bar{x} + b \tag{2.71}$$

Замењујући (2.70) и (2.71) у израз за варијансу обележја Y добиће се

$$S_y^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{y})^2 = \frac{1}{N}\sum_{i=1}^{N}(aX_i - a\bar{x})^2$$

тј.

$$S_y^2 = a^2 \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{x})^2$$

односно израз (2.69), што доказује Особина 3.

Особина 4. *Претпоставимо да један статистички подскуп има* N_1 *елемената са аритметичком средином* $\bar{x}_1$ *и варијансом обележја X једнаком* S_1^2*. Нека други статистички подскуп има* N_2 *елемената, аритметичку средину* $\bar{x}_2$ *и варијансу* S_2^2*. Тада је варијанса статистичког скупа једнака*

$$S^2 = \frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2} + \frac{N_1 N_2}{(N_1 + N_2)^2}(\bar{x}_1 - \bar{x}_2)^2 \qquad (2.72)$$

Доказ. Означимо са

$$X_1, X_2, ..., X_{N_1}$$

вредност обележја X на првом подскупу, а са

$$X'_1, X'_2, ..., X'_{N_2}$$

његове вредности на другом подскупу.

Аритметичка средина целог скупа једнака је пондерисаној аритметичкој средини

$$\bar{x} = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2}{N_1 + N_2} \qquad (2.73)$$

а његова варијанса је

$$S^2 = \frac{1}{N_1 + N_2}\left(\sum_{i=1}^{N_1} (X_i - \bar{x})^2 + \sum_{i=1}^{N_2} (X'_i - \bar{x})^2 \right) \qquad (2.74)$$

Додавањем и одузимањем $\bar{x}_1$ у загради код првог члана на десној страни израза (2.74), а у другом члану додавањем и одузимањем $\bar{x}_2$ добиће се

$$S^2 = \frac{1}{N_1 + N_2}\left(\sum_{i=1}^{N_1} (X_i - \bar{x}_1)^2 + \sum_{i=1}^{N_2} (X'_i - \bar{x}_2)^2 + N_1 (\bar{x}_1 - \bar{x})^2 + N_2 (\bar{x}_2 - \bar{x})^2 \right) \qquad (2.75)$$

Из (2.73) се добија

$$\bar{x}_1 - \bar{x} = \frac{N_2 (\bar{x}_1 - \bar{x}_2)}{N_1 + N_2} \qquad \bar{x}_2 - \bar{x} = \frac{N_1 (\bar{x}_2 - \bar{x}_1)}{N_1 + N_2}$$

па се после замене у (2.75) добија

$$S^2 = \frac{1}{N_1 + N_2}\left(N_1 S_1^2 + N_2 S_2^2 + \frac{N_1 N_2}{N_1 + N_2}\left(\bar{x}_1 - \bar{x}_2\right)^2 \right)$$

јер је

$$\sum_{i=1}^{N_1}\left(X_i - \bar{x}_1\right)^2 = N_1 S_1^2 \qquad \sum_{i=1}^{N_2}\left(X'_i - \bar{x}_2\right)^2 = N_2 S_2^2$$

Тиме је и доказана особина 4.

Ову особину можемо уопштити и на случај кад се статистички скуп састоји од више подскупова.

Особина 5. *Нека се статистички скуп састоји од k подскупова величина N_1, N_2, ... , N_k чије аритметичке средине су $\bar{x}_1, \bar{x}_2, ..., \bar{x}_K$, а варијансе*

$$S_1^2, S_2^2, ..., S_K^2$$

Тада је варијанса статистичког скупа једнака

$$S^2 = \frac{1}{N}\sum_{i=1}^{K} N_i S_i^2 + \frac{1}{N}\sum_{i=1}^{K} N_i \left(\bar{x}_i - \bar{x}\right)^2$$

при чему је $N = \sum_{i=1}^{k} N_i$ величина статистичког скупа, а $\bar{x}$ је његова аритметичка средина.

Доказ. Применом принципа математичке индукције лако се из Особине 4 добија Особина 5.

Коефицијент варијације

Често смо у ситуацији да треба да поредимо варијабилитет два различита обележја. На пример варијабилитет у висини и тежини ученика. Пошто су сви до сада наведени параметри варијабилитета изражени у јединицама у којима је мерено посматрано обележје, онда

немогуће је вредности тих параметара поредити за обележја која су мерена у различитим јединицама.

За оваква поређења погоднији су релативни показатељи, какав је *коефицијент варијације.*

Коефицијент варијације *је процентуално изражен однос стандардне девијације и аритметичке средине посматраног обележја X.*

Нека је $\bar{x}$ *аритметичка средина обележја X, а S његова standardna devijacija. Тада је* **коефицијент варијације V** *једнак*

$$V = 100 \cdot \frac{S}{\bar{x}} \tag{2.76}$$

На пример, нека је на једном скупу студената мерена њихова висина и тежина. Нека X представља висину, а Y тежину и нека је аритметичка средина $\bar{x}$, а S_x, њена стандардна девијација. Даље, нека је аритметичка средина $\bar{y}$, а њена стандардна девијација S_y. Претпоставимо да су се из резултата мерења добиле следеће вредности

$\bar{x}$ = 174 cm; S_x = 11 cm, $\bar{y}$ = 69 kg; S_y = 7 kg.

Интересује нас да ли је код студената већи варијабилитет у тежини од варијабилитета у висини, или је обрнуто?

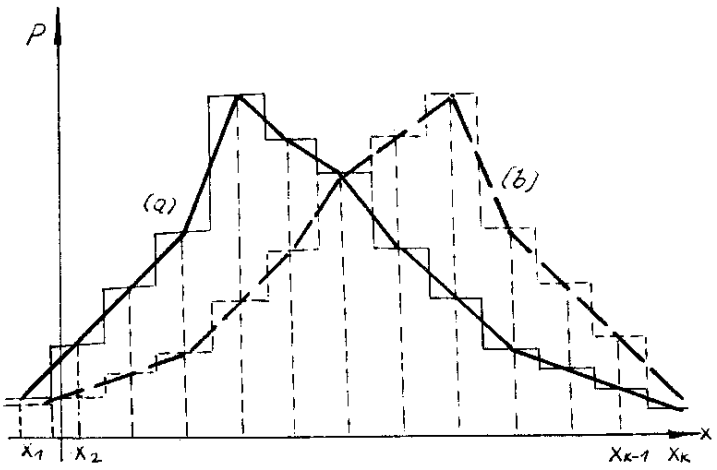
Пошто су S_x и S_y изражени у различитим јединицама, немогуће је њихово поређење. Зато треба одредити коефицијенте варијације.

За X се добија $V_1 = 100 \cdot \frac{11}{174} = 6{,}32\,\%$, а за Y $V_2 = 100 \cdot \frac{7}{169} = 10{,}14\,\%$, па се може закључити да је за посматрани скуп студената већи варијабилитет у тежини од варијабилитета у висини.

2.6. КОЕФИЦИЈЕНТИ АСИМЕТРИЈЕ И СПЉОШТЕНОСТИ

Битна карактеристика расподеле обележја X на статистичком скупу је симетричност. Може се десити да су поједине вредности обележја X симетрично распоређене у односу на неку одређену вредност. Мање вредности обележја X, међутим, могу имати мање фреквенције у односу на веће вредности или обрнуто. На Слици 2.7. су приказане две расподеле. Код расподеле (a) мање вредности обележја X имају веће фреквенције, а код расподеле (b) је обрнуто.

За обе расподеле кажемо да су асиметричне.



Слика 2.7. Асиметричне расподеле обележја X

За мерење асиметричности расподеле обележја X служи једноставан параметар који је базиран на разлици између аритметичке средине $\bar{x}$ и медијане M_e. Наиме, ако је расподела обележја X симетрична, онда су аритметичка средина и медијана једнаке. А кад постоји асиметричност онда ће постојати и разлика између $\bar{x}$ и M_e. Та разлика је већа што је асиметричност већа јер медијана зависи само од

фреквенција и помераће се ка вредностима за X које имају веће фреквенције. Аритметичка средина зависи и од фреквенција и од вредности обележја X па ће њено померање ка вредностима обележја X које имају веће фреквенције, бити спорије у односу на померање медијане.

Коефицијент асиметрије *је количник*

$$K_A = \frac{3(\bar{x} - M_e)}{S} \tag{2.77}$$

при чему је $\bar{x}$ *аритметичка средина,* M_e *медијана, а S стандардна девијација обележја X.*

Коефицијент асиметрије је једнак нули за симетричне расподеле. Ако је расподела *асиметрична у леву страну* (расподела a на Слици 2.7) коефицијент асиметрије K_A *је позитиван*, а ако је расподела *асиметрична у десну страну* (расподела b на Слици 2.7) коефицијент асиметрије K_A *је негативан*. Напоменимо да је коефицијент асиметрије K_A релативни број без димензије који нам омогућава поређења расподела обележја X и Y мерених у различитим јединицама мере.

Мада је коефицијент асиметричности веома једноставан за рачунање ипак се у пракси чешће користи тзв. Пирсон-ов (Pearson) први коефицијент као мера асиметрије. Дефинисан је на одступањима појединих вредности обележја X од њихове аритметичке средине.

Одредимо прво тзв. трећи централни моменат. То је аритметичка средина низа

$$(X_1 - \bar{x})^3, (X_2 - \bar{x})^3, ..., (X_N - \bar{x})^3$$

који представља трећи степен одступања појединих вредности обележја X од њihove aritmetičke sredine.

Трећи централни моменат *је једнак*

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{x})^3 \tag{2.78}$$

Ако је обележје X груписано и дато расподелом фреквенција

(X)	x_1	x_2	. . .	X_k
(f)	f_1	f_2	. . .	f_k

тада је трећи централни моменат једнак

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^3 \cdot f_i \tag{2.79}$$

или преко релативних фреквенција

$$\mu_3 = \sum_{i=1}^{k}(x_i - \bar{x})^3 \cdot p_i \tag{2.80}$$

Први Pearson-ов коефицијент *је дефинисан преко трећег централног момента и стандардне девијације обележја X као количник*

$$\beta_1 = \frac{\mu_3}{S^3} \tag{2.81}$$

Ако је расподела обележја X симетрична, коефицијент β_1 је једнак нули. Усвојићемо и обрнуто: ако је $\beta_1 = 0$ рећи ћемо да је расподела обележја X симетрична. Ако је расподела асиметрична у леву страну коефицијент β_1 је позитиван, а ако је расподела асиметрична у десну страну коефицијент β_1 је негативан. што је асиметричност већа то је апсолутна вредност коефицијента β_1 већа.

Пример: Нека је обележје X дато расподелом

(X)	-6	-5	-4	-3	-2	-1	0	1,5	3,5	5,5	7,5	9,5
(f)	7	15	34	19	19	19	6	9	10	6	1	3

Израчунати коефицијент асиметрије!

Табела 2.10. Рачунање коефицијента асиметрије

(X)	(f)	(C)	(Xf)	$(X-\bar{x})$	$(X-\bar{x})^3$	$\left((X-\bar{x})^3 \cdot f\right)$	$\left(X^2 f\right)$
-6	7	7	-4.2	-4.12	-68.92	-482.45	252
-5	25	32	-125	-3.1	-29.79	-744.78	625
-4	34	66	-136	-2.1	-9.26	-314.87	544
-3	19	85	-57	-1.1	-1.33	-25.29	171
-2	19	104	-38	-0.1	-0.00	-0.01	76
-1	19	123	-19	0.9	0.73	13.85	19
0	6	129	0	1.9	6.86	41.16	0
1.5	9	138	13.5	3.4	39.30	353.74	20.25
3.5	10	148	35.0	5.4	157.46	1574.64	122.50
5.5	6	154	33.0	7.4	405.22	2431.34	181.50
7.5	1	155	7.5	9.4	830.58	830.58	56.25
9.5	3	158	28.5	11.4	1481.54	4444.63	270.75
Укупно	158		-299.5			8122.54	2338.25

Из треће колоне Табеле 2.10. се види да медијана обележја X припада групном интервалу (-3,5:-2,5) тако да је медијана једнака

$$M_e = -3{,}5 + \frac{1}{19}(79-66) = -2{,}816$$

Из четврте колоне се добија да је аритметичка средина обележја X једнака

$$\bar{x} = \frac{1}{158}(-229{,}5) = -1{,}896$$

Варијансу и стандардну девијацију добијаћемо из последње колоне

$$S^2 = \frac{1}{158} 2338{,}25 - (-1{,}896)^2 = 11{,}204$$

тако да је стандардна девијација $S = 3{,}347$.

Коефицијент симетрије K_A има вредност

$$K_A = \frac{3(-1.896 + 2.816)}{3.347} = 0.824$$

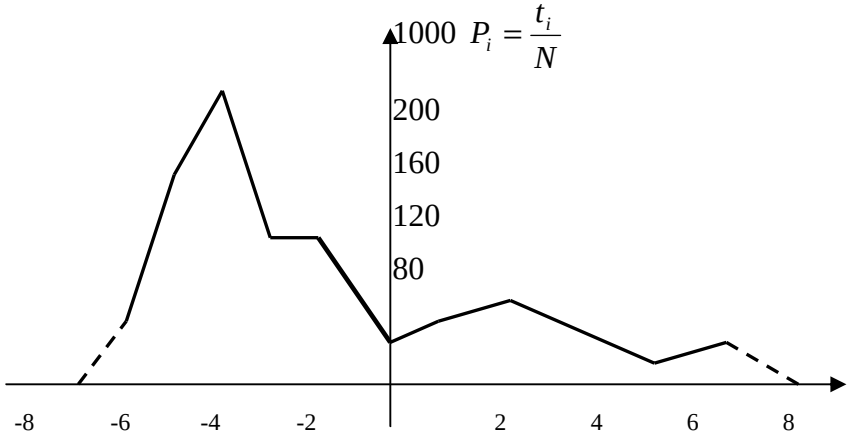
Из Табеле 2.10. добија се трећи централни моменат

$$\mu_3 = \frac{1}{158} 8122.54 = 51.41$$

па је први Pearson-ов коефицијент једнак

$$\beta_1 = \frac{51.411}{3.347^3} = 1.37$$

Као што се види оба коефицијента су позитивна, што значи да је расподела асиметрична у лево. То се види и на слици 2.8. на којој је приказана расподела посматраног обележја X.

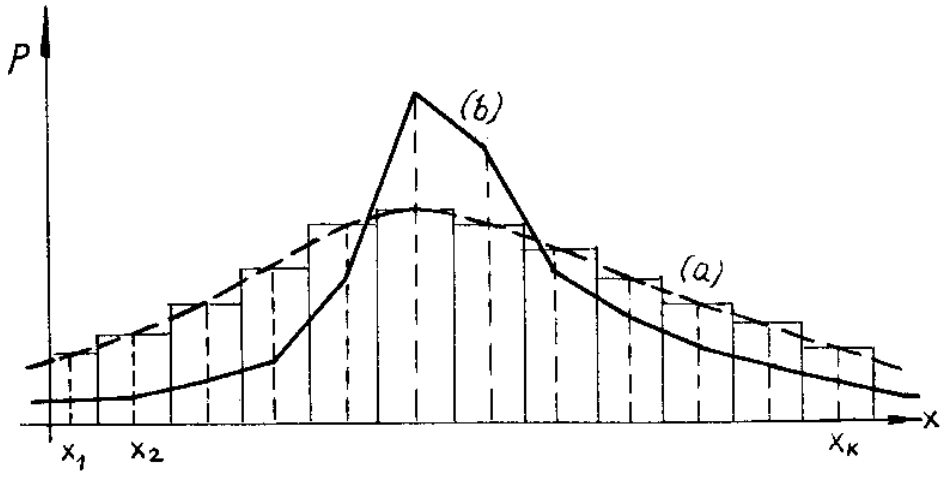


Слика 2.8. Расподела релативних фреквенција за X

Мада два обележја X и Y могу имати исту асиметричност или чак бити потпуно симетрична, њихове расподеле могу бити сасвим

различите. На Слици 2.9 приказане су две такве расподеле. Оне се разликују у спљоштености.

Коефицијент спљоштености *служи као мера спљоштености једне расподеле. Дефинише се на основу тзв. четвртог централног момента.*



Слика 2.9. Расподеле са различитим спљоштеностима

Посматрајмо одступања појединих вредности обележја X од њихове аритметичке средине и низ

$$(X_1 - \bar{x})^4, (X_2 - \bar{x})^4, ..., (X_N - \bar{x})^4 \tag{2.82}$$

Четврти централни моменат је аритметичка средина низа (2.82), тј.

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{x})^4 \tag{2.83}$$

За груписано обележје дато расподелом фреквенција четврти централни моменат се своди на

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^4 \cdot f_i \tag{2.84}$$

односно

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{k}\left(X_i - \bar{x}\right)^4 \cdot p_i \tag{2.85}$$

за расподелу релативних фреквенција.

Други Pearson-ов коефицијент *или* **коефицијент спљоштености** *је количник*

$$\beta_2 = \frac{\mu_4}{S^4} \tag{2.86}$$

при чему је S стандардна девијација обележја X.

Ако је коефицијент $\beta_2 = 3$, кажемо да *обележје X има нормалну спљоштеност*. Кад је $\beta_2 > 3$ кажемо да *X има већу спљоштеност од нормалне*, а кад је $\beta_2 < 3$ кажемо да је *спљоштеност обележја X мања од нормалне*.

Уместо коефицијента спљоштености β_2 може се користити коефицијент

$$\gamma_2 = \beta_2 - 3$$

који има вредност нулу кад обележје X има нормалну спљоштеност. У осталим случајевима коефицијент γ_2 је или позитиван или негативан.

3. СЛУЧАЈНИ ДОГАЂАЈИ И ВЕРОВАТНОЋА

Циљ дескриптивне статистике је био да се прикупе подаци о тзв. масовним појавама и да се изврши анализа посматраног статистичког скупа на основу прикупљених података. Таква анализа је у великој мери непотпуна јер не објашњава посматрану појаву или процес са становишта узрочно-последичних односа. Са друге стране, методе дескриптивне статистике не омогућавају утврђивање законитости које важе у масовним појавама, а то би требало да буде основни циљ сваког испитивања.

Поред тога многе податке је немогуће у потпуности обухватити због њихове масовности. Зато је потребно изучавати појаве на великим статистичким скуповима и то на основу мањег броја података, а затим извршити уопштавање добијених резултата и на цео статистички скуп.

Теорија вероватноће представља основну научну дисциплину која омогућава решавање ова два проблема статистичке анализе. Зато су следећа поглавља посвећена основама теорије вероватноће и то у оној мери која је потребна за даља статистичка истраживања.

Развој статистичке методологије почиње онда када су статистичари почели користити дефиниције, постулате и теореме математичке ***теорије вероватноће***. Ова теорија представља "костур" статистике, а почела се формирати примењујући математику у одређивању могућности у тзв. играма на срећу: коцка, карте, рулет и сл. У XVII веку у приватној кореспонденцији математичара Pascal-a и Fermat-a са Chevalier de Méré, која се односила на различита питања везана за посматрање игара, формиран је основ за развој теорије вероватноће.

Статистичари прихватају скоро фасцинирајуће могућности теорије вероватноће у решавању статистичких проблема. Развијају специјалну методологију која постаје најважније средство у техничким, биомедицинским и друштвеним наукама.

Математичка дисциплина која се бави израчунавањем законитости случајних појава назива се теорија вероватноће. Она је почела да се развија средином XVII века и то на основу извесних проблема везаних за хазардне игре. Захваљујући узајамном стимулативном деловању теорије и примене дошло је до развитка опште математичке теорије која има велику примену у веома широком спектру различитих области природних и друштвених појава.

3.1. СИГУРНИ, НЕМОГУЋИ И СЛУЧАЈНИ ДОГАЂАЈИ

Изучавање појава или процеса, било у природи или у друштву, врши се тако што се посматрају догађаји везани за ту појаву. Утврђивањем одређених релација међу разним догађајима описују се појаве и одређују законитости везане за ту појаву или процес.

Све догађаје везане за једну појаву или процес делимо у три класе:

- **Немогући догађаји,**
- **Сигурни догађаји, и**
- **Случајни догађаји**.

У природи постоје појаве које се могу потпуно описати помоћу сигурних и немогућих догађаја. То су појаве детерминистичког карактера. На пример, познавајући законе кретања небеских тела у сунчевом систему могуће је сасвим тачно утврдити положај сваке планете у сваком тренутку, предвидети помрачење сунца, утврдити удаљеност и брзину кретања планета и сл. Исто тако, знајући Newton-ов закон гравитације могуће је утврдити брзину падања тела, на основу времена и брзине утврдити положај тела и сл.

Проучавање оваквих појава се заснива на утврђивању законитости из којих следе сигурни или немогући догађаји. Међутим, у природи, а нарочито у друштву далеко је већи број појава које се не могу потпуно описати помоћу сигрних и немогућих догађаја.

Узмимо најједноставнији пример бацања коцке. Јасно је да се може утврдити да ће коцка сигурно пасти на сто, да је немогуће да се задржи у ваздуху и сл. Таква тврђења нам ни издалека не описују тај експеримент. Знамо да ће при бацању пасти један од бројева

$$\{1,2,3,4,5,6\}$$

али не знамо који. Ако при бацању коцке посматрамо, на пример, догађај: пао је број 4, јасно је да се тај догађај може, али и не мора остварити.

Овакву врсту догађаја који се у датој појави може остварити, али и не мора називамо **случајним догађајем.** *Појаве или процесе које ћемо проучавати преко случајних догађаја звaћемо* **експерименти**.

Скуп елементарних догађаја

У изучавањима случајних догађаја прво утврђујемо експеримент или појаву коју ћемо испитивати, а затим скуп могућих резултата посматраног експеримента. Посматрајмо следећих неколико примера:

При бацању правилног новчића утврђујемо да се експеримент састоји у бацању новчића и да је скуп могућих резултата експеримента:

$$\{\text{Писмо, Грб}\}$$

Расподела три службеника у три канцеларије врши се тако што се за првог службеника бира случајно једна од канцеларија, затим се за другог бира једна од преосталих и за трећег трећа канцеларија. Скуп могућих резултата овог експеримента је скуп следећих расподела службеника **a**, **b** и **c** по канцеларијама:

1.|**a**|**b**|**c**| 4. |**b**|**c**|**a**|

2.|**a**|**c**|**b**| 5. |**c**|**b**|**a**|

3.|**b**|**a**|**c**| 6. |**c**|**a**|**b**|

На једном месту се бележи број саобраћајних несрећа. Ако се посматрају дани у недељи и расподела **n** несрећа по данима, онда је скуп могућих резултата експеримента скуп расподела **n** несрећа у 7 дана. Различитих могућих резултата има укупно $\mathbf{n}^7$.

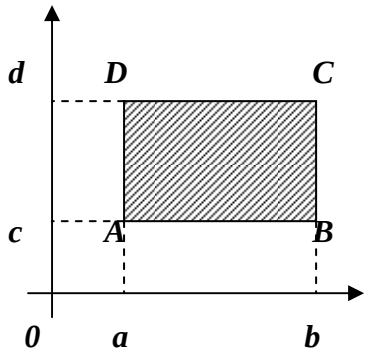
У мету се гађа 10 пута. Ако се бележи број погодака, скуп могућих резултата експеримента биће скуп бројева

$$\{0, 1,2, 3,4,5,6, 7,8,9, 10\}$$

Баца се коцка до првог падања шестице. Кад се бележи број бацања, скуп могућих резултата је скуп бројева

$$\{1,2,3,... \}$$

Из скупа брачних парова бира се на случајан начин један брачни пар и бележи се старост оба супружника. Скуп могућих резултата је било која тачка (x, y) из равни X0Y у правоугаонику ABCD датом на слици 3.1.



Слика 3.1.

Као што се види из ових примера, за сваки експеримент везујемо један скуп могућих резултата експеримента. У једном извођењу експеримента остварiће се тачно један од могућих резултата. Унапред се не може знати који ће то резултат бити. Кажемо да се ови могући резултати међусобно искључују. Поред тога сваки од њих представља случајан догађај јер се у извођењу експеримента може остварити, а и не мора.

Скуп могућих исхода експеримента зовемо **скуп елементарних догађаја**. *Било који резултат експеримента потпуно се описује једним и само једним елементом скупа могућих исхода.*

Из наведених примера се види да скуп елементарних догађаја могу бити слова, расподеле елемената, реални бројеви и сл. Поред тога скуп елементарних догађаја може бити: ***коначан; бесконачан, а пребројив; и бесконачно непребројив.***

Скуп елементарних догађаја може бити скуп тачака у једнодимензионалном простору, скуп тачака у дводимензионалном, тродимензионалном или вишедимензионалном простору.

На пример, при мерењу века трајања једне сијалице скуп елементарних догађаја је скуп тачака на реалној оси таквих да је $\{t \mid t \geq 0\}$.

Ако се на пример на случајан начин из скупа привредних предузетника бира једно предузеће и испитује се:

- број запослених,
- укупан доходак,
- трошкови пословања,
- просечан лични доходак,
- број повреда на раду;

онда ће скуп елементамих догађаја бити скуп тачака у петодимензионалном простору који одговара ограничењима везаним за свако посматрано обележје.

Посматрајмо сада произвољни случајни догађај А. За сваки елемент скупа елементарних догађаја знамо да ли његова реализација значи остварење догађаја А, или његово „не остварење". Другим речима, скуп елементарних догађаја можемо поделити на два подскупа:

1. подскуп елементарних догађаја чија реализација значи реализацију догађаја A, и
2. подскуп елементамих догађаја чија реализација значи да се догађај A није реализовао.

Први подскуп елементамих догађаја потпуно описује догађај A. Са друге стране, било какав подскуп скупа елементарних догађаја који садржи једну или више тачака може се посматрати као случајан догађај A који се остварује ако се оствари једна од тачака посматраног подскупа.

Према томе, случајни догађај се може дефинисати на следећи начин:

Дефиниција 3.1. Случајни догађај *је подскуп скупа елементарних догађаја.*

Случајни догађај може бити исказан речима. Његово проучавање могуће је једино ако одредимо одговарајући експеримент, скуп елементарних догађаја и подскуп који одговара датом случајном догађају.

На пример, нека случајни догађај A значи: при бацању коцке пао је број дељив са три. Овде је експеримент бацања коцке, а скуп елементарних догађаја је скуп бројева

$$\{1,2,3,4,5,6\}$$

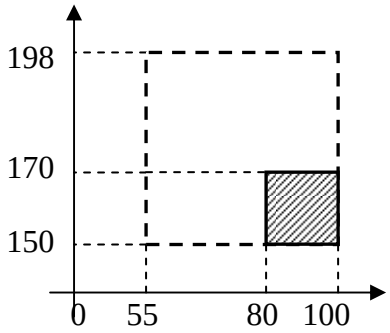
Случајном догађају A одговара подскуп чији су елементи бројеви:

$$\{3,6\}$$

јер остварење било ког од њих значи и остварење догађаја A. Са друге стране догађај A се остварује једино онда кад се оствари један од елемената посматраног подскупа.

Посматрајмо други пример. Нека се из једне велике групе студената бира, на случајан начин, један студент и нека се мере његова тежина X и висина Y. Ако се тежина свих посматраних студената креће у интервалу [55; 100], а висина у интервалу [150; 198], шта ће

представљати догађај A: „изабрани студент је тежи од 80 kg, а није виши од 180 cm".



Слика 3.2. Пример у 2 димензије

На слици 3.2. догађај A је скуп тачака у шрафираној површини.

Ако се у једном експерименту догађај не може остварити, тј. ако је то немогућ догађај, онда њему одговара подскуп скупа елементарних догађаја који нема ни један елемент, а то је празан подскуп. Према томе, немогућ догађај треба дефинисати на следећи начин:

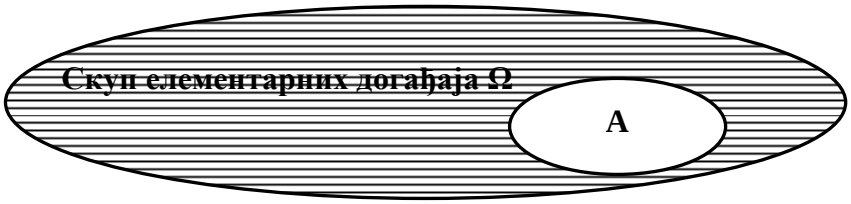
Дефиниција 3.2. **Немогућ догађај** *је подскуп скупа елементарних догађаја који нема ниједан елемент и означава се са Ø.*

Операције са случајним догађајима

Случајне догађаје дефинишемо као подскупове и међу њима можемо дефинисати извесне операције аналогне операцијама на скуповима. Наравно, о тим операцијама говорићемо као о операцијама међу случајним догађајима.

Дефиниција 3.3. *Нека је дат случајан догађај A.* **Супротан догађај** *је догађај* $\overline{A}$ *који се оствари онда и само онда кад се A не оствари.*

Супротном догађају $\overline{A}$ одговара подскуп елементарних догађаја који не припадају одговарајућем подскупу догађаја A. На слици 3.3. то су тачке у шрафираној површини.



Слика 3.3. Супротан догађај

На пример, бацају се истовремено четири новчића. Нека случајни догађај А значи: „ни једанпут није пало писмо". Тада ће супротан $\overline{A}$ догађај значити „бар једанпут је пало писмо".

Случајни догађај А је подскуп са једним елементом

$$A=\{0\}$$

а супротан догађај је подскуп

$$\overline{A}=\{1,2,3,4\}$$

Дефиниција 3.4. *Супротан догађај немогућег догађаја је* **сигуран догађај**. *Означаваћемо га са* **Ω**.

Немогућем догађају одговара празан подскуп. Сви елементи који нису у празном подскупу су елементарни догађаји. Зато је сигуран догађај уствари скуп елементамих догађаја.

Дефиниција 3.5. *Дата су два случајна догађаја А и В. Ако се сваки пут кад се оствари догађај А остварује и догађај В кажемо да* **А имплицира В** *и означавамо то са*

$$A \subseteq B \tag{3.1}$$

Релација (3.1) значи да сви елементи подскупа који одговара догађају А припадају и подскупу који одговара догађају В, тј. кад је А садржано у В.

Дефиниција 3.6. *За два случајна догађаја* **А и В** *кажемо да* **су једнаки** *ако истовремено А имплицира В и В имплицира А, тј. ако је*

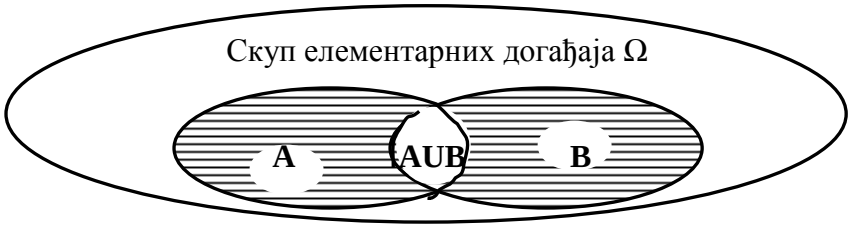
$$A \subseteq B \text{ и } B \subseteq A \tag{3.2}$$

Као што се на реалним бројевима дефинишу операције сабирања и множења тако се и на случајним догађајима дефинишу две операције. То су унија и пресек случајних догађаја.

Операција уније: *Нека су дати случајни догађаји A и B,* **унија догађаја A и B** *је догађај који се оствари онда и само онда кад се оствари бар један од догађаја A и B. Унију два догађаја означаваћемо са*

$$A \cup B \tag{3.3}$$

Случајним догађајима A и B одговарају два подскупа елементарних догађаја. Њиховој унији $A \cup B$ ће одговарати подскуп елементарних догађаја који припадају бар једном од подскупова везаних за догађаје A и B. На Слици 3.4. то су тачке шрафиране површине.



Слика 3.4. Унија два догађаја

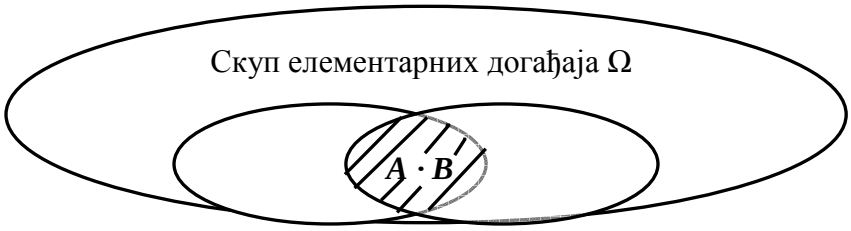
На пример, зна се да један производ може имати дефект прве врсте и дефект друге врсте. Нека случајни догађај A значи: „производ има дефект прве врсте", а догађај B значи: „производ има дефект друге врсте".

Њихова унија $A \cup B$ ће бити случајни догађај који ће значити: „производ има дефект ", или „производ је неисправан ".

Операција пресека: *Дата су два случајна догађаја A и B,* **пресек догађаја A и B** *је случајан догађај који се остварује онда и само онда кад се остваре оба догађаја, тј. кад се истовремено оствари и догађај A и догађај B. Пресек два догађаја означаваћемо са*

$$A \cdot B \text{ или } A \cap B \tag{3.4}$$

Догађајима А и В одговарају два подскупа елементарних догађаја. Њиховом пресеку $A \cdot B$ одговараће подскуп елементарних догађаја који истовремено припадају одговарајућим подскуповима догађаја А и В. То су уствари заједнички елементи подскупова А и В, као што се види на слици 3.5.



Слика 3.5. Пресек два догађаја

Нека се статистички скуп састоји од N студената међу којима је један број пушача. На случајан начин се бира један студент. Нека догађај А значи: „изабрана је студенткиња", а догађај В нека значи: „изабран је пушач". Њихов пресек $A \cap B$ биће случајан догађај који ће значити: „изабрана је студенткиња пушач".

Подскуп који одговара догађају А је подскуп студенткиња, а подскуп који одговара догађају В су пушачи (студенти и студенткиње). Подскуп који одговара пресеку $A \cap B$ су студенткиње пушачи, тј. они елементи који припадају А и В.

Дефиниција 3.7. *За два догађаја* **А и В** *кажемо да* **се међусобно искључују** *ако је њихов пресек немогућ догађај, тј. ако је*

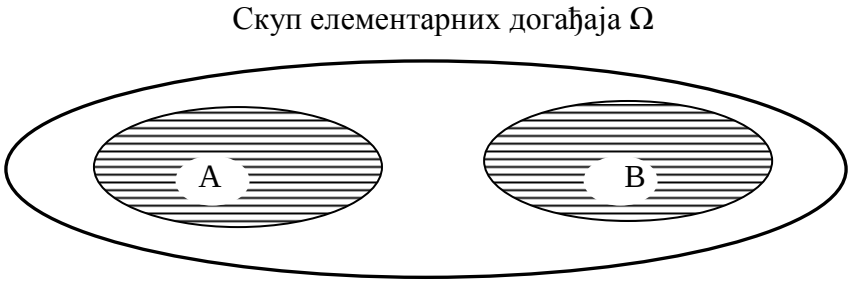
$$A \cap B = \varnothing \tag{3.5}$$

Догађајима који се међусобно искључују одговарају подскупови елементарних догађаја који немају заједничких тачака.

Ако се два догађаја А и В међусобно искључују тада је њихова унија збир тих догађаја, што означавамо са

$$A \cup B = A + B \tag{3.6}$$

Тада њиховој унији одговара подскуп елементарних догађаја који припадају или подскупу A или подскупу B. То су тачке у шрафираној површини на слици 3.6.



Слика 3.6. Збир два догађаја A+B

За операције уније и пресека који су дефинисани на случајним догађајима важе следећа три закона:

1. ***Комутативност:*** $A \cdot B = B \cdot A$ и $A \cup B = B \cup A$
2. ***Асоцијативност:*** $(A \cup B) \cup C = A \cup (B \cup C)$ и $(A \cdot B) \cdot C = A \cdot (B \cdot C)$
3. ***Дистрибутивност:*** $A \cdot (B \cup C) = (B \cdot A) \cup (A \cdot C)$.

За наведене дефиниције и операције на случајним догађајима важе следеће особине :

Особина 1. *A и* $\overline{A}$ *се међусобно искључују јер је*

$$A \cdot \overline{A} = \varnothing \tag{3.7}$$

Њихова унија је сигуран догађај

$$A + \overline{A} = \Omega \tag{3.8}$$

Особина 2. *Сваки случајан догађај A имплицира сигуран догађај* Ω

$$A \subseteq \Omega$$

Особина 3. *Ако A имплицира B, тада је*

$$A \cup B = B$$

$$A \cdot B = A$$

Особина 4. *Унија случајног догађаја A и немогућег догађаја* $\varnothing$ *је*

$$A \cup \varnothing = A$$

а њихов пресек је

$$A \cdot \varnothing = \varnothing$$

Особина 5. *Унија и пресек случајног догађаја A са самим собом је сам догађај A, тј.*

$$A \cup A = A$$

$$A \cdot A = A$$

Особина 6. *Посматрајмо случајни догађај*

$$A \cup B$$

који се остварује кад се оствари бар један од њих. Супротан догађај је догађај који се оствари кад се не оствари ниједан од њих, тј. кад се оствари пресек $\overline{A}$ *и* $\overline{B}$ *. Зато је*

$$\overline{A \cup B} = \overline{A} \cdot \overline{B}$$

На сличан начин се може закључити да је

$$\overline{A \cdot B} = \overline{A} \cup \overline{B}$$

У решавањима разних проблема везаних за случајне догађаје користе се наведене особине и операције и помоћу једноставнијих случајних догађаја описују се неки други случајни догађаји.

На пример, гађа се мета 3 пута и посматрају се случајни догађаји A_1, A_2 и A_3 који значе погађање мете у првом, другом и трећем гађању, респективно. Тада ће случајни догађај A који значи: „**мета је погођена са три поготка**" бити једнак

$$A = A_1 \cdot A_2 \cdot A_3$$

а догађај B: „**мета погођена бар једанпут**" једнак је

$$B = A_1 \cup A_2 \cup A_3$$

док случајни догађај C значи: „**мета није уопште погођена**" и једнак је

$$C = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$$

односно

$$C = \left(\overline{A_1 \cup A_2 \cup A_3}\right)$$

3.2. ВЕРОВАТНОЋА

Као што је већ речено, за описивање тзв. масовних појава и појава са случајним исходима служимо се случајним догађајима. Сад смо у стању да тачно опишемо случајне догађаје и да успоставимо извесне релације међу њима. Оно што је битно за посматрану појаву су законитости које треба да утврдимо. То ћемо моћи да урадимо онда кад будемо у стању да на известан начин **„меримо"** случајност код случајних догађаја. На основу **„измерених"** случајности могуће је квантитативно поређење случајних догађаја, а то ће нам омогућити и потпуније описивање посматране појаве.

Случајност код случајних догађаја се испитује и **„мери"** преко вероватноће која је у различитим тренуцима свог развоја била дефинисана различито. Ради бољег схватања суштине вероватноће, упознаћемо се са три дефиниције вероватноће и то редоследом историјског развоја ове теорије.

Класична дефиниција вероватноће

Класична дефиниција вероватноће одређује тзв. ***вероватноћу à priori*** и појам вероватноће се дефинише преко тзв. **једнако вероватних догађаја**. Класична дефиниција вероватноће омогућава да се одреди вероватноћа случајних догађаја једне релативно уске класе експеримената, а то је класа експеримената са **коначним бројем једнако вероватних могућих резултата.**

Посматрајмо експеримент коме одговара коначан број могућих резултата, рецимо **n**. Нека је сваки од **n** могућих резултата подједнако вероватан у следећем смислу: на основу експеримента немамо никаквог разлога да верујемо да ће **неки резултат наступити пре него неки други**.

На пример, ако бацамо **правилну** коцку, скуп могућих резултата су бројеви {1,2,3,4,5,6}.

Немамо никаквог разлога да верујемо да ће на пример, 2 пасти пре него 5 (под условом да је коцка правилног облика). Ако коцка није правилног облика не можемо унапред прихватити резултате као подједнако вероватне.

Дакле, претпоставићемо да скуп елементарних догађаја има **n** подједнако вероватних елемената. Посматрајмо случајни догађај A. Нека је **m** број елемената подскупа који одговара случајном догађају A. Кажемо да је **m** број „повољних" резултата експеримента за случајни догађај A.

Дефиниција 3.8. **Вероватноћа случајног догађаја A** *је однос броја „повољних" резултата за догађај A и броја могућих резултата експеримента. Вероватноћу случајног догађаја означаваћемо са*

$$P(\cdot)$$

при чему ћемо у загради уписати догађај и читати „Вероватноћа догађаја...".

Из дефиниције 3.8. следи да је

$$P(A)=\frac{m}{n} \tag{3.9}$$

при чему је **m** *број „повољних", а* **n** *број могућих резултата експеримента. Вероватноћа (3.9) се зове* **вероватноћа à priori .**

Пошто су **m** и **n** цели ненегативни бројеви, није тешко проверити да *вероватноћа à priori* има следеће особине:

Особина 1. *Вероватноћа било ког догађаја је ненегативан број, тј.*

$$P(A)\geq 0 \tag{3.10}$$

Особина 2. *Вероватноћа сигурног догађаја је једнака јединици. Ако је* Ω *сигурни догађај онда је за* Ω *број „повољних" резултата* $m = n$*, тако да је из (3.9)*

$$P(\Omega) = 1 \tag{3.11}$$

Особина 3. Ако се случајни догађаји A_1 и A_2 међусобно искључују, онда је број „повољних" резултата за њихову унију $A_1 + A_2$ једнак

$$m = m_1 + m_2$$

при чему је m_1 број „повољних" резултата за A_1, а m_2 број „повољних" за A_2.

На основу (3.9) добиће се да је

$$P(A_1 + A_2) = \frac{m_1}{n} + \frac{m_2}{n} \quad \text{односно}$$

$$P(A_1 + A_2) = P(A_1) + P(A_2) \tag{3.12}$$

Особина 4. *Вероватноћа немогућег догађаја једнака је нули, тј:*

$$P(\varnothing) = 0 \tag{3.13}$$

Особина 5. *Вероватноћа супротног догађаја једнака је*

$$P(\overline{A}) = 1 - P(A) \tag{3.14}$$

При одређивању à priori вероватноће треба бити доста опрезан. Прво што треба утврдити је број могућих резултата експеримента n. Друго, треба проверити да ли су ти резултати подједнако вероватни.

На пример, два пута се баца правилна коцка и бележи се збир тачкица у ова два бацања. Скуп могућих резултата су бројеви

$$\{2,3,4,5,6,7,8,9,10, 11, 12\}$$

Ови резултати нису подједнако вероватни, јер на пример резултат ће бити 2 ако је пала јединица у првом и другом бацању. Резултат бацања ће бити седам ако су пали следећи парови бројева у бацањима:

(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)

Зато је вероватније да ће резултат бити седам него да ће бити два. У овом примеру број могућих резултата је $n=6^2$, и то су парови бројева у бацањима

$$\begin{array}{c} (1.1)\ (2,1)\ \ldots\ (6,1) \\ (1.2)\ (2,2)\ \ldots\ (6,2) \\ \vdots\ \vdots\ \vdots \\ (1,6)\ (2,6)\ \ldots\ (6,6) \end{array} \qquad (3.15)$$

Сваки од резултата (3.15) је подједнако вероватан. Зато је вероватноћа да ће збир бити 2 једнака

$$P(2)=\frac{1}{36}$$

а вероватноћа да ће збир бити 7 једнака је

$$P(7)=\frac{6}{36}$$

Често се непотпуном анализом лако прихвати закључак о једнако вероватним исходима, што може довести до погрешних закључака.

Такав је био случај са тзв. проблемом **CHEAVALIER DE MÉRÉ**: Посматрају се два експеримента:

а) баца се правилна коцка **4 пута** и посматра се догађај да **бар једанпут падне *шестица;***

б) бацају се истовремено две коцке **24 пута** и посматра се догађај да **бар једанпут падне *пар шестица***.

Проблем је да се одреди који од два догађаја је вероватнији.

Извесно време се сматрало да су оба догађаја једнако вероватна јер се закључивало на следећи начин: број различитих резултата са две коцке 6 пута је већи од броја резултата са једном коцком. Пошто у а) бацамо коцку 4 пута а у б) две коцке бацамо 6 пута више (тј. 24 пута) онда су исте шансе да падне једна шестица у 4 бацања једне коцке као и падања пара шестица у 24 бацања пара коцки.

Међутим, кад је била организована игра на бази ова два експеримента утврђено је било да у великом броју играња чешћи је био добитак базиран на игри под а) од добитка базираног на игри под б).

Заиста, у бацању једне коцке број могућих резултата је

$$n = 6^4$$

а број резултата да ниједанпут не падне 6 је

$$m = 5^4$$

Зато је тражена вероватноћа

$$P = 1 - \left(\frac{5}{6}\right)^4 = 0.516 \tag{3.16}$$

као вероватноћа супротног догађаја.

У случају бацања пара коцки, број могућих резултата је

$$n = 36^{24}$$

а међу њима број резултата да не падне ниједанпут пар шестица је

$$m = 35^{24}$$

Тражена вероватноћа је једнака:

$$P = 1 - \left(\frac{35}{36}\right)^{24} = 0.491 \tag{3.17}$$

Из (3.16) и (3.17) се види да је вероватнији догађај под а) од догађаја под б), што је сасвим у складу са експериментално утврђеном чињеницом на бази игара.

Класичну дефиницију вероватноће могуће је у извесним случајевима проширити и на експерименте са бесконачно могућих резултата. Тако се добија тзв. ***геометријска вероватноћа***.

На пример, нека је у равни *X0Y* дата ограничена област G и унутар ње друга област g. Замислимо следећи експеримент: произвољно мала куглица се на случајан начин баци у област G и посматра се тачка у коју ће пасти куглица. Свака тачка области G је могући резултат експеримента и све тачке су ,,подједнако вероватне".

Посматрајмо случајни догађај A који значи: ,,бачена куглица је пала у тачку области g". Тада је вероватноћа догађаја A једнака

$$P(A)=\frac{Pov(g)}{Pov(G)} \tag{3.18}$$

тј. вероватноћа је једнака односу површина области g и G.

Дефиниција геометријске вероватноће се може проширити и на једнодимензионалан, тродимензионалан или вишедимензионалан простор. У једнодимензионалном простору посматра се коначан интервал дужине D и интервал дужине d. Ако је свака тачка интервала D ,,подједнако вероватна" као могућ резултат експеримента онда је вероватноћа да ће се случајно изабрана тачка наћи у интервалу чија је дужина d једнака

$$P(A)=\frac{d}{D} \tag{3.19}$$

при чему је интервал дужине d садржан у интервалу дужине D.

У тродимензионалном простору, геометријска вероватноћа је однос одговарајућих запремина

$$P(A)=\frac{V(g)}{V(G)} \tag{3.20}$$

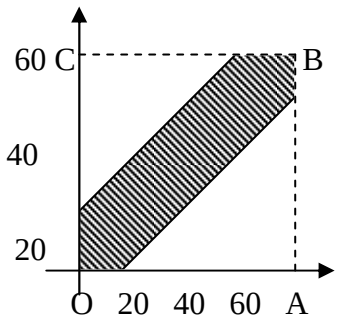
а у вишедимензионалном простору то је однос одговарајућих хиперзапремина.

На пример, у рулету кружница је подељена на 37 поља која су наизменично обојена црвеном и црном бојом. Једно поље је остало беле боје. Свако поље носи број 1, 2, ..., или 36, а бело поље има број 0. Учесници у игри стављају одређен износ на поље које желе. Куглица кружи ивицом круга који се такође врти, али у обрнутом смеру од куглице. Свака тачка кружнице је подједнако вероватна. Зато свако поље има исту вероватноћу која је једнака

$$P(A)=\frac{1}{37}$$

а то је однос дужине кружнице која одговара исечку и обима кружнице.

ПРОБЛЕМ СУСРЕТА. Посматрајмо следећи пример. Два лица се договоре да се нађу на договореном месту у временском интервалу од једног сата. Први који дође на уговорено место чека 20 минута, после чега одлази. Ако сваки од њих може доћи у било којем тренутку посматраног сата, а њихови доласци су независни међусобно, колика је вероватноћа сусрета? На слици 3.7. приказан је посматрани експеримент.



Slika 3.7 Проблем сусрета

На оси x нанешен је тренутак доласка првог лица, а на оси y тренутак доласка другог лица. Могући резултати су све тачке у правоугаонику OABC.

До сусрета ће доћи ако је

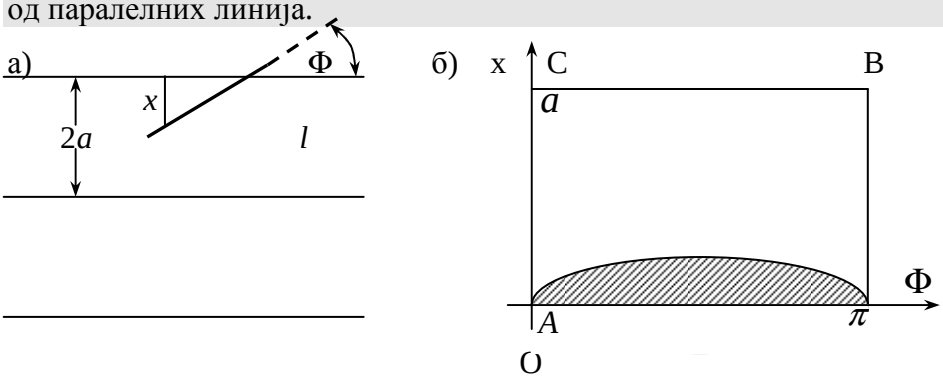
$$|x-y| < 20$$

а то су тачке у шрафираној површини. Зато је тражена вероватноћа

$$P = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}$$

Buffon-ов проблем: *Доста важну улогу у историји рачуна вероватноће имао је проблем познат под именом* **Buffon-ов проблем игле.**

Танка игла дужине $2l$ баца се на случајан начин на равну површину која је подељена паралелним линијама удаљеним за $2a$ једна од друге $(l < a)$. Тражи се вероватноћа да ће игла пресећи једну од паралелних линија.



Слика 3.8 Buffon-ов проблем

Означимо са x удаљеност средине игле од најближе праве линије, а са Φ угао који игла заклапа са правим линијама. Величина x може бити било који број из интервала $[0:a]$, а угао Φ може бити било који број из интервала $[0:\pi]$. На слици 3.8 б) приказан је правоугаоник OABC чија свака тачка може бити могућ резултат експеримента.

На слици 3.8 а) се види да ће игла пресећи праву једино кад је

$$x \le l \sin \Phi$$

а то су тачке у шрафираној површини слике 3.8 б).

Зато је тражена вероватноћа једнака

$$P = \frac{\int_0^{\pi} \sin \Phi d\Phi}{a\pi}$$

односно

$$P = \frac{2}{a\pi} \tag{3.21}$$

Из (3.21) добија се да је

$$\pi = \frac{2}{a} \cdot \frac{1}{P} \tag{3.22}$$

Израз (3.22) послужио је да се експериментално одреди вредност броја π. Ако се игла баци велики број пута и бележи се број бацања у којима игла пресече праву, вероватноћа P ће приближно бити једнака

$$P \approx \frac{\textit{број пресецања праве}}{\textit{број бацања игле}}$$

Тако је на пример 1894 године извршено 1120 бацања игле и добијена вредност за π је била

$$\pi = 3{,}1419$$

а 1901 године извршено је 3408 бацања и добијена је вредност

$$\pi = 3{,}1415929$$

Емпиријска вероватноћа

Посматрајмо случај кад изаберемо једно лице и испитујемо његову старост. Ако је T време које доживи посматрано лице, онда је скуп могућих резултата „експеримента" скуп ненегативних бројева t, при чему је

$$0 \le t \le T_0$$

где је T_0 највећа старост коју може доживети неко лице.

Из искуства знамо да је вероватније да ће посматрано лице доживети 60 година, него да ће доживети, рецимо, 100 година. Према томе, скуп елементарних догађаја не можемо одредити тако да сваки елемент буде подједнако вероватан, па самим тим не можемо одредити ни вероватноће случајних догађаја везаних за овај „експеримент".

Зато је и дефинисана *емпиријска вероватноћа* или ***вероватноћа à posteriori*** која ће нам омогућити одређивање вероватноћа и за ону класу експеримената који доводе до неједнако вероватних резултата.

Претпоставимо да експеримент можемо поновити неограничен број пута. Посматраћемо случајни догађај A везан за тај експеримент.

Ако експеримент поновимо ***n*** пута и при том се посматрани догађај A реализовао тачно ***n(A)*** пута, тада се количник

$$f = \frac{n(A)}{n} \tag{3.23}$$

назива *релативна фреквенција догађаја A*.

Дефиниција 3.9. *Ако број понављања експеримента неограничено расте и ако релативна фреквенција догађаја A тежи неком коначном броју, онда ту граничну вредност називамо вероватноћом догађаја A. Вероватноћа одређена дефиницијом 3.9. назива се* **емпиријска вероватноћа** *или* ***вероватноћа*** **à posteriori**. *Често се та вероватноћа зове и* **статистичка вероватноћа**.

У ствари, **емпиријска вероватноћа** случајног догађаја A је једнака

$$P(A) = \lim_{n\to\infty} \frac{n(A)}{n} \tag{3.24}$$

Основни проблем у дефинисању емпиријске вероватноће је проблем егзистенције и једнозачности граничне вредности (3.24). Ако та **гранична вредност не постоји и није једнозначна** за неке случајне догађаје, онда нема смисла ни вероватноћа дефинисана том граничном вредношћу.

Гранична вредност (3.24), односно *à posteriori* вероватноћа може се одредити експерименталним путем.

Тако је на пример, Buffon извео експеримент који се састојао у 4040 бацања правилног новчића. Добио је, да је релативна фреквенција појављивања „Писма" била

0,5080

Pearson је овај експеримент поновио 12000 пута и добио је релативну фреквенцију

0,5016

а на основу 24000 понављања овог експеримента добио је вредност

0,5005

На основу ових података се може закључити да је вероватноћа падања „Писма" при бацању правилног новчића једнака 0,5.

Или на пример, таблица случајних бројева се прави тако што се на случајан начин бирају бројеви

{0, 1, 2, 3, 4, 5, 6, 7, 8, 9}

и свакој цифри се приписује иста вероватноћа која је једнака 0,10.

Ако се у таблици случајних бројева посматра релативна фреквенција појављивања рецимо седмице добиће се следећи резултати:

Број цифара у таблици случајних бр.	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
Фреквенција појављивања броја 7	0,095	0,088	0,095	0,120	0,095	0,099	0,082	0,089	0,111	0,102

Види се да је гранична вредност релативних фреквенција појављивања седмице једнака 0,10, односно вероватноћа појављивања седмице је једнака 0,10.

Међутим, то ипак није потпуно решење проблема егзистенције емпиријске вероватноће. Ови примери потврђују претпоставку да постоји гранична вредност (3.24) за посматране експерименте и посматране случајне догађаје. Потпуно решење овог проблема могуће је имати тек онда кад се докаже да гранична вредност (3.24) постоји и за све друге експерименте и за све друге случајне догађаје. То ће бити предмет изучавања 4. поглавља (Закон великих бројева).

За *à posteriori* вероватноћу може се лако проверити да поседује све особине као і *à priori* вероватноћа:

1. $0 \le P(A) \le 1$,
2. $P(\Omega) = 1$,
3. $P(\varnothing) = 0$,
4. $P(A_1 + A_2) = P(A_1) + P(A_2)$,
5. $P(\overline{A}) = 1 - P(A)$,
6. $P(A \cup B) = P(A) + P(B) - P(AB)$.

Аксиоматска дефиниција вероватноће

Да би одредили емпиријску вероватноћу случајних догађаја потребно је извести велик број пута посматрани експеримент и тек онда утврдити граничну вредност (3.24), односно вероватноћу. Међутим, у пракси најчешће нисмо у стању да многе експерименте поновимо више пута.

На пример, ако се одређује вероватноћа судара космичког брода са васионским честицама немогуће је то урадити експерименталним путем. Исто тако, ако треба одредити вероватноћу да одређени производ има жељену тражњу никако то не можемо урадити експериментално.

Према томе, још увек остаје проблем како одредити вероватноће случајних догађаја везаних за неке експерименте.

Савремена дефиниција вероватноће омогућава развој теорије која даје решења за одређивање вероватноће било ког случајног догађаја. Посматра се скуп случајних догађаја и пресликавање тог скупа на скуп реалних бројева, тако што ће се сваком случајном догађају A придружити реалан број *P(A)*.

Дефиниција 3.10. *Вероватноћа случајних догађаја, је свака функција P дефинисана на случајним догађајима која пресликава случајне догађаје у реалне бројеве и која има следеће* ***три особине****:*

1. Ненегативност. *Сваком случајном догађају A функција P придружује ненегативан број P(A) тј.*

$$P(A) \geq 0 \tag{3.25}$$

2. Нормираност. *Сигурном догађају функција P придружује број један, тј.*

$$P(\Omega) = 1 \tag{3.26}$$

3. Адитивност. *Ако се случајни догађаји* $A_1, A_2, \ldots$ *међусобно искључују, онда функција P придружује њиховој унији број који је једнак збиру бројева које функција P придружује догађајима* $A_1, A_2, \ldots$, *тј.*

$$P(A_1 + A_2 + \cdots) = P(A_1) + P(A_2) + \cdots \tag{3.27}$$

Ове три особине представљају **аксиоме** који одређују вероватноћу. Све остале особине и примена целе теорије вероватноће базирана је на овим аксиомима.

Одмах се намеће ***питање:*** Да ли *à priori* вероватноћа и емпиријска вероватноћа имају особине 1. 2. и 3. којима је дефинисана вероватноћа? Другим речима, да ли се може сматрати да су те две вероватноће које смо раније дефинисали за одређене класе експеримената, уједно и вероватноће које одговарају савременој дефиницији вероватноће?

Одговор је позитиван. Из дефиниција 3.7. и 3.8. лако се проверава да ове две вероватноће имају особину ненегативности, нормираности и адитивности.

Из аксиома ненегативности, нормираности и адитивности следе особине вероватноће.

Особина 1. *Вероватноћа супротног догађаја је једнака*

$$P(\overline{A}) = 1 - P(A). \tag{3.28}$$

Заиста, посматрајмо догађаје A и $\overline{A}$.То су догађаји који се међусобно искључују. На основу аксиома адитивности имали би

$$P(A + \overline{A}) = P(A) + P(\overline{A}) \tag{3.29}$$

Са друге стране њихова унија је сигуран догађај тј.

$$\overline{A} + A = \Omega$$

па је на основу аксиома нормираности

$$P(A + \overline{A}) = 1 \tag{3.30}$$

Кад се (3.30) замени у (3.29) добиће се (3.28).

Особина 2. *Вероватноћа немогућег догађаја је једнака нули, тј.*

$$P(\varnothing) = 0 \tag{3.31}$$

Пошто се догађаји Ω и $\varnothing$ међусобно искључују и такви су да је

$$\Omega + \varnothing = \Omega$$

на основу аксиоме адитивности следи

$$P(\Omega) + P(\varnothing) = P(\Omega)$$

односно (3.31).

Особина 3. *Вероватноћа било ког случајног догађаја је број који се налази између 0 и 1, тј.*

$$0 \le P(A) \le 1 \tag{3.32}$$

Особина 4. *Ако догађај A имплицира догађај B, тада су њихове вероватноће такве да је*

$$P(A) \le P(B) \tag{3.33}$$

Заиста из $A \subseteq B$ следи да је

$$B = A + \overline{A} \cdot B$$

па на основу адитивности следи

$$P(B) = P(A) + P(\overline{A} \cdot B) \tag{3.34}$$

Како је на основу особине ненегативности

$$P(\overline{A} \cdot B) \ge 0$$

из (3.34) следи (3.33).

Особина 5. Вероватноћа уније два догађаја A и B је једнака

$$P(A \cup B) = P(A) + P(B) - P(A \cdot B) \tag{3.35}$$

Са слике 4.9. се види да је

$$A \cup B = A + \overline{A} \cdot B$$

(3.36)

$$B = A \cdot B + \overline{A} \cdot B$$

при чему се A и $\overline{A} \cdot B$ међусобно искључују. Исто тако се і $A \cdot B$ и $\overline{A} \cdot B$ такође међусобно искључују.

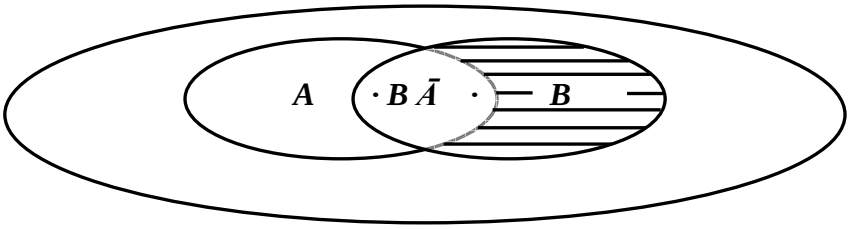
Због адитивности из (3.36) се добија

$$P(A \cup B) = P(A) + P(\overline{A} \cdot B)$$

(3.37)

$$P(B) = P(A \cdot B) + P(\overline{A} \cdot B)$$

Одузимањем од прве једначине другу у (3.37) добиће се (3.35).



Слика 4.9. Унија догађаја

3.3. УСЛОВНЕ ВЕРОВАТНОЋЕ И НЕЗАВИСНОСТ

У испитивањима одређених појава и случајних догађаја везаних за појаве често се поставља проблем међусобне зависности случајних догађаја, тј. поставља се проблем како реализација једног случајног догађаја утиче на реализацију другог.

Посматрајмо два случајна догађаја A и B. Постоје две крајности у њиховом међусобном односу. Прво, догађаји A и B су такви да из реализације догађаја B следи сигурна реализација догађаја A. Друго, реализација догађаја B искључује могућност реализације догађаја A. Остали међусобни односи догађаја A и B налазе се између ове две крајности.

Међусобна веза догађаја A и B у теорији вероватноће проучава се преко тзв. ***условне вероватноће***.

Означимо са

$$P(A/B)$$

условну вероватноћу реализације догађаја A под условом да се реализовао догађај B.

Посматрајмо пример једног експеримента са коначним бројем могућих исхода који су подједнако вероватни. Нека је ***n*** укупан број могућих исхода, а ***n(B)*** број исхода у којима се реализује догађај B. Нека је ***n(AB)*** број могућих исхода који доводе до реализације оба догађаја.

Тада је вероватноћа догађаја B једнака

$$P(B)=\frac{n(B)}{n} \tag{3.38}$$

а вероватноћа реализације и догађаја A и догађаја B је

$$P(AB) = \frac{n(AB)}{n} \tag{3.39}$$

Сад посматрајмо догађај A/B, тј. догађај да се А реализује под условом да се В реализовао. Ако је услов испуњен онда постоји *n(B)* могућих исхода а међу њима је *n(AB)* повољних за догађај А.

Зато је условна вероватноћа догађаја A/B једнака

$$P(A/B) = \frac{n(A \cdot B)}{n(B)} \tag{3.40}$$

Ако у изразу (3.40) и бројилац и именилац поделимо са n и заменимо вредности из (3.38) и (3.39) добићемо релацију

$$P(A/B) = \frac{P(A \cdot B)}{P(B)} \tag{3.41}$$

Одавде се види да условну вероватноћу треба дефинисати преко количника (3.41).

<u>Дефиниција 3.11.</u> *Условна вероватноћа догађаја А, под условом да се реализује догађај В једнака је количнику*

$$P(A/B) = \frac{P(A \cdot B)}{P(B)}$$

тј. једнака је количнику вероватноће истовременог остварења догађаја А и В и вероватноће догађаја В, $(P(B) > 0)$.

Условна вероватноћа задовољава сва три аксиома којима је дефинисана вероватноћа. Наиме, из количника (3.41) се види:

1. $P(A/B) \geq 0$
2. $P(A/B) = 1$ и то у случају кад догађај В имплицира А, тј. кад је $B \subseteq A$

3. Ако се А може изразити као унија догађаја A_1, A_2, ... који се међусобно искључују, онда се и условни догађај A/B може изразити као унија условних догађаја A_1/B, A_2/B, ..., па је

$$P(A/B)=\sum_k P(A_k/B)$$

За одређивање вероватноће пресека два догађаја користи се условна вероватноћа. Из (3.41) се добија да је

$$P(AB)=P(B)\cdot P(A/B) \tag{3.42}$$

Ако А и В замене места добиће се релација

$$P(AB)=P(A)\cdot P(B/A) \tag{3.43}$$

У практичним проблемима често је лакше одредити условне вероватноће, па се релације (3.42) и (3.43) могу користити за одређивање вероватноћа истовременог остварења догађаја А и В.

Формула потпуне (тоталне) вероватноће

У многим примерима вероватноћу случајног догађаја је тешко одредити без посматрања других случајних догађаја. На пример, из једне кутије у којој има 6 белих и 4 црне куглице извлаче се две куглице једна за другом без враћања и посматра се догађај да у другом извлачењу буде извучена бела куглица. Наравно, реализација тог догађаја зависи од тога која је куглица била извучена у првом извлачењу. Зато овај догађај морамо изразити преко догађаја везаних за прво извлачење, а затим на основу тих догађаја одредити и вероватноћу посматраног догађаја. За решавање оваквих проблема служи тзв. ***формула потпуне (тоталне) вероватноће:***

Нека су

$$B_1, B_2, ..., B_n$$

случајни догађаји који се међусобно искључују и чија унија је сигуран догађај, тј.

$$B_i \cdot B_j = \varnothing$$

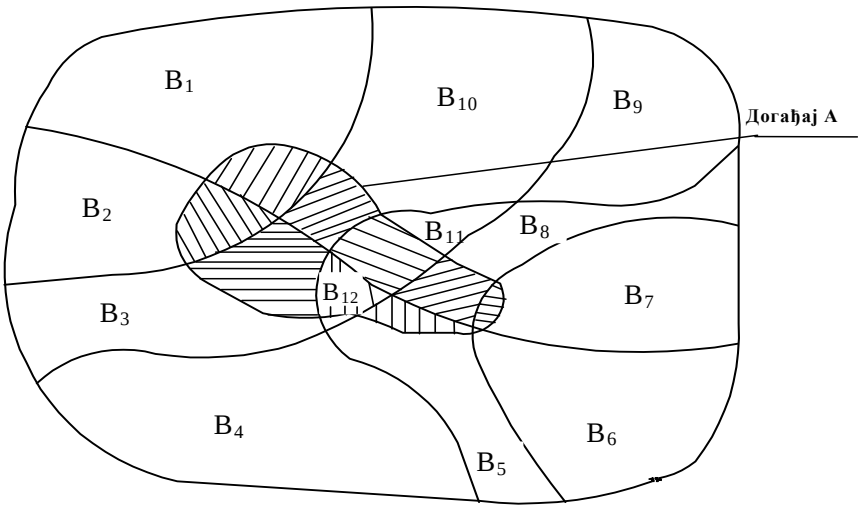
за свако $i \neq j,\ i, j \in \{1, 2, ..., n\}$,

$$\sum_{k=1}^{n} B_k = \Omega$$

Тада је вероватноћа догађаја A једнака

$$P(A) = \sum_{k=1}^{n} P(B_k) \cdot P(A/B_k) \qquad (3.44)$$

при чему су $P(B_k)$ вероватноће догађаја B_k, а $P(A/B_k)$ условне вероватноће догађаја A под условом да се реализовао B_k.



Слика 3.9.

Доказ. Догађаји B_k чине разбијање скупа могућих исхода Ω. Подскуп који одговара догађају A се састоји од подскупова који су заједнички са A и са B_k, (Слика 3.9.) тако да је

$$A = AB_1 + AB_2 + \cdots + AB_k \tag{3.45}$$

Вероватноћа сваког од догађаја на десној страни једначине (3.45) може се изразити преко условне вероватноће

$$P(AB_k) = P(B_k)P(A/B_k)$$

па на основу аксиома адитивности и релације (3.45) следи (3.44).

Узмимо следећи пример. Један производ се припрема за завршну операцију која се обавља успешно ако је припрема производа била добра. Припрема се обавља на једној од четири машине. Нека је поузданост машина 0,90; 0,91; 0,90 и 0,92 респективно за прву, другу, трећу и четврту машину.

Расподела производа по машинама за припрему се врши случајно и то тако да се другој и четвртој машини додељује по два посто више производа него машинама један и три. Колика је вероватноћа да ће производ бити успешно обрађен у завршној операцији?

Да би одредили тражену вероватноћу означићемо са:

- B_1 − догађај да ће производ у припреми доћи на 1. машину;
- B_2 − догађај да ће производ у припреми доћи на 2. машину;
- B_3 − догађај да ће производ у припреми доћи на 3. машину;
- B_4 − догађај да ће производ у припреми доћи на 4. машину.

Са A ћемо означити догађај да ће производ бити успешно обрађен. Тада је

$$P(B_1) = 0{,}244$$

$$P(B_2) = 0{,}256$$

$$P(B_3) = 0{,}244$$

$$P(B_4) = 0{,}256$$

а условне вероватноће су

$$P(A/B_1) = 0{,}90$$

$$P(A/B_2) = 0{,}91$$

$$P(A/B_3) = 0{,}90$$

$$P(A/B_4) = 0{,}91$$

Догађаји B_1, B_2, B_3 и B_4 се међусобно искључују и њихова унија је сигуран догађај. Зато се вероватноћа догађаја A може одредити преко формуле потпуне вероватноће, тако да је

$$P(A) = 0{,}244 \cdot 0{,}90 + 0{,}256 \cdot 0{,}91 + 0{,}244 \cdot 0{,}90 + 0{,}256 \cdot 0{,}91$$

тј.

$$P(A) = 0{,}90512$$

Посматраћемо још један **пример игре**. Баца се правилан новчић. Играч пре бацања „погађа" која ће страна новчића пасти. Ако погоди страну добија динар, а ако не погоди плаћа динар. Играч поседује X динара и решио је да игра све док не повећа свој износ на A динара. Колика је вероватноћа да ће играч изгубити целу суму X пре него што оствари жељу?

Означићемо са

$$p(X)$$

вероватноћу да ће играч изгубити суму X. После првог бацања његова сума или ће бити (X+1) динара или (X-1).

Вероватноћа да у наредним бацањима изгуби суму X+1 биће

$$p(X+l)$$

а вероватноћа да изгуби суму X-1 биће

$$p(X-1)$$

Нека је B_1 догађај да играч у првом бацању погоди, а B_2 догађај да не погоди. Тада се B_1 и B_2 искључују и њихова унија је сигуран догађај. Ако је A догађај да играч изгуби суму X онда је на основу формуле потпуна вероватноћа

$$p(X) = 0{,}5p(X+l)+0{,}5p(X-l)$$

Решење ове једначине је линеарна функција

$$p(X) = C_1+C_2X \tag{3.46}$$

при чему C_1 и C_2 треба одредити из почетних услова. Пошто је вероватноћа да играч изгуби суму од нула динара једнака

$$p(0) = 1 \tag{3.47}$$

а вероватноћа да изгуби суму од A динара

$$p(A) = 0 \tag{3.48}$$

(јер кад оствари суму од A динара престаје да игра), то значи да ће се из (3.46), (3.47) и (3.48) добити

$$C_1 = 1;\ C_2 = -\frac{X}{A}$$

Тражена вероватноћа губитка почетне суме од X динара је једнака

$$p(X) = 1-\frac{X}{A}$$

Bayes-ова формула

Посматрајмо пример продавнице која продаје један производ. Нека тај производ продавница набавља од два произвођача и то у односу 2:1. Зна се да је 1% производа првог произвођача неисправно, а 2% код другог произвођача. За један производ случајно изабран у продавници утврди се да је неисправан. Колика је вероватноћа да је то производ другог произвођача?

Означимо са B_1, B_2 и A следеће догађаје

- B_1: изабрани производ је производ првог произвођача
- B_2: изабрани производ је производ другог произвођача
- A: изабрани производ је неисправан.

Тада је вероватноћа догађаја B_1 једнака

$$P(B_1) = \frac{2}{3}$$

а вероватноћа догађаја B_2 је

$$P(B_2) = \frac{1}{3}$$

Условне вероватноће неисправности производа су

$$P(A / B_1) = 0,01$$
$$P(A / B_2) = 0,02$$

па је вероватноћа догађаја A

$$P(A) = \frac{2}{3}0,01 + \frac{1}{3}0,02$$

Међутим, зна се да се догађај A реализовао и тражи се вероватноћа да је изабрани производ другог произвођача тј. тражи се условна вероватноћа догађаја B_2 под условом кад се реализовао A. Значи тражи се

$$P(B_2 / A)$$

На основу дефиниције условне вероватноће, тражена вероватноћа је једнака

$$P(B_2 / A) = \frac{P(A \cdot B_2)}{P(A)}$$

односно

$$P(B_2 / A) = \frac{P(B_2) \cdot P(A / B_2)}{P(A)} \tag{3.49}$$

јер је $P(A \cdot B_2) = P(B_2) \cdot P(A / B_2)$.

На тај начин ћемо добити

$$P(B_2 / A) = \frac{\frac{1}{3} 0{,}02}{\frac{2}{3} 0{,}01 + \frac{1}{3} 0{,}02} = 0{,}5$$

Израз (3.49) познат је под називом **Bayes-ова формула**. *Служи за одређивање условне вероватноће извесних догађаја који су претходили реализацији једног случајног догађаја.*

Посматраћемо општији случај. Нека је дат низ случајних догађаја

$$B_1, B_2, ..., B_n$$

који се међусобно искључују и чија унија је сигуран догађај Ω, тј. таквих да је

$$B_i \cdot B_j = \varnothing$$

за свако $i \neq j$, $i, j \in \{1, 2, ..., n\}$,

$$B_1 + B_2 + \cdots + B_n = \Omega$$

Нека је A неки случајни догађај везан за посматрани експеримент. Тада је за сваки догађај B_k, условна вероватноћа једнака

$$P(B_k / A) = \frac{P(A \cdot B_k)}{P(A)} \tag{3.50}$$

Бројилац у изразу (3.50) може се изразити са

$$P(A \cdot B_k) = P(B_k) \cdot P(A / B_k)$$

а за именилац се може користити формула потпуне вероватноће. Тако се добија:

Bayes-ова формула. *За сваки догађај* B_k *условна вероватноћа његове реализације кад се реализовао догађај A је једнака*

$$P(B_k / A) = \frac{P(B_k) \cdot P(A / B_k)}{\sum_{i=1}^{n} P(B_i) \cdot P(A / B_i)} \tag{3.51}$$

Нека се на пример у једној лабораторији налазе три различите полице са епруветама у којима је крв из три различите крвне групе. У првој полици су 2 епрувете са првом крвном групом, 1 са другом и 3 са трећом. У другој полици су 3 са првом, 2 са другом и 4 са трећом, а на трећој полици су 4 са првом, 3 са другом и 2 са трећом. Позитивна реакција се постиже једино ако се помешају трећа и друга крвна група. Лаборант је на случајан начин изабрао једну полицу, а затим са те полице изабрао две епрувете. Експеримент је показао позитивну реакцију помешане крви из те две епрувете. Колика је вероватноћа да су то биле епрувете са прве полице?

Означимо са B_1, B_2 и B_3 случајни догађај да ће изабране епрувете бити са прве, друге, односно треће полице. Са A ћемо означити случајан догађај да ће се постићи позитивна реакција. Тада је

$$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$$

Условне вероватноће су

$$P(A / B_1) = \frac{3}{6} \cdot \frac{1}{5} + \frac{1}{6} \cdot \frac{3}{5} = \frac{1}{5}$$

$$P(A / B_2) = \frac{2}{9}$$

$$P(A / B_3) = \frac{1}{6}$$

Примењујући Bayes-ову формулу добиће се

$$P(B_1 / A) = \frac{18}{53}$$

Независни догађаји

Посматраћемо два експеримента који се изводе независно један од другог.

Нека је A_1 *случајни догађај везан за први експеримент, а* A_2 *случајни догађај везан за други експеримент. Тада реализација догађаја* A_1 *не утиче на реализацију догађаја* A_2 *и кажемо да су догађаји* A_1 *и* A_2 ***међусобно независни.***

Наравно, основни проблем везан за оваква два догађаја је проблем одређивања вероватноће њиховог истовременог остварења, тј. проблем одређивања вероватноће пресека независних догађаја A_1 и A_2.

Други проблем представља утврђивање независности за оне догађаје који су везани за један исти експеримент. Одредићемо вероватноћу пресека независних догађаја преко емпиријске вероватноће, а затим ћемо тај резултат уопштити и за остале случајеве.

Претпоставимо да два посматрана експеримента можемо поновити неограничен број пута. Нека смо истовремено поновили оба експеримента *n* пута.

Означимо са $n(A_2)$ број понављања експеримената у којима се реализовао догађај A_2, а са $n(A_1A_2)$ број понављања експеримената у којима су се реализовали истовремено и A_1 и A_2. Тада је емпиријска вероватноћа приближно једнака

$$P(A_2) \approx \frac{n(A_2)}{n}$$
$$P(A_1A_2) \approx \frac{n(A_1A_2)}{n} \tag{3.52}$$

А сад посматрајмо само она понављања у којима се реализовао догађај A_2. Њих има укупно $n(A_2)$. Међу њима биће и неких у којима се реализовао и догађај A_1. Таквих има укупно $n(A_1A_2)$. Ако је $n(A_2)$ довољно велико, тада је емпиријска вероватноћа догађаја A_1 приближно једнака

$$P\left(A_1\right) = \frac{n\left(A_1A_2\right)}{n\left(A_2\right)} \tag{3.53}$$

Вероватноћа пресека A_1A_2 у (3.52) се може написати у облику

$$P(A_1A_2) = \frac{n(A_1A_2)}{n(A_2)} \cdot \frac{n(A_2)}{n}$$

одакле се добија да је

$$P(A_1 \cdot A_2) = P(A_1) \cdot P(A_2) \tag{3.54}$$

Израз (3.54) служи за дефиницију независности догађаја.

Дефиниција 3.12. *Нека су A_1 и A_2 било каква два случајна догађаја. Рећи ћемо да су A_1 и A_2,* **међусобно независни,** *ако и само ако је*

$$P(A_1 \cdot A_2) = P(A_1) \cdot P(A_2) \tag{3.55}$$

Независност случајних догађаја може се посматрати и преко условних вероватноћа. Наиме, ако су догађаји A_1 и A_2 међусобно независни онда је условна вероватноћа, једнака

$$P(A_1 / A_2) = P(A_1)$$
$$P(A_2 / A_1) = P(A_2) \tag{3.56}$$

Обрнуто, ако је условна вероватноћа једнака

$$P(A_1 / A_2) = P(A_1)$$

тада је

$$P(A_1 \cdot A_2) = P(A_2) \cdot P(A_1 / A_2) = P(A_1) \cdot P(A_2)$$

тј. тада су A_1 и A_2 независни. Зато се може рећи:

Догађаји A_1 и A_2 су **међусобно независни**, *ако и само ако је условна вероватноћа једног од њих, под условом да се реализовао други, једнака вероватноћи тог догађаја.*

У неким примерима није лако одмах на основу експеримента утврдити независност случајних догађаја. Тада се мора проверавати тачност израза (3.55) или (3.56) и тек онда утврдити независност.

На пример, из шпила од 36 карата на случајан начин се бира карта. Нека догађај A_1 значи: „изабрана карта је треф"; а A_2: „изабрана карта је дама". На основу експеримента не можемо одмах рећи да ли су догађаји A_1 и A_2 независни. Проверићемо њихову независност на основу дефиниције (3.55).

Вероватноће догађаја A_1 и A_2 су

$$P(A_1) = \frac{9}{36},\ P(A_2) = \frac{4}{36} \tag{3.57}$$

Догађај $(A_1 \cdot A_2)$ значи да је изабрана карта дама треф и његова вероватноћа је

$$P(A_1 \cdot A_2) = \frac{1}{36} \tag{3.58}$$

Из (3.57) и (3.58) се види да је задовољена релација (3.55), што значи да су то независни догађаји.

Није тешко проверити и независност следећих догађаја. Ако су A_1 и A_2 међусобно независни тада су независни и следећи догађаји

$$\overline{A_1} \text{ i } \overline{A_2};\ A_1 \text{ i } \overline{A_2};\ \overline{A_1} \text{ i } A_2$$

Дефиниција независности два догађаја се може проширити и на више догађаја.

Нека је $A_1,\ A_2,\ \ldots$ *низ случајних догађаја. Ако је за сваки изабрани низ*

$$A_{i1},\ A_{i2},\ \ldots,\ A_{in}$$

вероватноћа пресека једнака производу вероватноће, тј. ако је

$$P(A_{i1} \cdot A_{i2} \cdot \ldots \cdot A_{in}) = P(A_{i1}) \cdot P(A_{i2}) \cdot \ldots \cdot P(A_{in}) \qquad (3.59)$$

тада кажемо да је низ $A_1,\ A_2,\ \ldots$ ***низ независних догађаја***.

На пример, истовремено се бацају две коцке. Нека су дати следећи догађаји:

- A_1: на првој коцки пао је непаран број,
- A_2: на другој коцки пао је непаран број,
- A_3: збир тачкица на обе коцке је непаран број.

Лако је видети да је

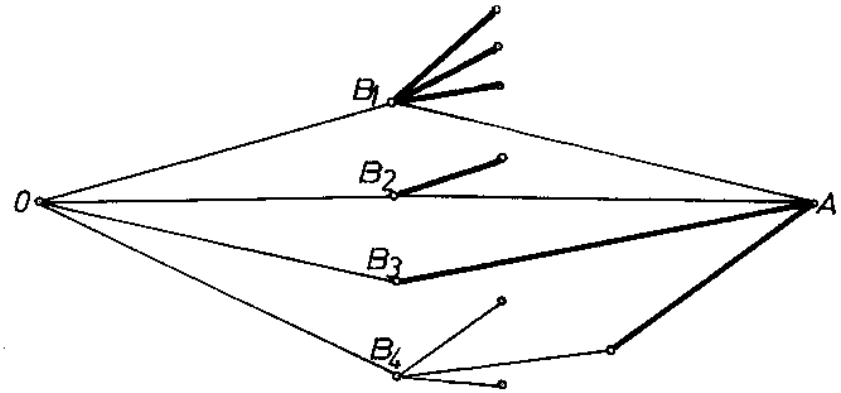
$$P(A_3 / A_1) = 0{,}5 \qquad P(A_3 / A_2) = 0{,}5 \qquad P(A_3) = 0{,}5$$

што значи да су догађаји A_1 и A_3 међусобно независни, а исто тако су независни и догађаји A_2 и A_3.

Међутим, ако су се истовремено остварили и A_1 и A_2, тада се догађај A_3 не може уопште остварити. То значи да се не може сматрати да је A_3 независан од A_1 и A_2 па се не може рећи да су догађаји A_1, A_2 и A_3 независни догађаји.

Пример 1. Путник креће из места О у место А. У сваком месту на случајан начин бира пут за даље кретање. Колика је вероватноћа да

ће путник стићи у место A без враћања назад ако су путеви дати на шеми?



Да би путник дошао у A мора проћи кроз једно од места B_1, B_2, B_3 или B_4. Зато је

$$P(B_i)=\frac{1}{4},\ i=1,2,3,4.$$

Условне вероватноће су

$$P(A/B_1)=\frac{1}{4} \qquad P(A/B_2)=\frac{1}{2}$$

$$P(A/B_3)=1 \qquad P(A/B_4)=\frac{1}{3}$$

Тражена вероватноћа је

$$P(A)=\frac{1}{4}\left(\frac{1}{4}+\frac{1}{2}+1+\frac{1}{3}\right)\frac{25}{48}$$

Пример2. У кутији се налази n_1 белих и n_2 црних куглица. Случајно се бира једна куглица, а затим се у кутију додаје m куглица оне боје које је била извучена куглица и k куглица друге боје (m и k могу бити и негативни бројеви што значи да се из кутије одузимају куглице). Одредити вероватноћу да се у другом извлачењу извуче бела

куглица. Ако је у другом извлачењу извучена црна куглица, колика је вероватноћа да је у првом извлачењу била извучена бела куглица?

Пример 3. Један стрелац гађа мету n пута. Вероватноћа поготка мете у сваком гађању је једнака p, $(0 < p < 1)$. Колика је вероватноћа да ће мета бити погођена бар једанпут? Колико треба извршити гађања да би са вероватноћом 0,999 тврдили да ће мета бити погођена бар једанпут?

Вероватноћа промашаја мете у сваком гађању је $(1-p)$, а вероватноћа промашаја у n гађања је

$$(1-p)^n$$

због независности. Тражена вероватноћа је

$$P(A) = 1-(1-p)^n \tag{3.60}$$

Да би одредили n, вероватноћу (3.60) изједначимо са 0,999. Тако добијамо да n треба да задовољава једнакост

$$(1-p)^n = 0{,}001$$

односно

$$n = \frac{\log 0{,}001}{\log(1-p)}$$

Пример 4. Наћи вероватноћу да се строј који ради у тренутку t, неће покварити до тренутка $t_0 + t$ ако:

а) вероватноћа квара зависи само од дужине временског интервала,

б) вероватноћа рада без квара за интервал Δt је пропорционална Δt са тачношћу до бесконачно мале вредности вишег реда,

в) исправност строја у интервалима који се не секу независне су међусобно.

Означимо са $p(t)$ вероватноћу квара строја од t_0 до $t_0 + t$. Тада је вероватноћа рада за време Δt једнака

$$1 - p(\Delta t) = a\Delta(t) + o(\Delta t)$$

при чему је a константа.

Рад од t_0 до $t_0 + t + \Delta t$ и вероватноћа квара строја у том интервалу је једнака

$$p(t + \Delta t)$$

Интервал $(t + \Delta t)$ се састоји од два интервала дужина t и Δt који се међусобно не секу. Зато је

$$p(t + \Delta t) = p(t)p(\Delta t)$$

тако да је

$$\frac{p(t + \Delta t) - p(t)}{\Delta t} = -ap(t) - o(1)$$

Гранична вредност овог израза, кад $\Delta t \to 0$, је једначина

$$\frac{dp(t)}{dt} = -ap(t) \tag{3.61}$$

Тражена вероватноћа ће бити решење једначине (3.61), а то је функција

$$p(t) = e^{-at}$$

Пример 5. Вероватноће за тзв. *таблице смртности*.

Претпоставке:

1. Вероватноћа да ће неко лице умрети у времену од t до $t + \Delta t$ једнака је

   $$p(t, t + \Delta t) = a(t)\Delta t + o(\Delta t) \tag{3.62}$$

   при чему је $a(t)$ ненегативна и непрекидна функција;

2. Сматра се да умирање неког лица у интервалу (t_1, t_2) не зависи од тога шта је било до момента t_1;

3. Вероватноћа смрти у моменту рађања једнака је нули.

Треба одредити вероватноћу да ће једно лице умрети кад доживи старост t. Означимо са $P(t)$ вероватноћу да ће лице доживети старост t. Са $P(t+\Delta t_1, t)$ означимо вероватноћу da će lice koje doživi starost t doživeti i starost $t+\Delta t$. Tada je

$$P(t+\Delta t) = P(t)P(t+\Delta t, t),$$

Из (3.62) добија се даје

$$P(t+\Delta t) = P(t)\{1 - p(t, t+\Delta t)\},$$

односно

$$P(t+\Delta t) = P(t)\{1 - a(t)\Delta t - o(\Delta t)\} \tag{3.63}$$

Из израза (3.63) добиће се да је

$$\frac{P(t+\Delta t) - P(t)}{\Delta t} = -a(t)P(t) - 0(1)$$

чија је гранична вредност кад $\Delta t \to 0$ једнака

$$\frac{dP(t)}{dt} = -a(t)P(t) \tag{3.64}$$

Решење диференцијалне једначине (3.64) је функција

$$P(t) = e^{-\int_0^t a(z)dz}$$

на основу које се добијају тражене вредности. Специјално, ако је $a(t) = \alpha + \beta e^{\gamma t}$ тада је тражена вероватноћа једнака

$$P(t) = e^{-\alpha t} - \frac{\beta}{\gamma}\left(e^{\gamma t} - 1\right)$$

4. ЗАКОН ВЕЛИКИХ БРОЈЕВА

Аксиоматска дефиниција вероватноће не одређује начин на који ће вероватноће случајних догађаја бити одређене у неком реалном експерименту. Зато треба наћи одговор на питање да ли се и под којим условима, на основу резултата извођења експеримента, може одредити одговарајућа вероватноћа посматраног случајног догађаја. Одговор на ово питање, важно у теорији вероватноће и статистици, дају тзв. закони великих бројева, о којима ће бити речи у поглављу које следи.

Најпре ће бити детаљније описани Бернулијева шема независних понављања експеримената и Гаусова крива, без којих развој теорије вероватноће и статистике не би био могућ.

У трећем делу поглавља ће бити изложен Бернулијев закон великих бројева и дата формулација Борел-Кантелијевог закона.

Локална и интегрална теорема Моавр-Лапласа, без доказа, су дате у четвртом делу поглавља. На основу ових теорема могуће је приближно одредити биномне вероватноће у случају када су вредности броја n велике, а Пуасоново решење за ретке догађаје је дато на крају поглавља.

4.1. БЕРНУЛИЈЕВА ШЕМА НЕЗАВИСНИХ ПОНАВЉАЊА ЕКСПЕРИМЕНАТА

Изузетно важну улогу у теорији вероватноће, њеној примени и развоју статистике, одиграла је једна класа експеримената познатих под именом ***Бернулијеви* (Bernoulli) *експерименти*** или класа независних понављања експеримента. Одређивање вероватноћа случајних догађаја везаних за Бернулијеве експерименте представља тзв. **Бернулијев проблем**.

Претпоставимо да је један експеримент могуће поновити неограничено пута. При томе се мисли да је могуће остварити неограничено пута један исти комплекс услова. Сваки пут кад се оствари тај комплекс услова, посматрани случајни догађај A се може остварити или не. Резултат експеримента, при томе, у једном понављању не утиче на резултате експеримента у осталим понављањима. Рећи ћемо да су понављања експеримента **међусобно независна**. Поред тога, предпоставља се да је у сваком понављању експеримента **вероватноћа остварења догађаја A иста**.

Наведимо неколико примера таквих експеримената.

Пример 1. Правилни новчић се баца неограничено пута и бележи се број појављивања „писма". Може се посматрати да ли постоји гранична вредност за релативну фреквенцију појављивања „писма" кад неограничено расте број бацања новчића (понављања експеримента) и, уколико постоји, колико износи.

Пример 2. Машина аутоматски обрађује производ са вероватноћом погрешне обраде једнаком p. Ако се у великој серији од 10000 или 100000 комада појави велики број неправилно обрађених комада, да ли то значи да се поузданост машине променила?

Пример 3. Зна се да је међу свим порођајима проценат близанаца једнак p. Ако се у једној клиници у коју долазе породиље изабране на случајан начин нађе 1000 породиља, колики се може очекивати број близанаца?

Пример 4. У статистичком скупу од N елемената, $100\,p\%$ има одређено својство, док преосталих $100(1-p)\%$ нема то својство. На случајан начин се из посматраног скупа бира један елемент и бележи се да ли он има посматрано својство или не. При томе се претпоставља да се извлачење врши са враћањем, тј. сваки пут се извучени елемент враћа у посматрани скуп, па се поново из целог скупа бира следећи елемент. Може нас занимати да одредимо вероватноћу да ће међу n извучених елемената бити тачно k са посматраним својством. Може нас, на пример, интересовати и вероватноћа догађаја да се количник k/n (релативна фреквенција појављивања елемента са посматраним

својством) разликује од p за не више од једног унапред датог броја. Специјално нас може интересовати чему ће тежити количник *k/n* када број извлачења n тежи бесконачности.

У свим наведеним примерима описани су експерименти у којима се посматра случајни догађај А, који се у једном извођењу експеримента остварује са вероватноћом p и не остварује са вероватноћом $q = 1 - p$. Низ независних понављања таквог експеримента представља тзв. Бернулијев експеримент.

Означимо са n број понављања посматраног експеримента. Најједноставнији и најинтересантнији проблем је одређивање вероватноће да ће се у n понављања експеримента посматрани догађај A реализовати тачно k пута. То значи да се у преосталих $(n-k)$ понављања догађај A неће реализовати. При томе је k један одређени број из скупа бројева {0, l, 2, ..., n}. Означићемо ту вероватноћу са $P_n(k)$.

Пре него пређемо на одређивање вероватноће $P_n(k)$ у општем случају, посматрајмо следећи пример:

Претпоставимо да се у једној кутији налази 5 сијалица и да је вероватноћа да је сијалица исправна једнака за све сијалице и износи 0,95. Одредимо вероватноћу да се у кутији налазе тачно три исправне и две неисправне сијалице.

Случајни догађај А који посматрамо у појединачном извођењу „експеримента" је догађај: „сијалица је исправна", а супротан догађај је $\overline{A}$: „сијалица је неисправна". При томе је $p = 0.95$, $q = 0.05$, а број понављања експеримента је једнак броју сијалица, тј. $n = 5$. Тражена вероватноћа је $P_5(3)$.

Ако за сваку сијалицу придружимо А или $\overline{A}$, онда ће се догађај да међу пет сијалица буду три исправне реализовати ако се реализовала једна од могућности писања низа од пет словних места у којима је на 3 места слово А, а на два места слово $\overline{A}$. Једна од могућности је, на пример, низ

$$AA\,\overline{A}\,A\,\overline{A}$$

Вероватноћа да се оствари овај низ је једнака

$$0{,}95 \cdot 0{,}95 \cdot 0{,}05 \cdot 0{,}95 \cdot 0{,}05 = 0{,}95^3 \cdot 0{,}05^2$$

јер су догађаји у овом низу независни. Оваквих могућих низова има укупно 10. То су

1. $AAA\,\overline{A}\,\overline{A}$
2. $AA\,\overline{A}\,A\,\overline{A}$
3. $A\,\overline{A}\,AA\,\overline{A}$
4. $\overline{A}\,AAA\,\overline{A}$
5. $AA\,\overline{A}\,\overline{A}\,A$
6. $A\,\overline{A}\,A\,\overline{A}\,A$
7. $\overline{A}\,AA\,\overline{A}\,A$
8. $A\,\overline{A}\,\overline{A}\,AA$
9. $\overline{A}\,A\,\overline{A}\,AA$
10. $\overline{A}\,\overline{A}\,AAA$

Сваки од ових низова има исту вероватноћу и она је једнака

$$0{,}95^3 \cdot 0{,}05^2$$

Наведени низови се међусобно искључују јер се у једном пакету од 5 сијалица може реализовати тачно један од њих.

Дефинисани догађај да се међу 5 изабраних сијалица налази тачно 3 неисправне, према томе, реализује се кад се реализовао један од наведених низова. Зато је његова вероватноћа једнака збиру вероватноћа сваког од њих, тј.

$$P_5(3) = 10 \cdot 0{,}95^3 \cdot 0{,}05^3$$

На сличан начин се могу одредити и вероватноће

$$P_5(0), P_5(\mathrm{l}), P_5(2), P_5(4), P_5(5).$$

Посматрајмо, сада, општи случај. Ако се n пута понавља експеримент, одредимо вероватноћа да ће се догађај A реализовати у k понављања, а у преосталих $(n-k)$ неће.

Сваком извођењу експеримента приписаћемо слово A или $\overline{A}$, зависно од тога да ли се догађај A реализовао или није. На тај начин

ћемо, низу од n понављања експеримента придружити низ од n словних места, који је састављен од слова A и $\overline{A}$. Посматрани догађај ће се реализовати ако се реализује низ у коме ће на k места бити слово A, а на преосталих $(n-k)$ места $\overline{A}$.

Вероватноћа сваког таквог низа је једнака:

$$p^k q^{n-k}$$

због независности догађаја.

Са друге стране, сви ти низови се међусобно искључују, јер једном Бернулијевом експерименту од n понављања одговара тачно један од тих могућих низова у којима се посматрани догађај A реализовао k пута. Таквих различитих низова има онолико колико има могућности да се од n словних места изабере k места на која се ставља слово A, (на преостала места се ставља $\overline{A}$), а тај број је једнак:

$$C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Низ вероватноћа:

$$P_n(0),\ P_n(1),\ldots,\ P_n(n)$$

где је:

$$P_n(k) = \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

је низ биномних вероватноћа *и представља вероватноће да се у Бернулијевом експерименту догађај A реализује тачно k пута, при чему је* $k=0,1,2,\ldots,n$.

Биномне вероватноће $P_n(k)$ имају мале вредности за велики број n. Неке вредности вероватноће за $n=50$ и $p=\frac{1}{3}$ су дате у Табели 4.1. Највећа вероватноћа једнака је *0,1178* (За $n=50$ и $p=0{,}5$, највећа

вероватноћа је мања од *0,08*). Из Табеле 4.1. види се да биномне вероватноће расту до извесне вредности, а затим опадају.

k	$P_n(k)$	k	$P_n(k)$
5	0,0001	17	0,1178
8	0,0033	18	0,1080
10	0,0157	19	0,0910
12	0,0470	20	0,0704
14	0,0879	22	0,0332
15	0,1077	25	0,0059
16	0,1178	30	0,0001

Табела 4.1. Биномне вероватноће за $n = 50$ *и* $p = \frac{1}{3}$.

Значи да ће се за неку вредност k добити највећа вредност биномних вероватноћа. Интересује нас која је то вредност.

Биномне вероватноће за k и $(k+1)$ су:

$$P_n(k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

$$P_n(k+1) = \frac{n!}{(k+1)!(n-k-1)!} p^{k+1} q^{n-k-1}$$

Њихов количник је број

$$\frac{P(k+1)}{P(k)} = \frac{n-k}{k+1} \frac{p}{q} \tag{4.1}$$

Ако је овај количник већи од један, значиће да са порастом k бинома вероватноћа расте. То ће бити случај кад је

$$\frac{n-k}{k+1}\frac{p}{q}\geq 1$$

односно кад је

$$k\leq np-q$$

За

$$k>np-q$$

количник (4.1) је мањи од један, што значи да ће за $k>np-q$ биномне вероватноће опадати.

Ако је $(np-q)$ цео број, количник биномних вероватноћа за $(np-q+1)$ и $(np-q)$ биће једнак

$$\frac{P_n(np-q+1)}{P_n(np-q)}=\frac{n-(np-q)}{np-q+1}\frac{p}{q}$$

Односно

$$\frac{P_n(np-q+1)}{P_n(np-q)}=1$$

На основу изложеног се може закључити:

Ако $np-q$ није цео број, онда је број понављања у којима се реализује случајан догађај A са највећом вероватноћом, једнак:

$$k=[np-q]+1$$

при чему је $[\ ]$ ознака за највећи цео број који није већи од $np-q$.

Ако је $np-q$ цео број, онда постоје две вредности k које су највероватније, а то су вредности:

$$k=np-q \qquad k+1=np-q+1$$

За те вредности, биномне вероватноће су међусобно једнаке, и веће су од свих осталих вероватноћа.

4.2. ГАУСОВА КРИВА

У примени теорије вероватноће велику улогу има функција позната под именом ***Гаусова крива (Gauss)*** *или* ***Гаус- Лапласова крива (Laplace)*** која се често зове и ***закон вероватноћа Нормалне расподеле***.

Гаусова крива је крива дата функцијом

$$\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \tag{4.2}$$

која је дефинисана за свако реално x из интервала $(-\infty,+\infty)$.

Функција $\phi(x)$ је симетрична функција која има максимум у тачки $x=0$. Максимална вредност ове функције је

$$\phi(0)=\frac{1}{\sqrt{2\pi}}\approx 0{,}3989$$

Превојне тачке функције $\Phi(x)$ су тачке $x=\pm 1$ у којима функција има вредност

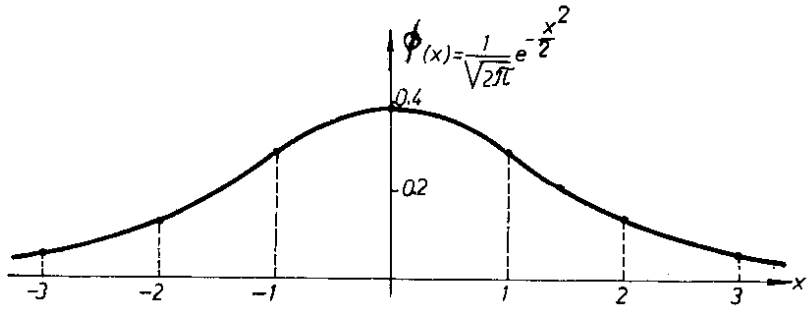
$$\phi(\pm 1)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}}\approx 0{,}2420$$

Поред тога је

$$\lim_{x\to+\infty}\phi(x)=0$$

$$\lim_{x\to-\infty}\phi(x)=0$$

Гаусова крива је приказана на слици 4.1.



Слика 4.1. Гаусова крива $\phi(x)$

Површина између криве $\phi(x)$ и осе x је једнака интегралу

$$\int_{-\infty}^{+\infty}\phi(x)dx$$

и има вредност једнаку јединици, тј.

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}dx=1 \qquad (4.3)$$

Вредности функције $\phi(x)$ дају се табеларно, најчешће за вредности *x* из интервала [0; 3,99]. За *x = 3,99* функција има вредност *0,0001*, а за вредности *x* веће од *3,99* функција $\phi(x)$ има мале вредности које су практично једнаке 0.

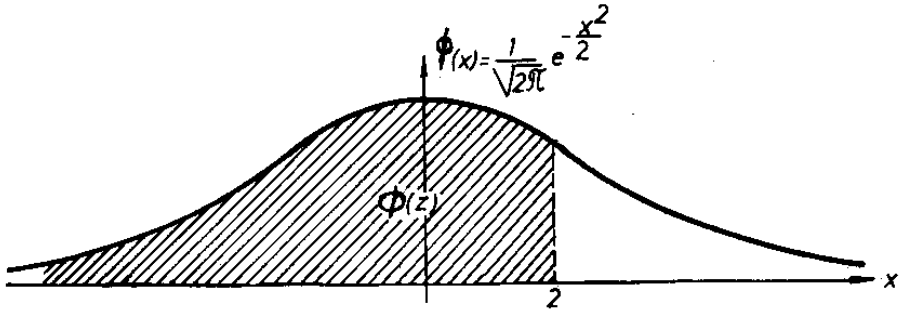
На основу Гаусове криве, дефинисана је тзв. **функција Нормалне расподеле** $\Phi(z)$ чије су вредности дате табеларно и која служи за одређивање површине између криве $\phi(x)$ и осе x у неком произвољном интервалу [a; b].

<u>Дефиниција 4.1.</u> Функција Нормалне расподеле $\Phi(z)$ *дефинисана је за свако* $z \in (-\infty, +\infty)$ *једначином*

$$\Phi(z) = \int_{-\infty}^{z} \phi(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx$$

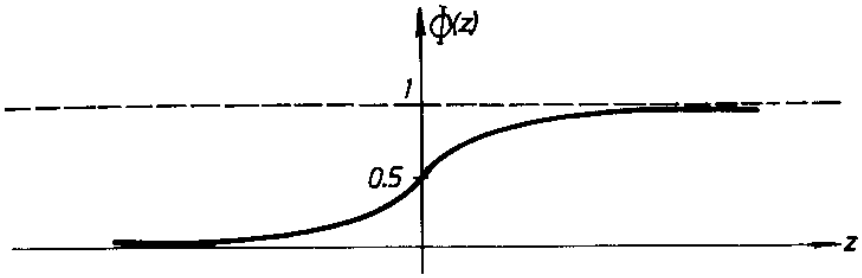
и представља површину између криве $\phi(x)$ *и осе x у интервалу* $(-\infty, z]$.

Графички приказ преко површине функције $\phi(x)$ дат је на слици 4.2.



Слика 4.2. Функција Нормалне расподеле приказана као површина

Функција Нормалне расподеле $\Phi(z)$ је монотоно растућа функција која има хоризонталне асимптоте *y=0,* када $x \to -\infty$, и *y = 1,* када $x \to \infty$. Графички је приказана на слици 4.3.



Слика 4.3. Функција $\Phi(z)$

Вредности функције нормалне расподеле $\Phi(z)$ дате су табеларно за позитивне вредности z. Због симетрије Гаусове криве, важи:

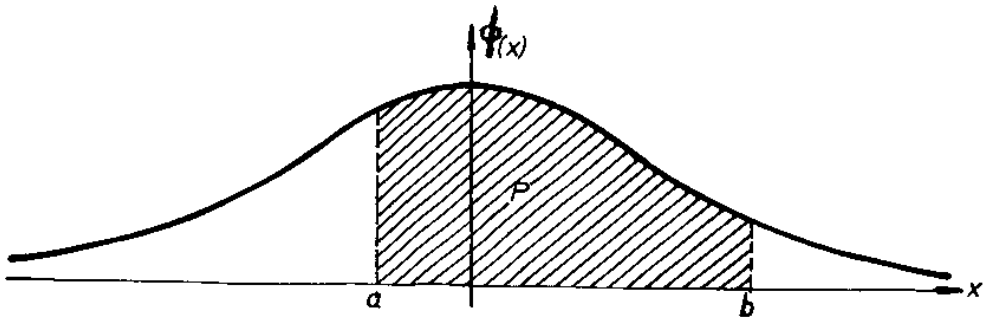
$$\Phi(-z)=1-\Phi(z)$$

што није тешко проверити, на основу дефиниције (4.1).

У табелама функције Нормалне расподеле најчешће су дате вредности функције за z из интервала [0; 3,49]. За вредности веће од 3,49 функција Ф практично има вредност 1.

Да би одредили ***површину између Гаусове криве и осе x*** *у произвољном интервалу* $(a, b]$ *користимо функцију* $\Phi(z)$. *Важи:*

$$P=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx=\Phi(b)-\Phi(a)$$



Слика 4.4. Приказ површине и функције $\phi(z)$

Најчешће се посматра симетрични интервал $[-z,\ z]$. Површина између функције $\phi(x)$ и осе x, у том интервалу једнака је

$$P = \frac{1}{\sqrt{2\pi}} \int_{-z}^{z} e^{-\frac{x^2}{2}} dx = 2\Phi(z) - 1.$$

Тако се добија да је површина у интервалу $(-1,\ 1]$ једнака:

$$P = 0{,}6826 = 68{,}3\%$$

а у интервалу $(-2,\ 2]$:

$$P = 0{,}9544 = 95{,}4\%$$

док је за интервал $(-3,\ 3]$

$$P = 0{,}9974 = 99{,}7\%$$

што значи да се 99,7% од укупне површине између Гаусове криве $\phi(x)$ и осе x, налази у интервалу (-3; 3] док је изван тог интервала свега 0,03% те површине.

4.3. БЕРНУЛИЈЕВ И БОРЕЛ-КАНТЕЛИЈЕВ ЗАКОН ВЕЛИКИХ БРОЈЕВА

Посматраћемо Бернулијеву шему независних понављања експеримента са случајним догађајем A и вероватноћом p реализације догађаја A у сваком понављању.

Посматрајмо следећи проблем: ***Ако се n пута понавља експеримент, да ли је између релативне фреквенције остварења посматраног догађаја и вероватноће p може уочити нека зависност?***

Одговор на постављено питање дао је J. Bernoulli (1654-1705. године) у познатом делу „*Ars Conjectandi*" објављеном 1713. године. Бернулијеве резултате уопштио је Моавр (Abraham De Moivre) 1718. године, а затим је Лаплас (Pierre S. Laplace) 1812. године дао најопштију формулацију одговора на постављено питање.

Ради бољег разумевања правог значења ових резултата, размотримо прво Бернулијеву формулацију.

Означимо са X број понављања експеримента у којима се реализовао догађај A. Ако се експеримент понови n пута X ће узети једну вредност из низа могућих вредности

$$\{0,1,2,...,n\}$$

Догађај $D=\{X=k\}$ је случајни догађај који се у Бернулијевом експерименту реализује са вероватноћом

$$P_n(k)=P\{X=k\}=\frac{n!}{k!(n-k)!}p^k q^{n-k} \tag{4.4}$$

Количник

$$\frac{X}{n} \tag{4.5}$$

је **фреквенција остварења догађаја А** *у* ***n*** *пута поновљеном експерименту.*

Фреквенција остварења догађаја А може узети једну вредност из низа могућих вредности

$$\{0, \frac{1}{n}, \frac{2}{n}, ..., 1\}$$

Догађај $D`=\left\{\frac{X}{n}=\frac{k}{n}\right\}$ је случајни догађај који се у Бернулијевом експерименту реализује са вероватноћом датом изразом (4.4).

Интересује нас разлика између фреквенције (4.5) и вероватноће p. Посматрајмо апсолутну вредност те разлике.

$$\left|\frac{X}{n}-p\right| \tag{4.6}$$

Ако експеримент поновимо n пута, разлика (4.6) ће узети неку вредност из интервала [0; 1]. Шта се дешава са разликом када се повећава број понављања експеримента?

Изаберимо произвољно мали позитиван број ε и посматрајмо догађај да ће разлика (4.6) бити већа од тога броја, тј. посматрајмо догађај

$$D=\left\{\left|\frac{X}{n}-p\right|>\varepsilon\right\} \tag{4.7}$$

и одредимо вероватноћу остварења овог догађаја у n понављања експеримента.

Догађај D ће се остварити кад је

$$X > n(p+\varepsilon) \tag{4.8}$$

или кад је

$$X < n(p-\varepsilon) \tag{4.9}$$

(Види слику 4.1.).



Слика 4.1.

Вероватноћа догађаја датог изразом (4.8) је једнака збиру биномних вероватноћа

$$P\{X > n(p+\varepsilon)\} = \sum_{i=r}^{n} P_n(i) \tag{4.10}$$

при чему је r најмањи цео број који је већи од производа $n(p+\varepsilon)$.

Могуће је показати да је посматрана вероватноћа ограничена са горње стране тако да је

$$P\{X > n(p+\varepsilon)\} < P_n(r)\frac{n(p+\varepsilon)+q}{n\varepsilon+q} \tag{4.11}$$

Када број понављања експеримента *n* неограничено расте, израз на десној страни (4.8) остаје ограничен. Из дефиниције биномних вероватноћа се закључује да је:

$$\lim_{n\to\infty} P_n(r) = 0$$

Следи да је

$$\lim_{n\to\infty} P\{X > n(p+\varepsilon)\} = 0 \tag{4.12}$$

На сличан начин се може утврдити да је гранична вредност вероватноће догађаја датог изразом (4.9) такође једнака нули, тј. да је

$$\lim_{n\to\infty} P\{X < n(p-\varepsilon)\} = 0 \tag{4.13}$$

Из наведеног следи да је:

$$\lim_{n\to\infty} P\left\{\left|\frac{X}{n} - p\right| > \varepsilon\right\} = 0$$

одакле следи:

Теорема 4.2. ***(Бернулијева теорема).*** *Вероватноћа да ће фреквенција остварења догађаја A одступити од његове вероватноће p за вредност мању од произвољно малог броја ε тежи јединици кад број понављања експеримента n неограничено расте.тј.*

$$\lim_{n\to\infty} P\{|\frac{X}{n} - p| \le \varepsilon\} = 1 \tag{4.14}$$

Теорема 4.2. представља **Бернулијев закон великих бројева**.

Ако је број понављања n довољно велики, теорема тврди да је скоро сигурно да ће разлика релативне фреквенције остварења догађаја A и његове вероватноће бити произвољно мала. То значи да се на основу релативне фреквенције може скоро са сигурношћу утврдити вероватноћа неког случајног догађаја.

Тек када је доказан овај закон, било је могуће прихватити дефиницију емпиријске вероватноће и на тај начин, експерименталним путем, одређивати вероватноће случајних догађаја.

Овај закон одиграо је пресудну улогу у развоју и примени теорије вероватноће, односно у развоју статистике, јер је тек његовим доказивањем постало јасно да постоје законитости код појава и процеса које нису детерминистичког карактера, тј. код појава и процеса стохастичког карактера.

Са развојем теорије вероватноће утврђене су и општије формулације закона великих бројева. Са становишта одређивања самог појма вероватноће и њене суштине, интересантно је уопштење које су дали **Борел и Кантели** и које је познато под називом „***јаки закон великих бројева***”, а гласи:

Ако је X број експеримента у којима се случајни догађај A остварио, вероватноћа да ће гранична вредност фреквенција остварења догађаја A, када се број понављања експеримената неограничено повећава, бити једнака вероватноћи p је једнака јединици, тј.

$$P\{\lim_{n\to\infty}\frac{X}{n}=p\}=1 \tag{4.15}$$

Закони дати изразима (4.14) и (4.15) потпуно описују фундаментално својство случајности које је могуће проучавати преко вероватноће.

4.4. МОАВР-ЛАПЛАСОВО РЕШЕЊЕ БЕРНУЛИЈЕВОГ ПРОБЛЕМА

У разматрањима Бернулијевог експеримента и закона великих бројева остао је нерешен проблема израчунавања биномних вероватноћа за велике вредности *n* и *k*. Било би пожељно да се утврди начин приближног рачунања ових вероватноћа, али тако да буде обезбеђена жељена прецизност.

Решења за овај проблем први је дао Моавр за случај када је $p = 0{,}5$, а затим је Лаплас то решење уопштио и за произвољне вредности *p*. Та решења дата су у Теоремама 4.3 и 4.4.

***Теорема 4.3. (Локална теорема Моавр-Лапласа)**: Нека је вероватноћа остварења неког догађаја A у сваком од n независних понављања једног експеримента једнака p, (0 < p < 1), и нека је X број понављања експеримента у којима се реализовао догађај A. Означимо са*

$$P\{X = k\}$$

вероватноћу да ће се догађај A реализовати k пута.

Тада је низ

$$\sqrt{npq}P\{X = k\} = \sqrt{npq}P_n(k)$$

униформно конвергентан за свако k за које се

$$\frac{k - np}{\sqrt{npq}} \tag{4.16}$$

налази у неком коначном интервалу. При томе је

$$\lim_{n\to\infty}\sqrt{npq}P_n(k)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(k-np)^2}{npq}}=\phi\left(\frac{k-np}{\sqrt{npq}}\right) \tag{4.17}$$

где је функција $\phi(x)$ *Гаусова крива.*

Из теореме 4.3. непосредно следи:

Биномне вероватноће $P_n(k)$ за довољно велико n имају приближне вредности

$$P_n(k)\approx\frac{1}{\sqrt{npq}}\phi(x) \tag{4.18}$$

при чему је

$$x=\frac{k-np}{\sqrt{npq}} \tag{4.19}$$

А функција $\phi(x)$ ја Гаусова крива

$$\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \tag{4.20}$$

односно закон вероватноћа Нормалне расподеле.

Приближне вредности биномних вероватноћа дате преко Гаусове криве изразом (4.18) су боље што је *n* веће. Међутим, чак и за релативно мале вредности n приближне вредности су довољно добра апроксимација.

Теорема 4.4. (Интегрална теорема Моавр-Лапласа). *Нека је X број понављања експеримента у којима се реализовао догађај A, а p вероватноћа његовог остварења у сваком понављању, (0 < p < 1). Тада ће вероватноћа да се вредност*

$$\frac{X-np}{\sqrt{npq}} \tag{4.21}$$

нађе у интервалу (a,b] тежити интегралу

$$\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx=\int_a^b \phi(x)dx$$

тј. тада је

$$\lim_{n\to\infty} P\left\{a<\frac{X-np}{\sqrt{npq}}\le b\right\}=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx=\int_a^b \phi(x)dx \qquad (4.22)$$

Интегрална теорема Моавр-Лапласа има веома широку примену у одређивању вероватноћа различитих догађаја везаних за Бернулијеву шему независних понављања једног експеримента.

На основу граничне вредности дате у теореми 4.4 можемо приближно одређивати збирове биномних вероватноћа. Наиме, из (4.22) следи да је за довољно велико n

$$P\left\{a<\frac{X-np}{\sqrt{npq}}\le b\right\}\approx\int_a^b \phi(x)dx \qquad (4.23)$$

Пошто је вероватноћа на левој страни (4.23) једнака збиру одговарајућих биномних вероватноћа, то значи да је

$$\sum_k P_n(k)\approx\int_a^b \phi(x)dx \qquad (4.24)$$

при чему збир иде по k од $(np-a\sqrt{npq})$ до $(np+b\sqrt{npq})$.

Интеграл на десној страни (4.23) и (4.24) може се одредити преко таблица за функцију Нормалне расподеле $\Phi(z)$. Вероватноћа у (4.23) је приближно једнака

$$P\left\{a < \frac{X - np}{\sqrt{npq}} \le b\right\} \approx \Phi(b) - \Phi(a) \tag{4.25}$$

тј.

$$\sum_k P_n(k) \approx \Phi(b) - \Phi(a)$$

4.5. ПУАСОНОВО РЕШЕЊЕ ЗА РЕТКЕ ДОГАЂАЈЕ

Одређивање биномних вероватноћа $P_n(k)$ у Бернулијевој шеми независних понављања експеримента није сасвим једноставно за велике бројеве n. Моавр-Лапласова теорема омогућава да се приближно одреде посматране вероватноће за случај када np и npq расту и кад $n \to \infty$.

Међутим, постоје случајеви када је Моавр-Лапласова апроксимација незадовољавајућа. За те случајеве користи се једина апроксимација која је математички много једноставнија од претходне.

Пуасон је дао формулу која представља веома добру апроксимацију биномних вероватноћа за ретке догађаје, тј. за догађаје чија вероватноћа реализације у једном извођењу је мала.

Означимо са λ производ

$$\lambda = np,$$

при чему је n број понављања експеримента, а p вероватноћа реализације догађаја A у сваком понављању експеримента. Тада се биномне вероватноће

$$P_n(k)$$

могу написати у облику:

$$\lim_{n \to +\infty} P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda} \tag{4.26}$$

Израз (4.26) омогућава да се изврши апроксимација биномних вероватноћа.

Ако је n довољно велико, а p мало, тако да је вредност производа np мала ($np<10$), тада се биномне вероватноће могу приближно одређивати по формули:

$$P_n(k)=\frac{\lambda^k}{k!}e^{-\lambda} \tag{4.27}$$

при чему је $\lambda=np$.

5. ЈЕДНОДИМЕНЗИОНАЛНЕ СЛУЧАЈНЕ ПРОМЕНЉИВЕ

Проучавање појава у природи или друштву, базирано је на успостављању и откривању релација између догађаја везаних за дате појаве. Потпуније изучавање доводи до три категорије догађаја:

- немогући;
- сигурни и
- случајни догађаји.

За описивање појава или експеримената који генеришу случајне догађаје користе се ***пробабилистички модели*** засновани на ***Теорији вероватноће.***

Најширу, и за праксу најважнију класу пробабилистичких модела, чини класа модела базирана на *случајним променљивим*. ***Случајне променљиве проучавају оне експерименте*** чији резултати се могу изразити нумеричким вредностима: У сваком „извођењу" експеримента „нешто" се пребројава или мери. Резултат експеримента је број из коначног, бесконачног и пребројивог или бесконачно-непребројивог скупа реалних бројева. Другим речима, резултат експеримента је тачка на реалној оси. Такву класу случајних променљивих представљају *једнодимензионалне случајне променљиве*.

Ако се резултат експеримента може изразити са више нумеричких карактеристика, онда су резултати експеримента тачке у вишедимензионалном простору чија димензија одговара броју мерених нумеричких карактеристика. Такви резултати се називају *вишедимензионалне случајне променљиве*, а пробабилистички модели који проучавају те експерименте, засновани су на мултидимензионалним случајним променљивим

Потребно је упознати се са основним елементима теорије вероватноће једнодимензионалних и вишедимензионалних случајних променљивих. То су они елементи ове теорије, који су *неопходни за*

примену у ***статистичким закључивањима*** - а то је главни циљ статистичких проучавања природних и друштвених промена.

5.1. РАСПОДЕЛА СЛУЧАЈНИХ ПРОМЕНЉИВИХ

Кад се баца коцка на равну површину, онда је скуп могућих резултата скуп бројева

$$\{1,2,3,4,5,6\} \tag{5.1}$$

Ако са X означимо резултат експеримента, онда кажемо да променљива величина X може узети једну вредност из скупа реалних бројева (5.1), и то на ***случајан начин.***

Нека је дат статистички скуп од N елемената и нека обележје X има расподелу дату табелом

X	x_1	x_2	...	x_n
f	f_1	f_2	...	f_n

при чему су бројеви

$$\{x_1, x_2, \ldots, x_n\}$$

могуће вредности обележја X на елементима статистичког скупа, а бројеви

$$\{f_1, f_2, \ldots, f_n\}$$

су фреквенције појављивања појединих вредности X у посматраном скупу. Ако на случајан начин бирамо један елемент из посматраног скупа и утврђујемо вредност обележја X код тог елемента, тада је променљива величина X случајна променљива која узима једну вредност из низа могућих *вредности статистичког скупа и то на случајан начин.*

На телефонској централи се посматра број позива у временском интервалу. Променљива величина, која представља број позива, може узети једну од вредности из низа реалних бројева.

$$\{0, 1, 2, \ldots\} \tag{5.2}$$

и то на случајан начин.

При обради једног производа, на тачност неке димензије утиче низ фактора. Зна се да ће димензија X бити у неким утврђеним границама, рецимо у интервалу

$$(x_1; x_2) \tag{5.3}$$

Код сваког обрађеног производа, димензија X ће *на случајан начин* узети једну вредност из интервала (5.3).

Посматра се век трајања једног уређаја. Резултат може бити било који реалан број из интервала

$$(0, \infty) \tag{5.4}$$

Сваки случајни догађај, везан за овакве појаве, може се изразити преко променљиве X. Тако се, на пример, у бацању коцке, случајни догађај A: „Пао је број не мањи од 3”, може изразити са

$$A = \{X \geq 3\} = \{3,4,5,6\}$$

Вероватноћа догађаја A може се одредити лако ако се зна вероватноћа са којом X узима поједине вредности из скупа могућих вредности $\{1,2,3,4,5,6\}$.

Ако је телефонска централа ограниченог капацитета, рецимо највише 100 позива у датом интервалу, онда се случајни догађај A: „Телефонска централа неће бити заузета”, може изразити преко променљиве X са

$$A = \{0 \leq X < 100\}$$

при чему је X број позива у посматраном временском интервалу. Ако знамо вероватноћу појединих вредности променљиве X, лако ћемо одредити вероватноћу случајног догађаја A.

У сваком експерименту, резултатима експеримента придружујемо промењљиву величину X која у једном извођењу експеримента може узети само једну вредност из датог скупа реалних бројева. Унапред не знамо коју ће вредност узети X. Једино што можемо утврдити је скуп могућих вредности за ту промењљиву.

У даљим разматрањима оваквих промењљивих величина, први задатак је утврдити *како често*, односно *са којом вероватноћом* ће промењљива X узимати поједине вредности из датог скупа могућих вредности.

У пракси се посматрају две категорије случајних промењљивих и то *прекидног* и *непрекидног* типа. Посматраћемо прво случајне величине прекидног типа и на основу њихове дефиниције и основних особина утврдити потребне елементе за дефиницију и изучавање непрекидних случајних промењљивих.

Закон вероватноћа прекидне случајне промењљиве

Дефиниција 5.1: *Промењљива величина* **X је случајна промењљива прекидног типа**, *ако вредности, које она може узети, образују коначан или пребројив низ реалних бројева*

$$\{x_1, x_2, x_3, \ldots\}, \tag{5.5}$$

а узимање сваке од ових вредности је случајан догађај са одређеном вероватноћом.

Посматрајмо низ случајних догађаја

$$D_1=(X=x_1),\ D_2=(X=x_2),\ D_3=(X=x_3),\ \ldots \tag{5.6}$$

и означимо са

$$p_1,\ p_2,\ p_3,\ \ldots \tag{5.7}$$

одговарајуће вероватноће догађаја $D_1, D_2, \ldots$, при чему је p_i вероватноћа да ће X узети вредност x_i, тј.

$$p_i = P(X = x_i),$$

за i=1,2, ...

Случајни догађаји (5.6) се међусобно искључују. Њихова унија је сигуран догађај, јер се у једном извођењу експеримента реализује тачно један од њих. Зато низ одговарајућих вероватноћа (5.7) представља низ ненегативних бројева чији је збир једнак јединици, тј. низ је такав да је

$$\sum_{i=1}^{\infty} p_i = 1 \tag{5.8}$$

У случају када је низ могућих вредности случајне променљиве (5.5) коначан низ од *n* бројева, збир у (5.8) биће збир од 1 до *n*, тако да је

$$\sum_{i=1}^{n} p_i = 1 \tag{5.9}$$

Дефиниција 5.2: *Скуп могућих вредности случајне променљиве X и одговарајућих вероватноћа, називамо закон вероватноћа случајне променљиве X и најчешће га приказујемо у виду табеле*

X	x_1	x_2	x_3	...
$P(x)$	p_1	p_2	p_3	...

при чему је

$$\forall i \in \{1,2,3,\cdots\} \Rightarrow p_i = P(X = x_i).$$

Кажемо да је законом вероватноћа одређена расподела случајне променљиве.

Расподела случајне променљиве *X* садржи максималну количину информација о посматраном експерименту, односно, о посматраној појави или процесу. Поред тога, ова расподела омогућава комплетну анализу посматране појаве. Другим речима, у расподели

случајне величине X садржана је довољна количина информација о посматраној појави.

На пример, бацају се истовремено два новчића и региструје број појављивања „писма”. Ако означимо са X резултат експеримента, онда ће X бити случајна променљива са могућим вредностима

$$\{0,1,2\}$$

и одговарајућим вероватноћама

$$\left\{\frac{1}{4},\frac{2}{4},\frac{1}{4}\right\}$$

Закон вероватноћа случајне променљиве X је дат табелом:

X	0	1	2
$p(x)$	0.25	0.50	0.25

Ако се, на пример, гађа неки циљ све до момента његовог погађања, може се посматрати број извршених гађања X. Нека је, у сваком гађању, вероватноћа поготка циља једнака p, при чему је $0<p<1$. Тада је, због независности, вероватноћа да број гађања X буде једнак, рецимо k, једнака

$$P(X=k)=(1-p)^{k-1}p$$

јер ће се догађај

$$D=(X=k)$$

реализовати једино ако је у $(k-1)$ гађања циљ промашен, а погођен у k-том гађању. Закон вероватноћа случајне променљиве X је дат следећом табелом:

X	1	2	3	...
$p(x)$	p	$(1-p)p$	$(1-p)^2p$	...

Посматрајмо статистички скуп од N елемената и расподелу обележја X на том скупу, дату табелом

X	x_1	x_2	...	x_n
F	f_1	f_2	...	f_n

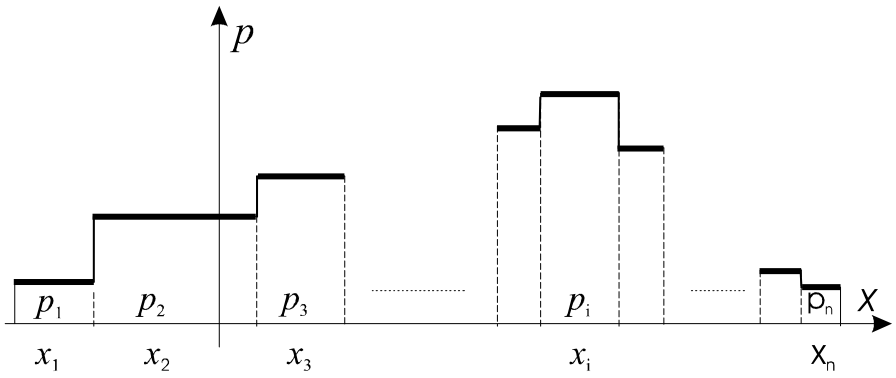
Ако на случајан начин из посматраног скупа бирамо један елеменат, тад је вредност обележја X на изабраном елементу случајна променљива са расподелом

X	x_1	x_2	...	x_n
$p(x)$	f_1/N	f_2/N	...	f_n/N

Ова расподела није ништа друго него расподела релативних фреквенција обележја X на посматраном статистичком скупу. Зато се сваки закључак који изведемо за случајну променљиву X односи и на расподелу обележја X на статистичком скупу.

Са друге стране, у изучавањима статистичких скупова и појава масовног карактера, служимо се одређеном методологијом и одређеним начином закључивања, и то у складу са реалним појавама које нас окружују. Зато ћемо случајне променљиве проучавати на исти такав начин, тј. у складу са захтевима који проистичу из реалности, а то ће омогућити широку примену изведене теорије. Истовремено резултати опште теорије потпуније ће описати посматране скупове, односно појаве масовног карактера.

Расподела случајне променљиве X представља се графички у виду **хистограма** приказаног на сл 5.1. На оси X се наносе могуће вредности случајне променљиве X, а затим се изнад сваке тачке конструише правоугаоник чија је површина једнака одговарајућој вероватноћи.



Слика 5.1. Закон вероватноћа случајне променљиве X

Пошто је

$$\sum_{i=1}^{n} p_i = 1$$

на хистограму расподеле случајне променљиве, цела површина између осе X и криве којом је ограничен хистограм, треба да је једнака јединици.

На основу расподеле случајне променљиве X прекидног типа, могуће је одредити расподелу било које случајне променљиве Y која је дефинисана преко X.

Ако између случајних променљивих X и Y постоји веза

$$Y = g(X)$$

тада је закон вероватноћа Y дат у следећој табели

Y	$y_1=g(x_1)$	$y_2=g(x_2)$	$y_3=g(x_3)$	...
$p(y)$	p_1	p_2	p_3	...

Најчешће се посматра линеарна функција

$$Y = aX + b$$

при чему је $a \neq 0$. Тада је закон вероватноћа случајне променљиве Y дат табелом.

Y	ax_1+b	ax_2+b	ax_3+b	...
$p(y)$	p_1	p_2	p_3	...

Ако се зна закон вероватноћа случајне променљиве X, могуће је лако одредити вероватноће било ког случајног догађаја везаног за тај експеримент. Наиме, случајни догађај је подскуп скупа могућих вредности, тј. подскуп скупа реалних бројева

$$\{x_1, x_2, x_3, \ldots\}$$

Када одредимо подскуп који одговара догађају, на основу вероватноћа појединих тачака тог подскупа, лако се одређује вероватноћа посматраног догађаја.

На пример, ако се сматра да је рађање девојчице случајан догађај са вероватноћом 0.5, а дечака такође, са вероватноћом 0,5, колика је вероватноћа да ће у породици са четворо деце бити бар две девојчице? Означимо са X број девојчица у породици са четворо деце. Тада X може узети једну вредност из скупа

$$\{0, 1, 2, 3, 4\}$$

са вероватноћама

$$\{0{,}0625;\ 0{,}2500;\ 0{,}3750;\ 0{,}2500;\ 0{,}0625\}$$

Тако је закон вероватноћа за X дат табелом

X	0	1	2	3	4
$p(x)$	0.0625	0.2500	0.3750	0.2500	0.0625

Посматрани догађаји ће се реализовати ако се реализује једна од вредности из подскупа

$$\{2, 3, 4\}.$$

Тражена вероватноћа је једнака

$$P(2 \le X \le 4) = P(X = 2) + P(X = 3) + P(X = 4)$$

односно,

$$P(2 \le X \le 4) = 0,3750 + 0,2500 + 0,0625 = 0,6875$$

Ако, у посматраном примеру, знамо да у породици има један дечак и једна девојчица колика би била вероватноћа да ће у тој породици бити бар два дечака? Тражена вероватноћа је условна вероватноћа

$$P(A/B)$$

при чему је

$$A = (0 \le X \le 2)$$
$$B = (1 \le X \le 3)$$

Пошто је

$$A \cdot B = (1 \le X \le 2)$$

из дефиниције условне вероватноће следи да је

$$P(A/B) = \frac{P(A \cdot B)}{P(B)} = \frac{0,2500 + 0.3750}{0,2500 + 0,3750 + 0,2500} = 0,7143$$

Узмимо следећи пример. Ради утврђивања квалитета у великој серији, узима се један производ и врши комплетна анализа. Ако изабрани производ, у потпуности, одговара жељеном квалитету, даља анализа се не врши, а ако не одговара, бира се следећи и врши његова анализа, итд. Нека је вероватноћа за сваки производ, да одговара жељеном квалитету, једнака *0.8*, а трошкови анализе су *2000* динара. Интересује нас вероватноћа да ће трошкови контроле серије бити мањи од *5000* динара.

Означимо са *X* број производа који ће у контроли серије, бити анализирани. Тада *X* има закон вероватноће

X	1	2	3	...
$p(x)$	0.8	$0.2 \cdot 0.8$	$0.04 \cdot 0.8$	...

Означимо са Y трошкове контроле серије. Тада је

$$Y = 2000 \cdot X$$

тако да је закон вероватноћа за трошкове Y дат табелом:

Y	2000	4000	6000	...
$p(y)$	0.8	0.16	0.032	...

Трошкови контроле ће бити мањи од 5000, ако је анализиран један или два производа, па је вероватноћа да трошкови контроле буду мањи од 5000, једнака

$$P(Y < 5000) = P(X=1) + P(X=2)$$

односно,

$$P(Y < 5000) = 0.96$$

Закон верованоћа непрекидне случајне промењиве

При гађању мете мери се удаљеност поготка од центра. Промењива величина X биће један од реалних бројева из интервала (0;a), а при једном „извођењу" експеримента (у једном гађању), X ће узети само једну вредност из тог интервала.

Век трајања сијалице је промењива величина која ће узети вредност из интервала $(0;\ +\infty)$. За једну сијалицу (једно „извођење" експеримента), век трајања ће бити само један број из овог интервала.

У оваквим експериментима, посматрана промењива величина X може узети једну вредност из неког интервала. Сматраћемо да је то бесконачни интервал

$$(-\infty,\ +\infty) \qquad (5.10)$$

Ако је у експерименту интервал коначан тада ћемо сматрати да су тачке изван тог интервала немогући догађаји са вероватноћама једанаким нули.

Случајну промењљиву непрекидног типа треба да одредимо тако да нам је могуће тачно утврдити њену расподелу, као и код променљивих прекидног типа.

По аналогији са случајним променљивим прекидног типа, и овде морамо дефинисати закон вероватноћа тако да нам је могуће да за сваки случајни догађај везан за тај експеримент, одредимо његову вероватноћу.

У овом случају, интервал $(-\infty;\ +\infty)$, поделимо на подинтервале произвољно мале дужине Δx. Уместо тачака $x \in (-\infty;\ +\infty)$ посматраћемо интервале $\left(x-\frac{\Delta x}{2};x+\frac{\Delta x}{2}\right)$. За тако одређене интервале могуће је одредити једну функцију $f(x)$, преко које ћемо дефинисати вероватноће.

Дефиниција 5.3: *За промењљиву величину X кажемо да је случајна промењљива непрекидног типа ако узимање било које вредности из интервала* $(-\infty;\ +\infty)$ *представља случајан догађај и ако постоји функција f(x) таква да је вероватноћа, да ће се X наћи у призвољно малој околини тачке x, тј. у интервалу*

$$\left(x-\frac{\Delta x}{2};x+\frac{\Delta x}{2}\right) \tag{5.11}$$

једнака производу $f(x)\cdot\Delta x$ *; тј ако је*

$$P\left(x-\frac{\Delta x}{2}<X\le x+\frac{\Delta x}{2}\right)=f(x)\cdot\Delta x \tag{5.12}$$

и за сваку произвољно малу вредност Δx.

Функција f(x) назива се закон вероватноћа случајне промењљиве X.

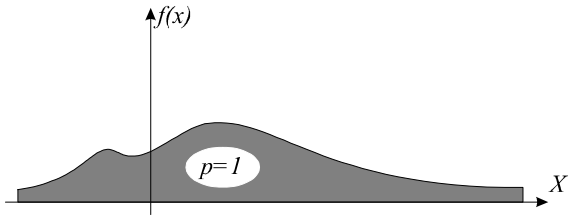
Према томе, случајна променљива непрекидног типа биће одређена функцијом $f(x)$, односно, *законом вероватноћа*.

Закон вероватноћа $f(x)$ мора задовољавати следећа два услова:

1. $f(x) \geq 0$, за свако $x \in (-\infty, +\infty)$
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$ (5.13)

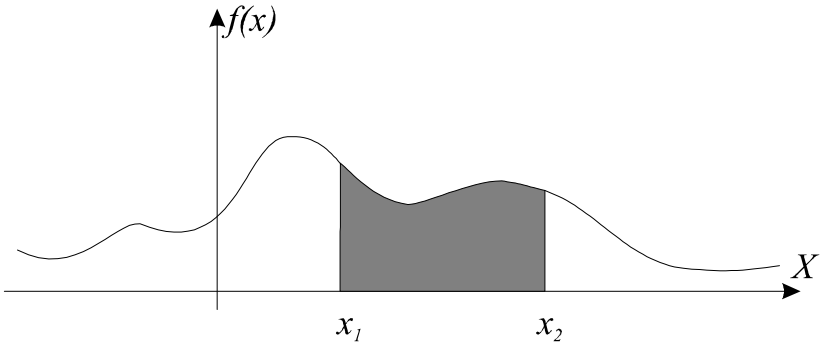
Први услов следи из (5.12) и особине ненегативности вероватноћа.

Пошто се у посматраном експерименту реализује тачно једна вредност x из $(-\infty; +\infty)$, онда је „збир" вероватноћа за све тачке у том интервалу једнак јединици (види сл. 5.2), а тај „збир" није ништа друго до интеграл (5.13). Геометријски, то је површина између X осе и закона вероватноћа $f(x)$.



Слика 5.2. Закон вероватноћа случајне променљиве X.

Најчешће нас интересује одређивање вероватноћа да ће се случајна променљива наћи у неком интервалу, рецимо интервалу $(x_1; x_2)$. Тражену вероватноћу ћемо наћи тако што ћемо, за све тачке из интервала $(x_1; x_2)$, „сабрати" одговарајуће вероватноће дате са (5.12). Тај збир није ништа друго до одређени интеграл функције $f(x)$ у границама $(x_1; x_2)$. На слици 5.3. то је шрафирана површина.



Слика 5.3. Вероватноћа да се X нађе у интервалу (x_1; x_2).

Тражена вероватноћа једнака је

$$P\left(x_1 < X \le x_2\right) = \int_{x_1}^{x_2} f\left(x\right) dx \tag{5.14}$$

Сваки случајни догађај, везан за посматрани експеримент, може се изразити преко интервала, а вероватноће интервала одредићемо преко (5.14). Зато ћемо, на основу закона вероватноћа $f(x)$, моћи одредити и вероватноће било ког случајног догађаја.

На пример, произвођач аутомобилских гума тврди да су његове гуме веома поуздане. Зна се да је за сваку гуму њен "век трајања" случајна променљива са законом вероватноћа

$$f(x) = \begin{cases} 25 \cdot 10^{-6} e^{25 \cdot 10^{-6} x} & за\ x > 0 \\ 0 & за\ x \le 0 \end{cases} \tag{5.15}$$

Функција (5.15) је, заиста закон вероватноћа. Задовољава услове:

1. $f(x) \ge 0$
2. $\int_{-\infty}^{+\infty} f\left(x\right) dx = \int_{-\infty}^{+\infty} 25 \cdot 10^{-6} e^{-25 \cdot 10^{-6} x} \cdot dx = 1$

Тражи се вероватноћа да ће купљена гума издржати 40.000 km, односно, $4\cdot10^4$ km. То ће се реализовати када је $X \geq 4\cdot10^4$, тако да је тражена вероватноћа једнака

$$P(X>40000)=25\cdot10^{-6}\int_{4\cdot10^4}^{\infty} e^{-25\cdot10^{-6}x}dx = 0.3678789$$

а вероватноћа да ће гума издржати 20.000 km, је

$$P(X > 20.000) = 0.6065$$

Узмимо још један пример. Нека X има закон вероватноћа

$$f(x) = \frac{2(b-x)}{b^2}$$

за $x \in [0;b]$. Потребно је одредити z тако да, са вероватноћом 0.5, тврдимо да ће случајна променљива X узети вредност већу од z.

Тражена вредност z је број за који је

$$P(X>z) = 0,5$$

односно,

$$\int_z^b f(x)\,dx = 0.5$$

Решење интеграла даје једначину

$$\left(\frac{z-b}{b}\right)^2 = 0,5$$

одакле се добија да је

$$z = b\left(1-\sqrt{0,5}\right)$$

Функција расподеле случајне променљиве

У практичним одређивањима вероватноћа случајних догађаја везаних за случајне променљиве, често смо заинтересовани за једну класу догађаја који се формулишу на следећи начин:

Случајна величина X узеће вредност не већу од z, при чему је *z* унапред одређен број из интервала $(-\infty, +\infty)$.

На пример, интересује нас догађај да ће тражња за једним производом бити мања од једне одређене вредности, рецимо 1.000 јединица у датом временском интервалу.

Или нас интересује догађај да ће, у аутоматској централи број позива бити мањи од неког броја, који представља капацитет централе.

За лакше одређвање вероватноћа оваквих догађаја, служи функција расподеле случајне променљиве *X*, која се дефинише на следећи начин:

Дефиниција 5.4.: *Функција расподеле случајне променљиве X је функција F(z) која, за свако*

$$z \in (-\infty;+\infty) \qquad (5.16)$$

представља вероватноћу да случајна променљива X неће узети вредност већу од z, тј.

$$F(z) = P(X \le z). \qquad (5.17)$$

Пошто су вредности функције расподеле *F(z)* вероватноће, то значи да ће та функција имати вредности између нула и један, тј. за свако $z \in (-\infty;+\infty)$

$$0 \le F(z) \le 1 \qquad (5.18)$$

Функција расподеле случајне променљиве *X* је функција са следећим особинама:

***Особина* 1.** $F(-\infty)=0$

$F(-\infty)$ је вероватноћа да ће X бити мање од $-\infty$, а то је немогућ догађај чија је вероватноћа једнака 0.

***Особина* 2.** $F(+\infty)=1$

$F(+\infty)$ је вероватноћа да ће X бити мање од $+\infty$, а то је сигуран догађај са вероватноћом једнаком један.

***Особина* 3.** *Функција F(z) је монотоно неопадајућа функција.*

Заиста, ако су z' и z'' две вредности, такве да је $z' \leq z''$, онда догађај

$$D' = (-\infty < X \leq z')$$

имплицира догађај

$$D'' = (-\infty < X \leq z'')$$

што значи да је

$$P(D') \leq P(D'') \tag{5.19}$$

Пошто је

$$P(D') = F(z')$$

и

$$P(D'') = F(z''),$$

онда, из (5.19) следи да је

$$F(z') \leq F(z'') \tag{5.20}$$

Неједнакост (5.20) одређује монотоност функције.

***Особина* 4.** *Код случајне промењиве X прекидног типа, функција расподеле F(z) је степенаста функција која у тачкама*

$$x_1, x_2, x_3, \ldots$$

има скокове једнаке одговарајућим вероватноћама.

Наиме, за свако

$$
\begin{aligned}
&z \in (-\infty;\, x_1) \Rightarrow F(z) = 0,\\
&z \in (x_1;x_2) \Rightarrow F(z) = P(X=x_1)=p_1,\\
&z \in (x_2;x_3) \Rightarrow F(z) = P(X=x_1)+P(X=x_2)=p_1 + p_2\\
&\ldots\\
&z \in (x_k;x_{k+1}) \Rightarrow F(z) = \sum_{i=1}^{k} P(X=x_i)=\sum_{i=1}^{k} p_i\\
&\ldots
\end{aligned}
\qquad (5.21)
$$

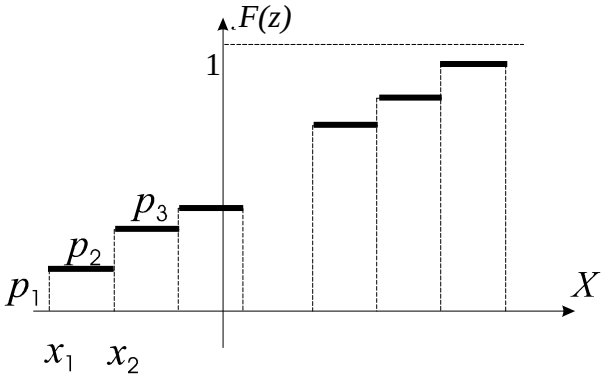
У овом случају, функција расподела *F(z)* је функција која има облик приказан на слици 5.4.

На основу израза (5.21) може се лако одредити фунцкија расподеле *F(z)*, када се зна закон вероватноћа случајне променљиве *X*.

Обрнуто, ако је позната функција расподела *F(z)*, онда се из (5.21) добија да је

$$
\begin{aligned}
p_1 &= F(x_1)\\
p_2 &= F(x_2)-F(x_1)\\
p_3 &= F(x_3)-F(x_2)\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
p_k &= F(x_k)-F(x_{k-1})\\
&\ldots\ldots\ldots\ldots\ldots\ldots,
\end{aligned}
\qquad (5.22)
$$

са чиме је потпуно одређен закон вероватноћа случајне промељиве *X*.



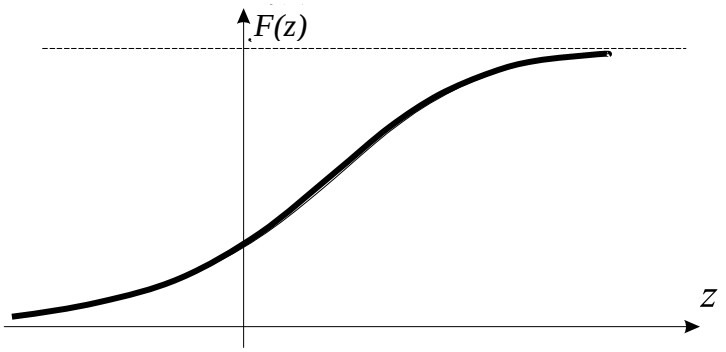
Слика 5.4. Функција расподеле случајне промењиве X прекидног типа.

***Особина* 5.** *Ако је X непрекидна случајна промељива са законом вероватноћа f(x), тада је, за свако* $z \in (-\infty;+\infty)$

$$F(z)=\int_{-\infty}^{z} f(x)\,dx \tag{5.23}$$

Функција расподеле *F(z)* је непрекидна и монотоно неопадајућа функција, или је прекидна за коначно много тачака прекида.

Функција расподеле непрекидне случајне промењиве *X* има облик криве, приказане на слици 5.5.



Слика 5.5. Функција расподеле непрекидне случајне промењиве X.

Када нам је познат закон вероватноће *f(x)* једне случајне променљиве, решавањем интеграла у (5.23), одредићемо њену функцију расподеле.

Обрнуто, ако нам је позната функција расподеле *F(z),* закон вероватноћа *f(x)* одредићемо преко извода функција *F(z).* Тада је за свако $z \in (-\infty;+\infty)$

$$f(x) = \frac{dF(z)}{dz}\Big|_{z=x} \tag{5.24}$$

***Закључак**: Под расподелом случајне променљиве X треба подразумевати, или њен **Закон вероватноћа**, или **Функцију расподеле**. Када знамо закон вероватноћа, лако одређујемо функцију расподеле, и обрнуто, када знамо функцију расподеле случајне промељиве X, лако одређујемо њен закон вероватноћа. Расподелом случајне променљиве одређена је максимална количина информација о експерименту са случајним исходима.*

Посматрајмо следећи пример. Познато је да је свака четврта особа алергична на антибиотике. Ради провере новог лека, на случајан начин је изабрана група од седам лица. Одредићемо закон вероватноћа случајне променљиве *X*, који представља број изабраних лица који су алергични на антибиотике.

Случајна променљива *X* је прекидног типа и може узети једну од вредности:

0, 1, 2, 3, 4, 5, 6, 7

За сваку изабрану особу, вероватноћа да буде алергична на антибиотике има вредност

$$p = 0.25$$

а избор особа вршимо независно. Значи, седам пута „понављамо" експеримент, и у сваком „понављању" посматрамо случајни догађај *A*: „Изабрано лице је алергично на антибиотике". Догађај *A* се, у сваком „понављању" реализује са вероватноћом 0,25. Случајна промељива *X*

представља број „понављања експеримента" у којима се догађај A реализовао. Зато је вероватноћа да X узме неку вредност k једнака

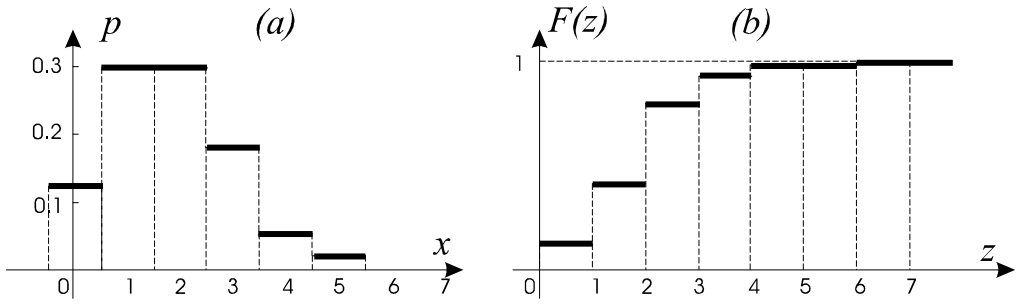
$$P(X=k)=P_7(k)=\binom{7}{k}\,0.25^{k}\;0.75^{7-k}$$

тј. вероватноће за поједине вредности случајне промељиве X су тзв. *биномне вероватноће* $P_7(k)$.

Закон вероватноћа случајне променљиве X дат је табелом:

X	0	1	2	3	4	5	6	7
$p(x)$	0.133	0.311	0.311	0.173	0.057	0.011	0.001	0.0001

Његов хистограм је приказан на слици 1.6.(а).



Слика 5.6. Закон вероватноћа и функција расподеле.

Функција расподеле ове случајне променљиве је дата са:

$$z \in (-\infty;\ 0) \Rightarrow F(z) = 0;$$
$$z \in [0;\ 1) \Rightarrow F(z) = 0{,}1335;$$
$$z \in [1;\ 2) \Rightarrow F(z) = 0{,}4450;$$
$$z \in [2;\ 3) \Rightarrow F(z) = 0{,}7565;$$

$$z \in [3;\ 4) \quad \Rightarrow F(z) = 0{,}9295;$$
$$z \in [4;\ 5) \quad \Rightarrow F(z) = 0{,}9872;$$
$$z \in [5;\ 6) \quad \Rightarrow F(z) = 0{,}9987;$$
$$z \in [6;\ 7) \quad \Rightarrow F(z) = 0{,}9999;$$
$$z \in [7;\ -\infty) \Rightarrow F(z) = 1{,}000;$$

Ова функција приказана је на Слици 5.6 (b).

На сличан начин, посматраћемо пример непрекидне случајне промељиве X са законом вероватноћа

$$f(x)=\begin{cases}\frac{1}{4}(3-x) & за\ 0<x<2\\ 0 & за\ x \le 0,\ x \ge 2\end{cases} \tag{5.25}$$

Функција расподеле $F(z)$ је одређена на следећи начин:

$$за\ z \in (-\infty;\ 0] \Rightarrow F(z) = 0$$

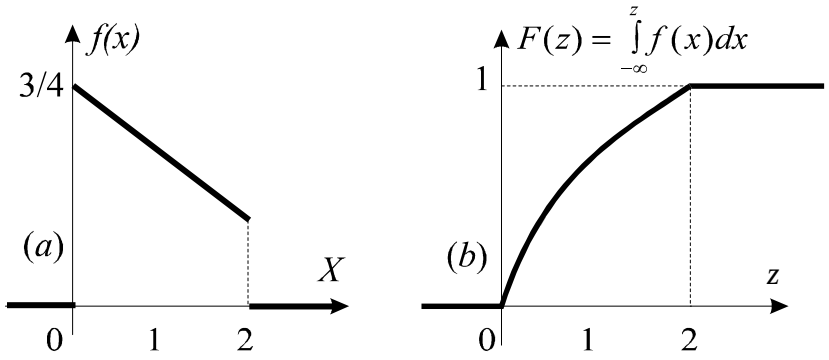
$$за\ z \in (0;2) \Rightarrow F(z) = \frac{1}{4}\int_0^z (3 - x)dx = \frac{z}{4}(3 - \frac{z}{2})$$

$$за\ z \in [2;+\infty) \Rightarrow F(z) = \frac{1}{4}\int_0^2 (3 - x)dx = 1$$

Према томе, функција расподеле $F(z)$, за случајну промељиву X, са законом вероватноћа (5.25), дата је функцијом

$$F(z)=\begin{cases}0 & за\ z \le 0\\ \frac{z}{4}\left(3-\frac{z}{2}\right) & за\ 0<z<2\\ 1 & за\ z \ge 2.\end{cases} \tag{5.26}$$

Функције (5.25) и (5.26) приказане су на слици 5.7. (a) и (b).



Слика 5.7. Закон вероватноћа и функција расподеле случајне промењиве.

5.2. АНАЛИЗА СЛУЧАЈНИХ ПРОМЕНЉИВИХ

Карактериситке једне случајне променљиве X садржане су у њеној расподели датој законом вероватноћа или функцијом расподеле. Зато се, у употреби случајних променљивих, мора радити са расподелом, а то отежава применљивост теорије вероватноће базираној на случајним променљивим.

Са друге стране, у пракси, најчешће не знамо расподелу случајне променљиве. У таквим случајевима се показало да је познавање одређених нумеричких карактеристика довољно за решавање практичних проблема примене теорије вероватноћа. Зато је потребно одредити нумеричке карактеристике по унапред одређеним правилима, па уместо целог закона вероватноћа, односно фунцкије расподеле случајне променљиве, посматрати одређени број параметара, који ће, у довољној мери, описивати посматрану променљиву.

Као што се, пре употребе неког грађевинског материјала, прво морају утврдити његове карактеристике, и то тако што ће се извршити мерење вредности посматраних карактеристика, тако се и у употреби случајних променљивих прво морају "измерити" одређене карактеристике. Претходно се мора обезбедити процедура или начин на који се врши „мерење” жељених карактеристика.

Одређивање појединих карактеристика, односно параметара случајне променљиве X, врши се преко ***очекиване вредности*** или ***математичког очекивања***. Зато ћемо се, укратко, упознати са овим појмом, а затим ћемо прећи на одређивање појединих параметара случајних променљивих.

Очекивана вредност

У анализи статистичких података на појавама масовног карактера, служили смо се расподелом фреквенција и могућим вредностима посматраног обележја. При томе смо посматрали поједине вредности, или њихове функције посматраног обележја, и њих смо множили апсолутним фреквенцијама, затим сабирали и делили са укупним бројем елемената статистичког скупа. Тако смо, на пример одредили аритметичку средину

$$\overline{x} = \sum_{i=1}^{n} x_i \cdot p_i \tag{5.28}$$

при чему су p_i relativne frekvencije ($p_i = \frac{f_i}{N}$).

Пошто се расподела статистичког скупа може сматрати расподелом случајне промењиве *X*, онда аритметичка средина може представљати „просечну вредност" случајне промењиве, па ће нам (5.28) послужити и за дефиницију очекиване вредности случајне промењиве.

Дефиниција 5.5.: *Нека је дата расподела случајне промењиве X*

X	x_1	x_2	x_3	...
$p(x)$	p_1	p_2	p_3	...

Ако ред

$$\sum_{i=1}^{\infty} x_i \cdot p_i$$

конвергира, његову суму називамо очекиваном вредношћу, или краће очекивањем случајне поменљиве X и означавамо га са $E(X)$ *, па је*

$$E(X) = \sum_{i=1}^{\infty} x_i \cdot p_i \tag{5.29}$$

Дефиниција 5.6.: *Нека је f(x) закон вероватноћа случајне променљиве X. Ако постоји интеграл*

$$\int_{-\infty}^{\infty} xf(x)\,dx$$

његова вредност се назива очекиваном вредношћу, или краће, очекивањем случајне променљиве X и означава се са $E(X)$*, тако да је*

$$E(X) = \int_{-\infty}^{\infty} xf\,(x)\,dx \tag{5.30}$$

Ако се, преко случајне променљиве *X*, дефинише нова случајна променљива *Y* функцијом

$$Y = g(X)$$

тада је *очекивана вредност променљиве Y* одређена преко расподеле променљиве *X*, као сума реда

$$E\left[g(X)\right] = \sum_{i=1}^{\infty} g(x_i)\,p_i \tag{5.31}$$

односно, као интеграл

$$E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)\ f(x)\,dx \tag{5.32}$$

Очекивање случајне променљиве *X* и очекивање функција случајне променљиве *X*, је основни апарат у анализи случајних променљивих. Мада се појам очекиване вредности, у почетку, везао за хазардне игре, његова велика употреба је проширила овај појам толико да је постао свакодневни термин. Сада се доношење најједноставнијих

пословних одлука не може ни замислити без употребе очекивања везаних за резултате одређених акција.

Пример. Типичан проблем **залиха** је проблем одређивања броја јединица неког артикла. Тражња за тим артиклом се може сматрати случајном променљивом. Потражње су нам алтернативе за набавку, али нам нису познати трошкови, односно, зараде за поједине алтернативе. Очекивани трошкови, односно, очекивана зарада ће нам послужити као критеријум за избор најбоље алтернативе.

Предпоставимо да продавница треба да купи 50, 60 или 70 сезонских хаљина, а да је њена зарада 100 динара по комаду. На крају сезоне непородате хаљине се могу продати по сниженој цени која ће довести до губитака од 75 динара по непродатој хаљини. Нека се зна да је вероватноћа да се прода 50 хаљина, 0.30, а да се прода 60, је 0.50, док је вероватноћа продаје 70 хаљина 0.20. У Табели 5.1. дате су очекиване зараде за посматране одлуке.

Из последње врсте Табеле 5.1, види се да је најбоља алтернатива набавити 60 хаљина. Код ње је очекивана зарада 5475 динара.

Тражња	Вероватноћа	Алтернативе					
		Набавити 50 хаљина		Набавити 60 хаљина		Набавити 70 хаљина	
		Условна зарада	Очекивана зарада	Условна зарада	Очекивана зарада	Условна зарада	Очекивана зарада
50	0,30	5000	1500	4250	1275	3500	1050
60	0,50	5000	2500	6000	3000	5250	2625
70	0,20	5000	1000	6000	1200	7000	1400
Укупна очекивана зарада			5000		5475		5075

Табела 5.1. Рачунање очекиване зараде

Ако случајна величина X има расподелу

X	10	20	30	40
$p(x)$	0.1	0.3	0.5	0.1

тада је њено очекивање једнако

$$E(X) = 10 \cdot 0.1 + 20 \cdot 0.3 + 30 \cdot 0.5 + 40 \cdot 0.1 = 26$$

а очекивана вредност функције

$$g(X) = 2X^2 - 900$$

једнака је

$$E\left[g(X)\right] = (2 \cdot 10^2 - 900) \cdot 0.1 + (2 \cdot 20^2 - 900) \cdot 0.3 +$$
$$+ (2 \cdot 30^2 - 900) \cdot 0.5 + (2 \cdot 40^2 - 900) \cdot 0.1 = 560$$

За случајну променљиву X, код које је закон вероватноћа

$$f(x) = \begin{cases} \frac{1}{50} x & za\ 0 < 1 \le 10 \\ 0 & za\ x \le 0, x > 10 \end{cases}$$

Очекивана вредност случајне променљиве X је једнака

$$E(x) = \frac{1}{50} \int_0^{10} x^2 dx = 20/3$$

Очекивана вредност функције

$$g(x) = 2X^2 - 4$$

је једнака интегралу

$$E\left[2X^2 - 4\right] = \frac{1}{50} \int_0^{10} (2x^2 - 4)x\, dx = 28/3$$

Ради лакшег одређивања очекиване вредности, навешћемо неке њене особине које ће, уједно, помоћи и бољем разумевању самог појма.

Особина 1: *Очекивана вредност константе c једнака је тој константи c.*

Наиме, ако случајна величина X може узети само једну вредност, онда је њена расподела

X	c
$p(x)$	1

тако да је њена очекивана вредност једнака

$$E(x) = c \cdot 1 = c \tag{5.33}$$

Особина 2: *Очекивање производа константе c и функције* $g(X)$ *једнако је производу те константе c и очекивања функције* $g(X)$, *тј.*

$$E\left[c \cdot g(X)\right] = c \cdot E\left[g(X)\right] \tag{5.34}$$

Заиста, ако X има закон вероватноћа $f(x)$, тада је

$$E\left[c \cdot g(X)\right] = \int_{-\infty}^{\infty} cg(X)\, f(x)\, dx = c \int_{-\infty}^{\infty} g(X)\, f(x)\, dx = c \cdot E\left[g(X)\right]$$

На сличан начин се може проверити особина 2 и за случајну величину X прекидног типа.

Особина 3: *Очекивање збира две функције случајне промењиве, једнако је збиру њихових очекивања.*

Наиме, ако су $g(X)$ и $h(X)$ функције, тада је

$$E\left[g(X) + h(X)\right] = E\left[g(X)\right] + E\left[h(X)\right] \tag{5.35}$$

Из ових особина следи следећа

Особина 4: *Очекивања линеарне функције случајне промењиве једнако је линеарној функцији очекивања те промењиве.*

Нека је Y промењива дата са

$$Y = aX + b$$

тада је њена очекивана вредност једнака

$$E(Y) = aE(X) + b \tag{5.36}$$

Особина 5: *Очекивање линеарне комбинације две случајне променљиве, једнако је линеарној комбинацији њихових очекивања.*

Наиме, ако је

$$Z = a_1 X + a_2 Y$$

očekivanje Z biće jednako

$$E(Z) = a_1 E(X) + a_2 E(Y) \tag{5.37}$$

Специјално, очекивање збира, или разлике две случајне променљиве једнако је збиру, односно, разлици њихових очекивања, тј.

$$E(X \pm Y) = E(X) \pm E(Y) \tag{5.38}$$

На пример, бележећи потражњу за својим производима, пекар је утврдио да она варира на следећи начин

Потражња *(комада)*	10000	11000	12000	13000	14000
Број дана	18	90	120	60	12

Користећи ***закон великих бројева*** можемо посматрати потражњу X као случајну променљиву са распоредом

X	10000	11000	12000	13000	14000
$P(x)$	18/300	90/300	120/300	60/300	12/300

Очекивана потражња једнака је

$$E(X) = 11.860$$

Нека су трошкови производње 0.15 дин/ком а продајна цена 0.20 динара. Непродати производи се други дан продају по цени 0.09 динара. Одредићемо очекивану зараду. Ако је Q произведена количина, X потражња а Y зарада, добиће се

$$Y = 0.05 \qquad \textit{за } X \geq Q$$
$$Y = 0.05 - 0.06(Q - X) \qquad \textit{за } X < Q$$

У првом случају је очекивана зарада

$$E(Y) = 0.05 \cdot Q \tag{5.39}$$

а у другом случају

$$E(Y) = 0.11E(X) - 0.06 \cdot Q = 1304.6 - 0.06 \cdot Q \tag{5.40}$$

Користећи (5.39) и (5.40), може се за сваку произведену количину, одредити очекивана зарада, а на основу тога изабрати најбоља одлука о количини коју треба производити.

Посматрајмо случајну промењиву X са законом вероватноћа

$$f_1(x) = \begin{cases} 2x & za\ 0 < x < 1 \\ 0 & za\ x \leq 0;\ x \geq 1 \end{cases}$$

и случајну промењиву Y са законом вероватноћа

$$f_2(y) = \begin{cases} e^{-y} & za\ y > 0 \\ 0 & za\ y \leq 0 \end{cases}$$

Очекивана вредност функције

$$Z = 3X + 2Y$$

једнака је

$$E(Z) = 3E(X) + 2E(Y) = 3\int_0^1 2x^2 dx + 2\int_0^\infty xe^{-x} dx = 2 + 2 = 4$$

Моменти случајне променљиве

Веома важну класу очекивања функција случајних променљивих, чине тзв. *моменти*. Практично, скоро цела анализа случајних променљивих базирана је на моментима и функцијама момената.

Разликују се две врсте момената: *обични и централни моменти*.

Дефиниција 5.7.: *Обични моменат k-тог реда случајне променљиве X је математичко очекивање k-тог степена те променљиве.*

Обичне моменте реда k означићемо са m_k, тако да је

$$m_k = E(X^k) \qquad за\ k = 0,\ 1,\ 2,\ \ldots \tag{5.41}$$

Ако је X прекидна случајна променљива, обични моменат реда k, једнак је збиру

$$m_k = \sum_{i=1}^{\infty} x_i^k\, p_i \tag{5.42}$$

а ако је X непрекидног типа , обични моменат је једнак интегралу

$$m_k = \int_{-\infty}^{\infty} x^k\, f(x)\, dx \tag{5.43}$$

Обичан моменат нултог реда једнак је јединици, јер је

$$m_0 = E(X^0) = E(1) = 1$$

Обични моменат првог реда је, уствари, *очекивана вредност* или краће, *очекивање* случајне величине X. Означаваћемо га са

$$m = E(X) \tag{5.44}$$

Дефиниција 5.8.: *Централни моменат k-tog реда случајне променљиве X, је очекивана вредност k-тог степена разлике X и очекиване вредности m.*

Централни моменат реда *k* означаваћемо са μ_k, тако да је

$$\mu_k = E(X-m)^k \qquad za\ k = 0,\ 1,\ 2,\ \ldots \tag{5.45}$$

За случајну промењиву прекидног типа, централни моменти су једнаки збировима

$$\mu_k = \sum_{i=1}^{\infty} (x_i - m)^k p_i \tag{5.46}$$

а за непрекидну случајну промењиву, централни моменти су једнаки интегралима

$$\mu_k = \int_{-\infty}^{\infty} (x-m)^k f(x)\,dx \tag{5.47}$$

И централни моменат нултог реда је једнак један, јер је

$$\mu_0 = E(X-m)^0 = E(1) = 1$$

а ***централни моменат првог реда*** *једнак је нули*. Наиме, из особина очекивања, следи да је

$$\mu_1 = E(X-m) = E(X) - m = m - m = 0$$

Централни моменти се лакше одређују преко обичних момената. Ако k-ti степен разлике $(X-m)$, изразимо преко биномног образца добићемо

$$(X-m)^k = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} m^{k-i} X^i$$

па је, због особина очекиване вредности

$$E(X-m)^k = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} m^{k-i} E(X^i)$$

односно

$$\mu_k = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} m^{k-i} m_i \qquad \text{за } k=0, 1, 2, \ldots \tag{5.48}$$

Специјално

$$\begin{aligned} \mu_2 &= m_2 - m^2 \\ \mu_3 &= m_3 - 3m_2 m + 2m^3 \\ \mu_4 &= m_4 - 4m_3 m + 6m_2 m^2 - 3m^4 \end{aligned} \tag{5.49}$$

Обрнуто, користећи везе (5.48), можемо из централних момената одредити обичне моменте. Зато се, практично, из расподеле случајне променљиве X, одреди једна класа момената, а затим се из израза (5.48), одређује друга класа.

На пример, нека случајна променљива X има тзв. *хипергеометријску расподелу*

X	0	1	2	...	k
$p(x)$	p_0	p_1	p_2	...	p_k

при чему је за свако i=0,1,2,3....,k.

$$p_i = \frac{\binom{m}{i}\binom{n}{k-i}}{\binom{m+n}{k}} \tag{5.50}$$

Одредићемо њена прва два обична и централна момента.

Због особине закона вероватноћа, из (5.50) добија се

$$\sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i} = \binom{m+n}{k}$$

Први обичан моменат је

$$E(X)=\frac{\sum_{i=0}^{k} i\binom{m}{i}\binom{n}{k-i}}{\binom{m+n}{k}}=\frac{1}{\binom{m+n}{k}}\sum_{i=1}^{k} m\binom{m-1}{i-1}\binom{n}{k-i}=$$

$$=\frac{1}{\binom{m+n}{k}}\sum_{j=0}^{k-1} m\binom{m-1}{j}\binom{n}{k-1-j}=\frac{m}{\binom{m+n}{k}}\binom{m-1+n}{k-1}=\frac{mk}{m+n}$$

Сличним операцијама може се одредити и други моменат, с тим што у збиру

$$\sum_{i=0}^{k} i^2 p_i$$

замењујемо

$$i^2=i(i-1)+i$$

а збир растављамо на два збира. Тако ће се добити

$$m_2=E(X^2)=\frac{m(m-1)\,k(k-1)}{(m+n)(m+n-1)}+\frac{mk}{m+n}$$

Први централни моменат је једнак нули, а други централни моменат је

$$\mu_2=m_2-\left[E(X)\right]^2=\frac{mnk(m+n-k)}{(m+n)^2(m+n-1)}$$

На сличан начин би одредили трећи и четврти обични и централни моменат.

Посматрајмо пример непрекидне случајне промењиве са законом вероватноћа

$$f(x)=\begin{cases}2(1-x) & \text{за } 0<x<1\\ 0 & \text{за } 0\le x\le 1\end{cases}$$

Обични моменат k-тог реда, једнак је

$$m_k = E(X^k) = 2\int_0^1 x^k(1-x)dx = \frac{2}{(k+1)(k+2)}$$

Специјално

$$m_1 = \frac{2}{2\cdot 3} = \frac{1}{3} \qquad m_2 = \frac{2}{3\cdot 4} = \frac{1}{6}$$
$$m_3 = \frac{2}{4\cdot 5} = \frac{1}{10} \qquad m_4 = \frac{2}{5\cdot 6} = \frac{1}{15}$$

тако да је

$$\mu_1 = 0$$
$$\mu_2 = \frac{1}{6} - \frac{1}{9}$$
$$\mu_3 = \frac{1}{10} - 3\frac{1}{6}\cdot\frac{1}{3} + 2\cdot\left(\frac{1}{3}\right)^3$$
$$\mu_4 = \frac{1}{10} - 4\frac{1}{10}\cdot\frac{1}{3} + 6\cdot\frac{1}{6}\left(\frac{1}{3}\right)^2 - 3\left(\frac{1}{3}\right)^4$$

Функција генератриса

Функција генератриса омогућава лакше одређивање *момената*, а исто тако служи за доказивање одређених теорема математичке статистике. Функција генератриса се дефинише преко функције

$$Y = e^{tX}$$

Дефиниција 5.9: *За свако реално t, очекивана вредност функције e^{tX} назива се функција генератриса случајне променљиве X.*

Функцију генератрису означавамо са g(t). Дефинисана је изразом

$$g(t) = E\left[e^{tX} \right] \tag{5.51}$$

Ако је X прекидног типа, функција генератриса је збир

$$g(t) = \sum_{i=1}^{\infty} e^{tx_i} p_i \tag{5.52}$$

а ако је непрекидног типа, функција генератриса је интеграл

$$g(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx \tag{5.53}$$

Функција генерариса g(t) се може одредити и преко момената. Наиме, ако се функција e^{tx} развија у ред, добићемо

$$e^{tX} = \sum_{i=0}^{\infty} X^i \frac{t^i}{i!}$$

тако да је

$$E(e^{tX}) = \sum_{i=0}^{\infty} \frac{t^i}{i!} E(X^i)$$

односно

$$g(t) = \sum_{i=0}^{\infty} \frac{t^i}{i!} m_i \tag{5.54}$$

Обрнуто, ако одредимо k-ти извод функције (5.54) по параметру t, и ставимо $t = 0$, добићемо

$$\left. \frac{d^k g(t)}{dt^k} \right|_{t=0} = m_k \tag{5.55}$$

што значи да је *k*-ти *извод функције генератрисе, у тачки* $t = 0$ *једнак обичном моменту реда k*. Одређивање момената преко функције генератрисе је веома једноставно.

Проблем инверзије. Видели смо да се на основу расподеле могу одредити моменти m_1, m_2, m_3,... . У применама, често, имамо велики број података добијених мерењем случајне променљиве *X*, али не знамо тачно њену расподелу. На основу података можемо приближно „оценити" моменте, па се поставља питање да ли се на основу момената може одредити расподела и да ли је то могуће учинити на једнозначан начин?

За све расподеле које се користе у пракси, одговор на постављено питање је позитиван. (***Расподеле одређене Pearson-овим системом кривих***).

Iz (5.54) следи да се функција генератриса g(t) може одредити на основу момената, а за функцију g(t) важи следећа теорема.

Теорема 5.1. Закон вероватноћа, односно функција расподеле случајне променљиве *X*, једнозначно је одређена функцијом генератрисом (ако та функција постоји).

Пример. Случајна променљива *X* има расподелу

X	*0*	*1*	*2*	...	*n*
p(x)	p_0	p_1	p_2	...	p_n

при чему је, за свако k = 0, 1, 2,... ,n

$$p_k = \binom{n}{k} p^k \; q^{n-k} \qquad p + q = 1$$

Њена функција генератриса ће бити

$$g(t) = \sum_{i=0}^{n} e^{t \cdot k} \binom{n}{k} p^k \; q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \; (pe^t)^k \; q^{n-k} = (q + pe^t)^n$$

тј.

$$g(t) = (q + pe^t)^n$$

Пример. Нека X има закон вероватноћа

$$f(x) = \begin{cases} e^{-x} & za\ 0 \le x < \infty \\ 0 & za\ x < 0 \end{cases}$$

функција генератриса ће бити

$$g(t) = \int_0^{\infty} e^{tx} e^{-x} dx = \frac{1}{1-t}$$

Параметри случајних променљивих

Како смо испитивали статистичка обележја преко параметара које смо рачунали по одређеним правилима, тако ћемо и случајне променљиве испитивати преко параметара добијених из њихових расподела. Биће то групе параметара које ће изражавати одређене особине расподела, а то су: Мере централне тенденције; Мере варијабилитета; Мере асиметрије и Мере спљоштености.

Мере централне тенденције

Прву групу параметара случајних променљивих чине параметри који показују *централну тенденцију расподеле*. То су: средине (очекивана, хармонијска и геометријска средина), модус и медијана. Сви ови параметри су дефинисани за статистичка обележја. Овде ћемо само уопштити дате дефиниције.

Очекивана вредност, или краће ***средина***, је очекивана вредност случајне величине X, тј.

$$m = E(X) \tag{5.56}$$

Хармонијска средина је дата преко очекивања реципрочне вредности случајне променљиве, изразом

$$\frac{1}{H} = E\left(\frac{1}{X}\right) \tag{5.57}$$

Геометријска средина је дефинисана преко очекиване вредности логаритма случајне променљиве *X*, тако да је

$$\log G = E(\log X) \tag{5.58}$$

Модус случајне променљиве *X* је она вредност M_0 за коју расподела, односно закон вероватноћа има највећу вредност.

Медијана случајне променљиве *X* је она вредност (означимо је са M_e) за коју је

$$P\{X \le M_e\} = P\{X > M_e\}$$

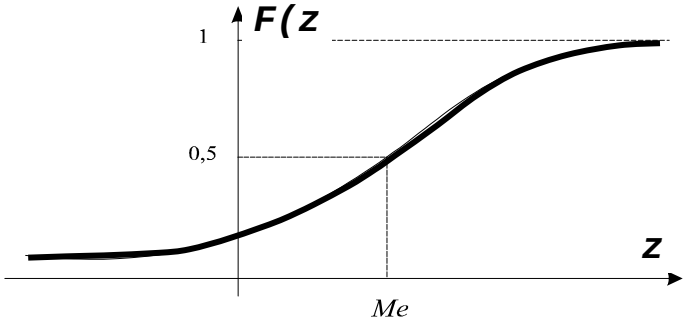
Вероватноћа да ће *X* бити мање од медијане, једнака је 0.5, а толика је и вероватноћа да ће *X* бити веће од медијане.

За случајну променљиву непрекидног типа, медијана је решење једначине дате интегралом

$$\int_{-\infty}^{M_e} f(x)\,dx = 0.5$$

Ако је позната функција расподеле $F(z)$, онда је медијана она вредност за коју је $F(M_e) = 0.5$.

То је, у ствари, апсциса пресечне тачке праве $F(z) = 0.5$ и функције расподеле $F(z)$ (види слику 5.8).



Слика 5.8. Медијана и функција расподеле.

Мере варијабилитета

Друга група параметара случајних променљивих су параметри који показују варијабилитет. То су следећи параметри: средња девијација, варијанса, стандардна девијација и коефицијент варијације.

Средња девијација је очекивана вредност апсолутне вредности разлике X и његовог очекивања, тј.

$$e_m = E\left(\left|X - m\right|\right) \tag{5.65}$$

Ако је X прекидног типа, средња девијација је једнака збиру

$$e_m = \sum_{i=1}^{\infty} \left|x_i - m\right| p_i$$

а ако је непрекидног типа, средња девијација је једнака интегралу

$$e_m = \int_{-\infty}^{\infty} \left|x - m\right| f(x)\, dx$$

Варијанса случајне променљиве *X је други централни моменат променљиве X.*

Означавамо је са σ^2. У ствари, варијанса је дефинисана изразом

$$\sigma^2 = \mu_2 = E(X - m)^2 = m_2 - m^2 \tag{5.66}$$

Lako je proveriti da varijansa ima sledeće osobine:

***Особина 1*:** Варијанса случајне променљиве *X* је једнака нули онда и само онда када је $X = const$.

***Osobina 2*:** Ako je slučajna promenljiva *Y* linearna funkcija

$$Y = aX + b$$

тада је варијанса *Y* једнака производу a^2 и варијансе *X*, тј.

$$\sigma^2(Y) = a^2\sigma^2(X)$$

Особина 3: *Очекивана вредност квадрата разлике X и константе c, имаће минимум ако је та константа* $c = m$, *а та минимална вредност је једнака варијанси случајне променљиве X, тј.*

$$\min_c E(X - c)^2 = E(X - m)^2 = \sigma^2$$

Стандардна девијација случајне променљиве X је позитивна вредност корена варијансе, тј.

$$\sigma = +\sqrt{\sigma^2}$$

Коефицијент варијације *случајне променљиве X је процентуално изражени количник стандардне девијације и средине случаја променљиве X, тј.*

$$V = \frac{\sigma}{m} \cdot 100\%$$

Коефицијенти асиметрије и спљоштености

*Као **мера асиметрије** расподеле, најчешће се користи тзв. први Pearson-ов коефицијент дат количником*

$$\beta_1 = \frac{\mu_3}{\sigma^3}$$

где је μ_3, трећи централни моменат случајне променљиве *X*. Ако је расподела симетрична, β_1 има вредност нула. Усвојићемо је обрнуто, ако је $\beta_1 = 0$ рећи ћемо да *X* има симетричну расподелу.

Други Pearson-ов коефицијент је количник

$$\beta_2 = \frac{\mu_4}{\sigma^4}$$

где је μ_4, четврти централни моменат. Представља ***меру спљоштености***. Код расподеле са „Нормалном спљоштеношћу", β_2 је једнако 3. У противном, β_2 је, или веће или мање од 3.

Чебишевљева теорема

Често се у пракси, одређује приближна вредност вероватноће да ће се случајна величина *X* наћи у некој околини свог очекивања. Чебишевљева теорема омогућава да се одреди та вероватноћа.

Теорема 5.2. Код сваке случајне променљиве *X*, која има коначну варијансу σ^2, за свако произвољно ε испуњена је следећа неједначина.

$$P\left\{\left|X - m\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{\varepsilon^2} \qquad (5.67)$$

Доказ. Нека је X непрекидног типа са законом вероватноће $f(x)$. Тада је, због ограничења функције

$$|X - m| \geq \varepsilon$$

односно,

$$(X - m)^2 \geq \varepsilon^2$$

$$\int_{-\infty}^{\infty} (x - m)^2 f(x)\, dx \geq \int_{(x-m)^2 \geq \varepsilon^2} (x - m)^2 f(x)\, dx \qquad (5.68)$$

па је

$$\int_{-\infty}^{\infty} (x - m)^2 f(x)\, dx \geq \varepsilon^2 \int_{(x-m)^2 \geq \varepsilon^2} f(x)\, dx$$

Интеграл на левој страни (5.68) је варијанса случајне променљиве X, а на десној страни, интеграл представља вероватноћу дату у (5.67). Зато из (5.68) следи (5.67).

Из Чебишевљеве теореме добија се следећа неједнакост за вероватноћу да ће се X наћи у интервалу

$$\left(m - \alpha\sigma;\ m + \alpha\sigma\right]$$

при чему је α произвољно одабрани број,

$$P\{m - \alpha\sigma < X \leq m + \alpha\sigma\} \geq 1 - \frac{1}{\alpha^2} \qquad (5.69)$$

Специјално

$$P\{m - 2\sigma < X \leq m + 2\sigma\} \geq 0.75$$

$$P\{m - 3\sigma < X \leq m + 3\sigma\} \geq 0.8889$$

шо представља тзв. „**правило две (односно 3) сигме**”.

5.3. ЗАДАЦИ

1. Случајна променљива X има закон вероватноћа

$$f(x)=\begin{cases}\dfrac{1}{(x+1)^2} & \text{за } x>0\\ 0 & \text{за } x\le 0.\end{cases}$$

Нека су дати следећи догађаји:

$$\mathrm{A}_1=(-\infty<\mathrm{X}<0)$$
$$\mathrm{A}_2=(0\le\mathrm{X}\le\infty)$$
$$\mathrm{A}_3=(0\le\mathrm{X}\le 1)$$
$$\mathrm{A}_4=(-6\le\mathrm{X}\le 0)$$

Наћи вероватноће следећих случајних догађаја:

a) $\overline{A}_1$; *d)* $A_3\cup A_1$; g) $(\mathrm{A}_3\cup\mathrm{A}_4)(\mathrm{A}_1\cup\mathrm{A}_2)$;

b) $\overline{A}_1\cup A_2$; *e)* $A_1\cup A_2$; h) $\overline{\mathrm{A}}_3$;

c) $A_2\cup A_3$; *f)* $A_1\cup\overline{A}_4$; i) $\overline{\mathrm{A}_1\cup\mathrm{A}_3}$.

2. За случајну променљиву X чији закон вероватноћа је

$$f(x)=\begin{cases}e^{-x} & \text{за } x>0\\ 0 & \text{за } x\le 0,\end{cases}$$

одредити функцију расподеле. Нацртај *f(x)* и функцију расподеле, а затим одреди вероватноћу случајних догађаја датих у задатку 1.

3. Аутоматски струг исеца кружну лимену плочицу. Дијаметар исечене плочице је случајна променљива са законом вероватноћа

$$f(x)=\begin{cases} c(x-0.24)^2(x-0.26)^2 & за\ 0.24\le x\le 0.26 \\ 0 & за\ x<0.24 \quad x>0.26. \end{cases}$$

а) Одредити константу c.

б) Плочице се морају дорађивати ако је њихов дијаметар одступио од прописане вредности 0.25 за више од 0.008. Колики се може очекивати проценат плочица за дораду?

4. Баца се новчић све док не падне писмо. Одреди расподелу случајне променљиве која представља број бацања новчића.

5. Играч A има 2 динара, а играч B 1 динар. Један од играча баца новчић, а други погађа страну новчића која ће пасти у бацању. Уколико погоди, добија динар од другог играча. У противном, плаћа динар. Игра се завршава када један од играча добије сва 3 динара. Одреди расподелу случајне променљиве која представља број бацања новчића до завршетка игре.

6. Машина производи ексере са 1% дефектних у укупној производњи. Како изгледа расподела броја неисправних ексера у узорку од 60 комада.

7. Правилна коцка се баца све док се не појави 6. Одреди вероватноћу да ће се коцка бацати више од пет пута.

8. Закон вероватноћа случајне променљиве X је функција

$$f(x)=\begin{cases} cx^3 & за\ 0<x<1 \\ 0 & за\ x\le 0; x\ge 1 \end{cases}$$

а) Одредити константу c.

б) Одреди ону вредност X за коју ће вероватноћа да случајна променљива X буде већа од ње, бити једнака вероватноћи да ће X бити мање од те вредности.

в) Одреди број A тако да са вероватноћом 0.05 тврдимо да ће случајна променљива бити већа од A.

9. Случајна променљива X има тзв. ***Гама расподелу*** дату законом вероватноћа

$$f(x)=\begin{cases}\dfrac{\lambda^{\beta}}{\Gamma(\beta)}\,x^{\beta-1}e^{-\lambda x} & \text{за } x\geq 0\\ 0 & \text{за } x<0\end{cases}$$

Испитај функцију $f(x)$ и провери да је то заиста функција која представља закон вероватноћа.

10. На основу модела из задатка бр. 9.

а) за $\beta = 640,\ \lambda = 20,$ провери да је

$$P(X = 30) = 0.057$$

б) Ако је $\lambda = 20,$ а вероватноћа

$$P(X \leq 30) = 0.01$$

колико треба да је β *?*

11. Одреди параметре расподеле случајне величине X са законом вероватноћа

$$f(x)=ce^{-\frac{|x-a|}{b}},\ b>0$$

12. Апсолутна вредност брзине кретања молекула је случајна променљива са тзв. *Maksvel*-овом расподелом датом функцијом

$$f(x)=\frac{4x^2}{a^3\sqrt{\pi}}e^{-\frac{x^2}{a^2}}\quad \text{за } x>0$$

Наћи средњу брзину молекула и њену варијансу.

13. Случајна променљива X може узети само две вредности: 1 и 2. Њена очекивана вредност је 3/2. Одредити закон верованотће за X.

14. Правилан новчић се баца 10 пута. Одреди очекивање броја појављивања „писма”, функцију генератрису за тај број, а затим одреди параметре расподеле из функције генератрисе.

15. Ако је функција генератриса дефинисана преко очекиване вредности функције

$$Y = e^{t\,(X-m)}$$

шта ће бити k-ти извод те функције у тачки t=0?

16. Провери да случајна величина *X*, која има тзв. *Cauchy*-јеву *расподелу* дату законом вероватноћа

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

нема очекивану вредност.

17. Да ли, за ***Cauchy*-јеву расподелу**, постоји функција генератриса?

18. Одреди функцију генератрису за расподелу дату законом вероватноћа

$$f(x) = \begin{cases} x & за\ 0 < x < 1 \\ -x+2 & за\ 1 < x < 2 \\ 0 & за\ x \le 0,\ x \ge 2 \end{cases}$$

19. Случајна величина *X* има тзв. *Лог-нормалну расподелу* дату законом вероватноћа

$$f(x) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\alpha)^2}{2\sigma^2}} & за\ x > 0 \\ 0 & за\ x \le 0 \end{cases}$$

Наћи очекивану вредност и варијансу ове случајне променљиве.

20. Ако *X* има тзв. ***Нормалну расподелу*** дату законом вероватноће

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$$

одреди њену функцију генератрису, а затим, из ње, одреди параметре случајне промењиве *X*.

21. Одреди параметре расподеле случајне величине *X* са законом вероватноћа

$$f(x)=ce^{-\frac{|x-a|}{b}},\ b>0$$

22. Апсолутна вредност брзине кретања молекула је случајна промењива са тзв. ***Maksvel*-овом расподелом** датом функцијом

$$f(x)=\frac{4x^2}{a^3\sqrt{\pi}}e^{-\frac{x^2}{a^2}}\quad za\ x>0$$

Наћи средњу брзину молекула и њену варијансу.

23. Случајна промењива *X* може узети само две вредности: 1 и 2. Њена очекивана вредност је 3/2. Одредити закон верованотће за *X*.

24. Правилан новчић се баца 10 пута. Одреди очекивање броја појављивања „писма”, функцију генератрису за тај број, а затим одреди параметре расподеле из функције генератрисе.

25. Ако је функција генератриса дефинисана преко очекиване вредности функције

$$Y=e^{t\,(X-m)}$$

шта ће бити k-ти извод те функције у тачки t=0?

26. Провери да случајна величина *X*, која има тзв. *Caychy-јеву расподелу* дату законом вероватноћа

$$f(x)=\frac{1}{\pi}\frac{1}{1+x^2}$$, нема очекивану вредност.

27. Да ли, за Caychy-јеву расподелу, постоји функција генератриса?

28. Одреди функцију генератрису за расподелу дату законом вероватноћа

$$f(x)=\begin{cases} x & за\ 0<x<1 \\ -x+2 & за\ 1<x<2 \\ 0 & за\ x\le 0,\ x\ge 2 \end{cases}$$

29. Случајна величина X има тзв. ***Лог-нормалну расподелу*** дату законом вероватноћа

$$f(x)=\begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(\ln x-\alpha)^2}{2\sigma^2}} & за\ x>0 \\ 0 & за\ x\le 0 \end{cases}$$

Наћи очекивану вредност и варијансу ове случајне променљиве.

30. Ако X има тзв. *Нормалну расподелу* дату законом вероватноће

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$$

одреди њену функцију генератрису, а затим, из ње, одреди параметре случајне променљиве X.

6. ВИШЕДИМЕНЗИОНАЛНЕ СЛУЧАЈНЕ ПРОМЕНЉИВЕ

Основу статистичке анализе чини онај део Теорије вероватноће који разматра вишедимензионалне случајне промењиве. Зато ћемо, укратко, у овој глави изложити најосновније појмове који ће се касније користити. Ради лакшег разумевања посматраћемо прво, дводимензионалне случајне промењиве и користити се геометријским интерпретацијама и објашњењима неких резултата. Тако прихваћени основни појмови за две димензије, омогућиће да се доста лако схвате уопштења и на вишедимензионалне случајеве.

6.1. РАСПОДЕЛЕ ВИШЕДИМЕНЗИОНАЛНИХ СЛУЧАЈНИХ ПРОМЕНЉИВИХ

Посматрајмо неки статистички скуп. На пример, скуп држава у Европи, скуп предузећа у Србији, и сл. Код сваког елемента статистичког скупа посматраћемо два обележја. Означићемо их са X и Y.

На пример, код предузећа, посматраћемо следећа обележје:

- X - број запослених;
- Y - остварени доходак.

Слично можемо урадити и за земље Европе и посматрати обележја:

- X – извоз,
- Y – увоз.

Код сваког елемента посматраног статистичког скупа, обележја X и Y ће узети одређене вредности. Тако ће се, за статистички скуп од H елемената, добити скуп парова вредности обележја X и Y. Рећи ћемо да тај скуп вредности одређује расподелу дводимензионалног обележја (X,Y) на посматраном скупу.

Да би потпуније одредили расподелу овог обележја (X,Y), прво ћемо претпоставити да су обележја X и Y прекидног типа.

Нека је

$$\{x_1, x_2, \ldots, x_n\}$$

скуп могућих вредности X, при чему је

$$x_i < x_{i+1}$$

за свако $i \in \{1,2,3,\ldots,n-1\}$. Поред тога, нека је скуп могућих различитих вредности обележја Y скуп

$$\{y_1, y_2, \ldots, y_m\}$$

при чему је

$$y_j < y_{j+1}$$

за свако $j \in \{1,2,3,\ldots,m-1\}$.

Означићемо са f_{ij} број елемената статистичког скупа код којих обележје X има вредност x_i, а обележје Y вредност y_j. Сад се све вредности посматраног обележја (X,Y) могу приказати у виду следеће дводимензионалне табеле:

Y / **X**	y_1	y_2	...	y_m
x_1	f_{11}	f_{12}	...	f_{1m}
x_2	f_{21}	f_{22}	...	f_{2m}
⋮				
x_n	f_{n1}	f_{n2}	...	f_{nm}

Табела 6.1. Расподела обележја (X,Y)

Табела 6.1. представља расподелу апсолутних фреквенција дводимензионалног обележја (*X*,*Y*) на посматраном статистичком скупу. Бројеви унутар табеле представљају фреквенције и њихов збир је једнак броју елемената статистичког скупа. Зато је

$$\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij} = N$$

Кад се апсолутне фреквенције, у Табели 6.1, поделе са *N*, добићемо табелу расподела релативних фреквенција обележја (*X*,*Y*):

Y / **X**	y_1	y_2	...	y_m
x_1	p_{11}	p_{12}	...	P_{1m}
x_2	p_{21}	p_{22}	...	P_{2m}
⋮				
X_n	p_{n1}	p_{n2}	...	p_{nm}

Табела 6.2. Расподела обележја (X,Y)

При томе смо означили са

$$p_{ij} = \frac{f_{ij}}{N}$$

за свако $i \in \{1,2,3,\ldots,n\}$ и $j \in \{1,2,3,\ldots,m\}$.

Јасно је да релативне фреквенције у Табели 6.2, задовољавају следећа два услова:

$$p_{ij} \geq 0 \tag{6.1}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = 1 \tag{6.2}$$

за свако $i \in \{1,2,3,\ldots,n\}$ и $j \in \{1,2,3,\ldots,m\}$,

Рећи ћемо да је Табелом 6.2. дата расподела дводимензионалног обележја (X,Y), на посматраном статистичком скупу.

Закон вероватноћа дводимензионалних случајних промењљивих

Сад ћемо посматрати следећи експеримент: из статистичког скупа бирамо на случајан начин један елеменат и утврђујемо код њега вредности обележја *X* и *Y*. Нека сваки елеменат скупа има подједнаку вероватноћу да буде изабран. Означићемо са *X* вредност првог посматраног обележја код изабраног елемента, а са *Y* вредност другог обележја. Тада, сваки пар бројева (x_i, y_j) за $i \in \{1,\ldots,n\}$ и $j \in \{1,\ldots,m\}$,

има одговарајућу вероватноћу да то буде вредност посматраних обележја код изабраног елемента. Другим речима, *X* и *Y* могу имати вредности

$$X = x_i \qquad Y = y_j$$

и при томе је вероватноћа да се то реализује, једнака p_{ij}, тј.

$$P\{X = x_i,\ Y = y_j\} = p_{ij} \tag{6.3}$$

за свако $i \in \{1,\ldots,n\}$ и $j \in \{1,\ldots,m\}$.

У овом **експерименту,** пар променљивих (X,Y), представља **дводимензионалну случајну промењиву**. Скуп тачака (x_i, y_j) је скуп тачака у равни, а (X,Y) може, у једном „извођењу" експеримента бити само једна тачка, и то са одговарајућом вероватноћом.

Можемо уместо статистичког скупа и овог експеримента везаног за статистички скуп, посматрати било какав експеримент чији ће резултат бити пар променљивих (X,Y). На пример, ако се бацају две коцке истовремено и бележи број тачкица на првој коцки X, а број тачкица на другој Y. Тада је резултат експеримента један од парова вредности

(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)

(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)

(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)

(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)

(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)

(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)

Или, можемо на равни XOY бирати на случајан начин једну тачку, тако што бацамо бесконачно малу куглицу и бележимо координате тачке у којој ће се куглица зауставити. Тада је свака тачка (x,y) равни XOY могући резултат експеримента. У оваквим експериментима може се унапред одредити скуп могућих вредности једне променљиве и скуп могућих вредности друге променљиве. Скуп парова ових могућих вредности представљаће скуп могућих резултата дводимензионалне променљиве (X,Y). У једном извођењу експеримента, променљива (X,Y) ће узети само један пар могућих вредности.

Узимање било ког пара могућих вредности променљиве (X,Y), представља случајни догађај са одређеном вероватноћом реализације. Пар променљивих (X,Y), оваквог типа, представља дводимензионалну случајну промењиву.

Слично као и код једнодимензионалних променљивих, посматраћемо две категорије: прекидне (дискретне) и непрекидне (континуиране) дводимензионалне случајне променљиве.

Нека је скуп могућих резултата променљиве X скуп реалних бројева

$$\{x_1, x_2, ...\}$$

а скуп могућих резултата променљиве Y је скуп

$$\{y_1, y_2, ...\}$$

при чему су то коначни скупови бројева, или бесконачни, али пребројиви.

Тада је скуп могућих резултата експеримента скуп парова вредности

$$(x_1, y_1), (x_1, y_2), ...$$

$$(x_2, y_1),\ (x_2, y_2), \ldots$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

а то истовремено скуп могућих вредности дводимензионалне променљиве (X,Y).

Да би се могли проучавати овакви експерименти потребно је одредити вероватноће за сваки могући резултат, односно, потребно је одредити расподелу променљиве (X,Y).

Означимо са p_{ij} вероватноћу да ће резултат експеримента бити пар (x_i, y_j), тј. p_{ij} је вероватноћа случајног догађаја

$$D = \{X = x_i,\ Y = y_j\} \tag{6.4}$$

Може се написати у облику

$$p_{ij} = P\{X = x_i, Y = y_i\} \tag{6.5}$$

за свако $i \in \{1,2,\ldots\}$ и $j \in \{1,2,\ldots\}$.

Ове вероватноће можемо дати у виду следеће табеле:

Y / **X**	y_1	y_2	...	y_j	...
x_1	p_{11}	p_{12}	...	p_{1j}	...
x_2	p_{21}	p_{22}	...	p_{2j}	...
⋮					
x_i	p_{i1}	p_{i2}	...	p_{ij}	...
..	...	...	...	...	...

Табела 6.3. Закон вероватноће случајне променљиве (X,Y)

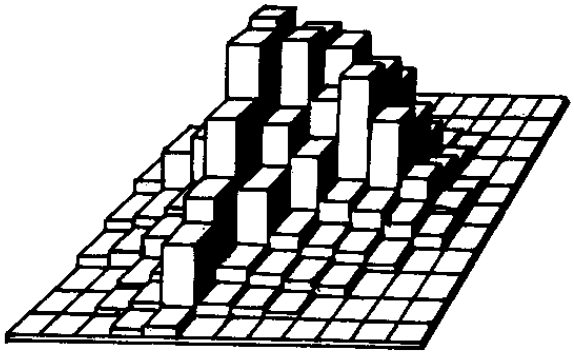
Табела 6.3. одређује закон вероватноћа дводимензионалне случајне променљиве (X,Y) прекидног типа. При томе су вероватноће садржане у табели такве да је

$$p_{ij} \geq 0 \tag{6.6}$$

за свако $i \in \{1,2,\ldots\}$ и $j \in \{1,2,\ldots\}$, а њихов збир је једнак

$$\sum_{i=1}\sum_{j=1} p_{ij} = 1 \tag{6.7}$$

На основу закона вероватноћа датог Табелом 2.3, могуће је одредити вероватноћу било ког случајног догађаја D, везаног за овај експеримент. Закон вероватноћа променљиве (X,Y) прекидног типа, може се приказати графички као на Слици 6.1.



Слика 6.1. Закон вероватноћа дводимензионалне случајне променљиве прекидног типа.

Око сваке тачке $\left(x_i, y_j\right)$ конструише се паралелепипед, чија ће запремина бити једнака вероватноћи p_{ij}, а њихова укупна запремина једнака је јединици.

На пример, ако бацамо истовремено две коцке, онда је закон вероватноћа дводимензионалне случајне променљиве (*X*,*Y*), која представља број тачкица на првој и другој коцки, дат табелом:

Y / **X**	1	2	3	4	5	6
1	p	p	P	p	p	p
2	P	p	p	p	p	p
3	P	p	p	p	p	p
4	P	p	p	p	p	p
5	P	p	p	p	p	p
6	P	P	p	p	p	p

Табела 6.4. Пример расподеле (X,Y)

при чему је $p = 1/36$. Ако нас интересује вероватноћа да ће збир тачкица бити једнак 8, онда је то

$$P = P\{X=2,Y=6\} + P\{X=3,Y=5\} + P\{X=4,Y=4\} + \\ + P\{X=5,Y=3\} + P\{X=6,Y=2\}$$

односно, $P = 6p = 1/6$.

На сличан начин, из Табеле 2.4, могли би одредити вероватноће догађаја:

- Збир тачкица је паран број;
- Збир тачкица није мањи од 5;
- Количник броја тачкица је цео број, и сл.

Другој категорији случајних променљивих припадају променљиве непрекидног типа.

Нека је могућа вредност за променљиву X било који број из коначног или бесконачног интервала (a,b). Поред тога, нека је интервал (c,d) скуп могућих вредности променљиве Y. Тада је свака тачка правоугаоника, одређеног са ова два интервала, могућа вредност дводимензионалне променљиве (X,Y).

Интервал (a,b) поделићемо у подинтервале

$$(a_0,a_1],(a_1,a_2],\ldots,(a_{n-1},a_n]$$

тако да је $a_0 = a,\ a_n = b.$

Слично ћемо урадити и за интервал (c,d). Поделићемо га на подинтервале

$$(c_0,c_1],(c_1,c_2],\ldots,(c_{m-1},c_m]$$

тако да је $c_0 = c,\ c_m = d.$

Посматраћемо случајне догађаје

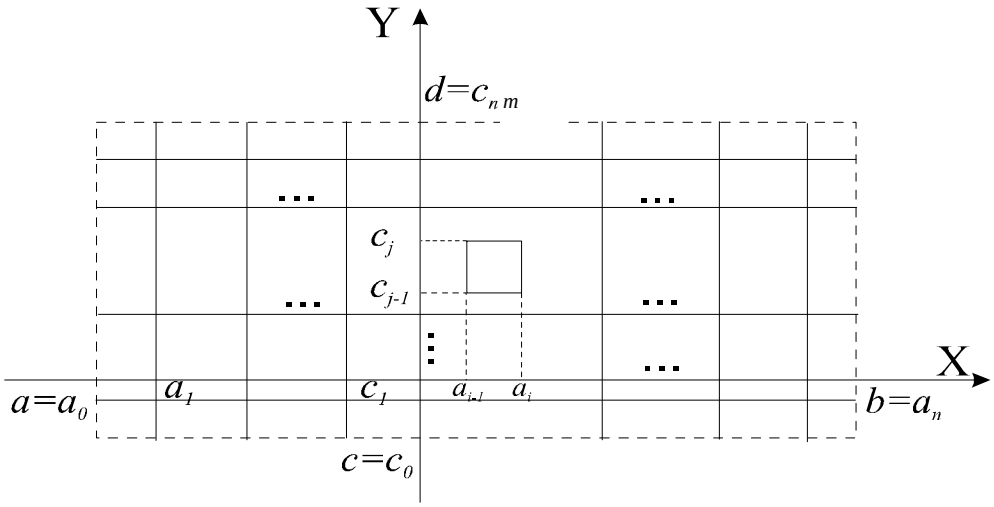
$$D_{ij}=\{a_{i-1}<X \le a_i;\ c_{j-1}<Y \le c_j\}$$

за $i\in\{1,2,\ldots,n\}$ и $j\in\{1,2,\ldots,m\}$. Уствари, D_{ij} су догађаји да ће резултат експеримента бити тачка у правоугаоницима датим на Слици 6.2.

Означимо њихове вероватноће са p_{ij}. Тако ћемо имати

$$p_{ij}=\{a_{i-1}<X \le a_i;\ c_{i-1}<Y \le c_i\}$$

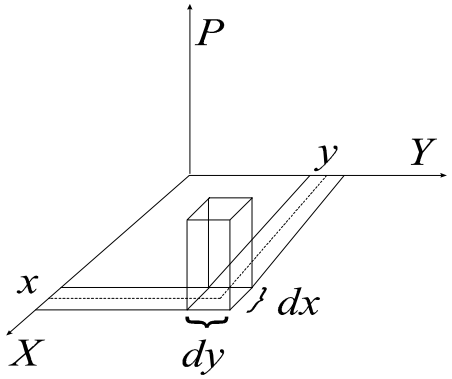
за свако $i\in\{1,2,\ldots,n\}$ и $j\in\{1,2,\ldots,m\}$.



Слика 6.2. Подела правоугаоника на подинтервале.

Подела интервала (a,b) и (c,d) може се вршити на произвољне начине. Поставља се питање, шта ће се десити ако подела интервала (a,b) буде таква да се око сваке тачке $x\in(a,b)$ посматра подинтервал, произвољно мале дужине *dx*, а интервал (c,d) поделимо тако да око сваке тачке $y\in(c,d)$ конструишемо подинтервале произвољно мале

дужине *dy*. Тако ћемо, око сваке тачке (x, y), имати правоугаоник дужине *dx* и ширине *dy* (Слика 6.3).



Слика 6.3. Расподела (X,Y).

Посматраћемо догађај да ће резултат експеримента бити тачка у овом правоугаонику, тј. догађај

$$D = \left\{ x - \frac{dx}{2} < X \le x + \frac{dx}{2};\ y - \frac{dy}{2} < Y \le y + \frac{dy}{2} \right\} \qquad (6.8)$$

Ако постоји функција $\phi(x, y)$ таква да је гранична вредност количника

$$\lim_{\substack{dx\to 0 \\ dy\to 0}} \frac{P\{D\}}{dx \cdot dy} = f(x, y), \qquad (6.9)$$

тада кажемо да је $f(x, y)$ закон вероватноћа дводимензионалне случајне промењиве (X,Y). При том је вероватноћа догађаја D једнака

$$P\left\{ x - \frac{dx}{2} < X \le x + \frac{dx}{2};\ y - \frac{dy}{2} < Y \le y + \frac{dy}{2} \right\} = f(x,y)dxdy\,. \qquad (6.10)$$

Функција $f(x, y)$ је закон вероватноћа или функција густине дводимензионалне случајне промењиве (X,Y), а вероватноћа

$$f(x,y)dx\,dy \tag{6.11}$$

назива се ***елементарна вероватноћа****. Она представља запремину паралелограма око тачке (x,y), као што је приказано на Слици 6.3.*

Ако за све тачке (x, y) у равни XOY, „саберемо" одговарајуће вероватноће, добиће се

$$\int_a^b \int_c^d f(x,y)dx\,dy = 1 \tag{6.12}$$

Посматрајмо неки случајни догађај D везан за овај експеримент. Тада догађају D одговара један скуп тачака S у равни XOY. Вероватноћа догађаја D биће одређена преко интеграла

$$P(D) = \int_S \int f(x,y)\,dx\,dy \tag{6.13}$$

Најчешће, је догађај D скуп тачака у равни, одређен интервалима (x_1, x_2) i (y_1, y_2), тј.

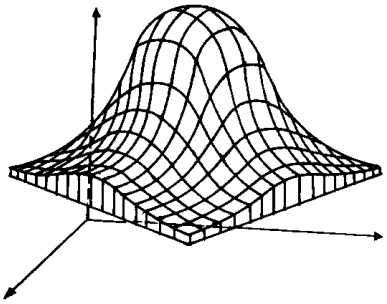
$$D=\{x_1 < X \le x_2; y_1 < Y \le y_2\}$$

Вероватноћа овог догађаја једнака је двоструком интегралу

$$P\{x_1 < X \le x_2;\ y_1 < Y \le y_2\} = \int_{x1}^{x2} \int_{y1}^{y2} f(x,y)\,dx\,dy$$

Геометријски, функција $f(x, y)$ представља неку површину у тродимензионалном простору, као што је на пример површина на Слици 6.4.

Цела запремина испод површине, одређена законом вероватноћа $f(x, y)$ и равни XOY, једнака је јединици. Вероватноћа неког догађаја, одређена је запремином ограниченом површином S у равни XOY и функцијом која представља закон вероватноћа.



Слика 6.4. Закон вероватноћа дводимензионалне случајне промењиве.

Закон вероватноћа k-димензионалне случајне промењиве

Закон вероватноћа дводимензионалне случајне промењиве може се лако уопштити и на вишедимензионалну промењиву.

Претпоставимо да на сваком елементу статистичког скупа посматрамо више обележја $X_1, X_2,\ldots, X_k$. Тада ће, сваком елементу посматраног скупа, одговарати једна тачка у *k*-димензионалном простору.

На сличан начин, у једном експерименту, резултат експеримента може бити само једна тачка *k*-димензионалног простора, при чему је догађај да ће резултат експеримента бити одређена тачка, случајан догађај. Тада кажемо да је $(X_1, X_2,\ldots, X_k)$, вишедимензионална случајна промењива.

Нека резултату експеримента одговара коначан или пребројив скуп тачака у *k*-димензионалном простору. Ако свакој тачки из тог скупа

$$(x_{1i_1}, x_{2i_2},\ldots,x_{ki_k})$$

одговара вероватноћа

$$p_{i_1 i_2 \ldots i_k}$$

при чему је то вероватноћа да ће се реализовати догађај

$$D=\left\{X_1=x_{1i_1},\, X_2=x_{2i_2},...,X_k=x_{ki_k}\right\}$$

tj.

$$P_{i_1 i_2 ... i_k}=P\left\{X_1=x_{1i_1},\, X_2=x_{2i_2},\,...,\, X_k=x_{ki_k}\right\}$$

тада се скуп тачака $(x_{1i_1}, x_{2i_2},..., x_{ki_k})$ *и одговарајућих вероватноћа* $P_{i_1\, i_2\, \ldots\, i_k}$ *назива* ***закон вероватноћа k-димензионалне случајне промењиве*** $(X_1, X_2,..., X_k)$.

Претпоставимо да је резултат експеримента било која тачка у k-димензионалном простору. Нека постоји функција $f(x_1, x_2,..., x_k)$, таква да је

$$\lim_{dx_i\to\infty}\frac{P\left\{x_1-\frac{dx_1}{2}<X_1\le x_1+\frac{dx_1}{2},...,\, x_k-\frac{dx_k}{2}<X_k\le x_k+\frac{dx_k}{2}\right\}}{dx_1, dx_2,..., dx_k}:$$
$$=f(x_1, x_2,..., x_k), \qquad (6.14)$$
i = 1,...., k

за сваку тачку $(x_1, x_2,..., x_k)$. *Тада се функција* $f(x_1, x_2,..., x_k)$ *назива законом вероватноће вишедимензионалне случајне промењиве. При томе је*

$$P\left\{x_1-\frac{dx_1}{2}<X_1\le x_1+\frac{dx_1}{2}, x_2-\frac{dx_2}{2}<X_2\le x_2+\frac{dx_2}{2},...,x_k-\frac{dx_k}{2}<X_k\le x_k+\frac{dx_k}{2}\right\} \qquad (6.15)$$
$$=f(x_1,x_2,...,x_k)dx_1...dx_k$$

тј. вероватноћа да ће се тачка $(X_1, X_2,..., X_k)$, као резултат експеримента, наћи у бесконачно малој околини тачке $(x_1, x_2,..., x_k)$ једнака је производу вредности функције $f(x_1, x_2,..., x_k)$ у тој тачки, и одговарајућих диференцијала $dx_1, dx_2,..., dx_k$.

Маргинални закони вероватноћа

У проучавањима и примени вишедимензионалних променљивих, потребно је, често, посматрати само једну компоненту променљиве, без обзира на вредности осталих компоненти, тј. потребно је посматрати маргиналне промене једне компоненте вишедимензионалне променљиве.

За овакве анализе треба одредити тзв. маргинални закон вероватноћа. Нека је (X,Y) дводимензионална случајна променљива прекидног типа са законом вероватноће:

Y / **X**	y_1	y_2	...
x_1	p_{11}	p_{12}	...
x_2	p_{21}	p_{22}	...
...	...	...	...

Табела 6.5. Закон вероватноћа променљиве (X, Y)

Одредићемо маргинални закон вероватноћа случајне променљиве X. Из Табеле 6.5, види се да X узима вредности из низа бројева $\{x_1, x_2, \ldots\}$.

Потребно је да одредимо вероватноће са којима X узима поједине вредности из овог низа. Догађај

$$D = \{X = x_1\}$$

може се изразити преко уније дисјунктних догађаја

$$D_1 = \{X = x_1, Y = y_1\},\ D_2 = \{X = x_1, Y = y_2\}, \ldots$$

тј.

$$\{X = x_1\} = \{X = x_1, Y = y_1\} \cup \{X = x_1, Y = y_2\} \cup \ldots \qquad (6.16)$$

На основу аксиома којима је дефинисана вероватноћа, следи да је вероватноћа догађаја на левој страни израза (6.16), једнака збиру вероватноћа догађаја на десној страни, тј.

$$P\{X = x_1\} = \mathrm{P}\{X = x_1, Y = y_1\} + P\{X = x_1, Y = y_2\} + \ldots \qquad (6.17)$$

Ако означимо са $p_{1\bullet}$ вероватноћу догађаја $D = \{X = x_1\}$, онда се из (2.17), добија да је

$$p_{1\bullet} = p_{11} + p_{12} + \ldots$$

Сличним закључивањем добиће се да је

$$p_{2\bullet} = p_{21} + p_{22} + \ldots,$$

при чему је $p_{2\bullet}$ вероватноћа догађаја $D = \{X = x_2\}$, итд.

Због тога је ***маргинални закон вероватноћа случајне променљиве X****, дат са*

X	x_1	x_2	...
п	$p_{1\bullet} = \sum_{j=1} p_{1j}$	$p_{2\bullet} = \sum_{j=1} p_{2j}$	...

(6.18)

при чему су p_{ij} вероватноће дате у Табели 6.5.

Из (6.18) се види да је маргинални закон вероватноћа променљиве X добијен као збир вероватноћа датих по врстама у Табели 6.5.

На сличан начин, из Табеле 6.5, добили би и ***маргинални закон вероватноћа случајне променљиве Y***:

Y	y_1	y_2	...
р	$p_{\bullet 1} = \sum_{i=1} p_{i1}$	$p_{\bullet 2} = \sum_{i=1} p_{i2}$	...

(6.19)

У ствари, маргинални закон вероватноћа промењиве *Y* добија се као збир вероватноћа датих у колонама Табеле 6.5.

При одређивању вероватноћа догађаја везаних за вредности само једне промењиве, служимо се маргиналним законима вероватноћа (6.18) и (6.19).

На пример, посматра се 1.000 домаћинстава и испитује приход и уштеђевина домаћинства. Расподела фреквенција дата је у Табели 6.6. Претпоставимо да на случајан начин бирамо једно домаћинство. Интересује нас колика је вероватноћа да ће изабрано домаћинство бити са приходима мањим од 3.000 динара.

Y / **X**	0	1-500	501-2000	2001-5000	5001-
1001-3000	117	56	41	22	14
3001-5000	72	94	71	25	18
5001-	28	100	102	67	63

Табела 6.6. Расподела прихода (X) и уштеђевине (Y)

Из Табеле 6.6, добија се следећа маргинална расподела за *X*:

X	-1000	1001-3000	3001-5000	5001-
p	0,110	0,250	0,280	0,360

Вероватноћа траженог догађаја је

$$p = 0.110 + 0.250$$

односно $p = 0.360$.

А сада ћемо посматрати дводимензионалну случајну промељиву (X,Y) непрекидног типа са законом вероватоћа $f(x, y)$.

Маргинални закон вероватноћа непрекидне случајне промењиве X *је функција* $f_1(x)$*, добијена преко интеграла*

$$f_1(x)=\int_{-\infty}^{\infty} f(x,y)dy$$

а **маргинални закон вероватноћа променљиве Y** *је функција* $f_2(y)$, *добијена преко интеграла*

$$f_2(x) = \int_{-\infty}^{\infty} f(x,y)dx$$

На пример, ако је закон вероватноћа променљиве (*X*,*Y*) дат функцијом

$$f(x,y) = \begin{cases} e^{-x-y} & \text{za } x>0 \text{ i } y>0 \\ 0 & \text{za ostale vrednosti } x \text{ i } y \end{cases}$$

тада је маргинални закон вероватноће променљиве *X* једнак

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = e^{-x}\int_{0}^{\infty} e^{-y}dy = e^{-x}\left[e^{-y}\right]_0^{\infty}$$

односно,

$$f_1(x) = \begin{cases} e^{-x} & za\ x>0 \\ 0 & za\ x\le 0 \end{cases}$$

На сличан начин ћемо добити и маргинални закон вероватноћа променљиве *Y*

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)dx = e^{-y}\int_{0}^{\infty} e^{-x}dx$$

односно,

$$f_2(y) = \begin{cases} e^{-y} & \text{za } y>0 \\ 0 & \text{za } y\le 0 \end{cases}$$

Посматраћемо, сад, *k*-димензионалну случајну променљиву $(X_1, X_2, \ldots, X_k)$ прекидног типа са законом вероватноћа

$$\left[(X_{1i_1}, X_{2i_2}, \ldots, X_{ki_k});\ p_{i_1\ldots i_k}\right]$$

при чему је

$$p_{i_1 \ldots i_k} = P\left\{ X_1 = x_{1i_1}, \ldots, X_k = x_{ki_k} \right\}$$

за $i_1 \in \{1,2,\ldots\};\ i_2 \in \{1,2,\ldots\}, \ldots, i_k \in \{1,2,\ldots\}$.

Можемо посматрати (*k-1*)-димензионалну промењиву $(X_2, \ldots, X_k)$, и одредити њен закон вероватноћа. То ће бити

$$\left[(x_{2i_2}, \ldots, x_{ki_k});\ p_{\bullet i_2 \ldots i_k} \right]$$

при чему је

$$p_{\bullet i_2 \ldots i_k} = \sum_{i_1=1} p_{i_1 i_2 \ldots i_l}$$

На сличан начин би могли, „занемарујући" било коју компоненету, сабирањем заједничких вероватноћа по одговарајућем индексу, добити одговарајући маргинални закон вероватноћа преосталих компоненти.

Поред тога, можемо „занемаривати" више компоненти. Сабирањем вероватноћа по одговарајућим индексима, добили би маргиналне законе вероватноћа преосталих компоненти.

За непрекидну k-димензионалу случајну промењиву $(X_1, X_2, \ldots, X_k)$ *са законом вероватноћа*

$$f(x_1, x_2, \ldots, x_k)$$

маргинални закон вероватноћа (k-r)-димензионалне случајне промењиве, *добиће се преко интеграла*

$$f^{(r)}(x_1, \ldots, x_{k-r}) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_k) dx_{k-r+1} dx_{k-r+2} \ldots dx_k,$$

за свако $r \in \{1,2,\ldots,(k-1)\}$.

Условни закони вероватноћа

Кад се посматрају вишедимензионалне случајне променљиве, најчешће се посматра узајамна повезаност појединих компоненти случајне променљиве. Прва етапа у проучавањима такве врсте, заснива се на тзв. условним законима вероватноћа.

Посматраћемо, прво, дводимензионалну случајну променљиву (X,Y) прекидног типа. Нека је њен закон вероватноћа дат табелом

Y / **X**	y_1	y_2	...
x_1	p_{11}	p_{12}	...
x_2	p_{21}	p_{22}	...
...	...	...	...

Интересују нас промене случајне променљиве X, под условом да Y има једну одређену вредност, рецимо $Y = y_j$. Треба да одредимо које ће вредности променљиве X узимати и са којим вероватноћама. Означимо са $p_{i/j}$, (читај p_i под условом j), вероватноћу да ће бити $X = x_i$, под условом да је $Y = y_j$.

Посматрајући догађаје $\{X = x_i\}$ и $\{Y = y_j\}$, а на основу дефиниције условне вероватноће случајних догађаја, добиће се тражена вероватноћа

$$p_{i/j} = \frac{P(X = x_i,\ Y = y_j)}{P(Y = y_j)}$$

Са p_{ij} смо означили вероватноћу бројиоца, а са $p_{\bullet j}$ означена је маргинална вероватноћа случајне променљиве Y, тј.

$$p_{ij} = P(X = x_i,\ Y = y_j)$$

$$p_{\bullet j} = P(Y = y_j)$$

тако да је условна вероватноћа

$$p_{i/j} = \frac{p_{ij}}{p_{\bullet j}}$$

односно,

$$p_{i/j} = \frac{p_{ij}}{\sum_{i=1} p_{ij}}$$

Сада можемо написати условни закон вероватноћа у следећем облику

X/Y=y_i	x_1	x_2	...
$p_{i/j}$	$p_{1/j} = \frac{p_{1j}}{p_{\bullet j}}$	$p_{2/j} = \frac{p_{2j}}{p_{\bullet j}}$	...

(6.20)

Оваквих закона вероватноћа биће онолико колико има могућих вредности промењиве *Y*, јер за сваку њену вредност можемо одредити одговарајући условни закон вероватноћа промењиве *X*.

На сличан начин се одређују и условни закони вероватноћа промењиве *Y*, под условом да промењива *X* има једну одређену вредност. Тако ћемо добити следећи условни закон вероватноћа.

Y/X=x_i	y_1	y_2	...
$p_{j/i}$	$p_{1/i} = \frac{p_{i1}}{p_{i\bullet}}$	$p_{2/i} = \frac{p_{i2}}{p_{i\bullet}}$	...

(6.21)

На пример, посматра се променљива (X, Y) са законом вероватноћа

Y / **X**	0,1	1	2
100	0,15	0,05	0,00
200	0,15	0,25	0,05
350	0,05	0,15	0,15

Да би одредили условни закон вероватноћа случајне променљиве Y, под условом да је $X = 350$, прво ћемо одредити маргинални закон вероватноћа за X. Из табеле се добија да је

X	100	200	350
$p_{i\bullet}$	0,20	0,45	0,35

Условни закон вероватноћа биће одређен вероватноћама

$$P(Y = 0.1 / X = 350) = \frac{0.05}{0.35} = 0.1428$$

$$P(Y = 1 / X = 350) = \frac{0.15}{0.35} = 0.4286$$

$$P(Y = 2 / X = 350) = \frac{0.15}{0.35} = 0.4286$$

тако да је тражени закон вероватноћа

Y/X=350	0,1	1	2
$p_{j/i}$	0,1428	0,4286	0,4286

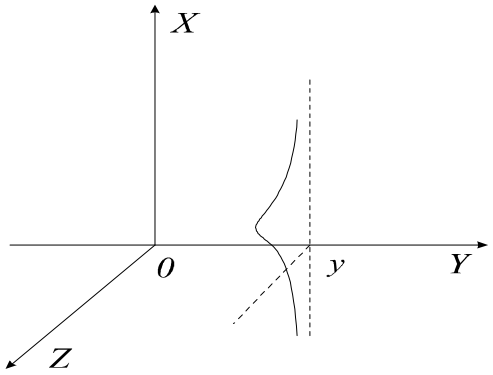
Сада ћемо посматрати дводимензионалну случајну променљиву (X, Y) непрекидног типа. Нека је њен закон вероватноћа дат функцијом

$$f(x,y)$$

а маргинални закони вероватноћа су

$$f_1(x) \;\; i \;\; f_2(y)$$

Претпоставимо да желимо испитати случјну променљиву *X*, под условом да *Y* има једну одређену вредност, рецимо *Y*=y. (Сл. 6.5).



Слика 6.5. Условни закон вероватноћа

Условни закон вероватноћа променљиве *X* биће представљен у равни паралелној *ZOX* равни, која пролази кроз тачку *Y*=y. То ће бити крива одређена једначином

$$f(x/y)$$

Ако је $f_2(y) \neq 0$, *тада је условни закон вероватноћа променљиве X дат функцијом*

$$f(x/y) = \frac{f(x,y)}{f_2(y)} \qquad (6.22)$$

На сличан начин се одређује и условни закон вероватноћа промењиве *Y*, под условом да је *X*=x . Тај закон вероватноћа је дат функцијом

$$f(x/y)=\frac{f(x,y)}{f_2(y)} \tag{6.23}$$

под условом да је $f_2(y)\neq 0$.

Није тешко проверити да функције (6.22) и (6.23) задовољавају следеће услове:

а) $f(x/y)\geq 0, \quad f(y/x)\geq 0$

б) $\int_{-\infty}^{\infty} f(x/y)dx = 1, \quad \int_{-\infty}^{\infty} f(y/x)dy = 1.$

Пример. Нека је закон вероватноћа променљиве (*X*,*Y*) дат функцијом

$$f(x,y)=\begin{cases}\frac{1}{20}x(x+2y) & за \quad 0<x<1,\ 0<y<6\\ 0 & за\ све\ остале\ тачке\end{cases}$$

Маргинални закони вероватноћа се добијају из интеграла

$$f_1(x)=\frac{1}{20}\int_0^6 x(x+2y)dy$$

$$f_2(y)=\frac{1}{20}\int_0^1 x(x+2y)dx$$

Условни закони вероватноћа су дати функцијама

$$f(x/y)=\frac{f(x,y)}{f_2(y)}=\frac{3x(x+2y)}{1+3x}, \quad f(y/x)=\frac{f(x,y)}{f_1(x)}=\frac{x+2y}{6(x+6)}$$

Условни закони вероватноћа могу се лако уопштити и на *k*-димензионалне случајне променљиве. Уопштење се врши аналогно дводимензионалној случајној променљивој, с тим што се посматрају количници заједничких закона вероватноћа и одоговарајућих маргиналних закона вероватноћа.

Независне случајне променљиве

Посматрајмо дводимензионалну случајну променљиву (X,Y). Нека је D_x било какав случајни догађај који је везан само за вредности променљиве X, а D_y случајни догађај који се односи на вредности променљиве Y. Ако је вероватноћа истовременог остварења догађаја D_x и D_y, једнака производу њихових вероватноћа, онда кажемо да су D_x и D_y независни догађаји.

Ако је овај услов задовољен за било која два догађаја D_x и D_y, рећи ћемо да су променљиве X и Y независне случајне променљиве.

Наравно, независност случајних променљивих може бити, на погоднији начин, изражена преко њихових закона вероватноћа.

Нека је дата дводимензионална случајна променљива (X,Y) прекидног типа са законом вероватноћа

Y / **X**	y_1	y_2	...
x_1	p_{11}	p_{12}	...
x_2	p_{21}	p_{22}	...
...	...	...	...

Означили смо маргиналне законе вероватноћа за X и Y са

X	x_1	X_2	...
$p_{i\bullet}$	$p_{1\bullet}$	$p_{2\bullet}$	...

Y	y_1	Y_2	...
$p_{\bullet j}$	$p_{\bullet 1}$	$p_{\bullet 2}$	...

Рећи ћемо да су **X и Y независне случајне променљиве**, *ако је, за свако* $i \in \{1,2,\ldots\}$ *и за свако* $j \in \{1,2,\ldots\}$, *заједничка вероватноћа* p_{ij} *једнака производу одговарајућих маргиналних вероватноћа, тј. ако је*

$$p_{ij} = p_{i\bullet} \cdot p_{\bullet j} \tag{6.24}$$

Посматрајмо, сад, дводимензионалну случајну променљиву (X,Y) непрекидног типа. Означили смо са $f(x,y)$ њен закон вероватноћа, а са $f_1(x)$ *i* $f_2(y)$ маргиналне законе вероватноћа променљивих X и Y.

Рећи ћемо да су **X и Y независне случајне променљиве**, *ако је њихов заједнички закон вероватноћа једнак производу маргиналних закона вероватноћа, тј. ако је*

$$f(x,y) = f_1(x) \cdot f_2(y) \tag{6.25}$$

и то за сваку тачку (x,y) у равни XOY.

Пример. Случајна променљива (X,Y) има закон вероватноћа дат функцијом

$$f(x,y) = \begin{cases} c(2xy + 2x + y + 1) & za\ 0 < x < 1,\ 0 < y < 1 \\ 0 & za\ ostale\ tacke. \end{cases}$$

Да би функција $f(x,y)$ била закон вероватноћа, потребно је да је

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1,$$

односно,

$$\int_{0}^{1}\int_{0}^{1} c(2xy + 2x + y + 1)dxdy = 1.$$

Одавде се добија да је

$$c = \frac{1}{3}$$

Маргинални закони вероватноћа су

$$f_1(x) = \frac{2x+1}{2}$$

$$f_2(y) = \frac{2}{3}(y+1)$$

Види се да је

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

што значи да су X и Y независне случајне промењиве.

Ако су случајне промењиве X и Y међусобно независне, онда су условни закони вероватноћа, $f(x/y)$ *i* $f(y/x)$, једнаки маргиналним законима вероватноћа, тј. онда је

$$f(x/y) = f_1(x)$$

$$f(y/x) = f_2(y)$$

Обрнуто, ако су условни закони вероватноћа једнаки маргиналним, онда су X *и* Y *међусобно независне случајне промењиве.*

Појам независних случајних промењивих може се лако уопштити и на k-димензионалне случајне промењиве.

Нека је $f(x_1, x_2, \ldots, x_k)$ закон вероватноћа промењиве $(X_1, X_2, \ldots, X_k)$, а $f_1(x_1)$, $f_2(x_2)$, ..., $f_k(x_k)$, маргинални закони вероватноћа.

Рећи ћемо да су $X_1, X_2, \ldots, X_k$ *међусобно независне случајне промењиве ако, и само ако је*

$$f(x_1, x_2, \ldots, x_k) = f_1(x_1) \cdot f_2(x_2) \cdot \ldots \cdot f_k(x_k)$$

и то за сваку тачку $(x_1, x_2, \ldots, x_k)$ *k-димензионалног простора.*

Често се посматра зависност једне случајне променљиве Y и k-димензионалне променљиве $(X_1, X_2, \ldots, X_k)$.

Нека је њихов заједнички закон вероватноћа функција

$$f(y, x_1, x_2, \ldots, x_k)$$

а маргинални закони вероватноћа су

$$f_1(y) \quad i \quad f_2(x_1, x_2, \ldots, x_k)$$

Рећи ћемо да су Y и $(X_1, X_2, \ldots, X_k)$ *независне ако и само ако је*

$$f(y, x_1, x_2, \ldots, x_k) = f_1(y) \cdot f_2(x_1, x_2, \ldots, x_k)$$

6.2. ПАРАМЕТРИ РАСПОДЕЛА ВИШЕДИМЕНЗИОНАЛНИХ СЛУЧАЈНИХ ПРОМЕНЉИВИХ

Комплетне карактеристике једне случајне промењиве садржане су у њеном закону вероватноћа, чије познавање је довољно за добијање свих потребних информација о случајној промењивој, њеним компонентама и карактеристикама вишедимензионалне случајне промењиве.

Међутим, у пракси најчешће нисмо у стању да одредимо тачан закон вероватноћа, па самим тим не можемо доћи до потребних информација о случајној промењивој.

Поред тога често је закон вероватноћа одређен функцијама које зависе од једног или више параметара. За конкретну анализу потребно је, пре свега, утврдити вредности тих параметара. Такав је случај са најчешће коришћеном тзв. Нормалном расподелом код које је потребно претходно утврдити средине, варијансе и коваријансе посматране вишедимензионалне случајне промењиве.

И на крају, чак и у ретким случајевима, кад нам је потпуно познат закон вероватноћа, потребна је прилично компликована математичка апаратура за једноставну анализу.

Зато случајне промењиве и њихове расподеле морамо проучавати преко одређеног броја нумеричких карактеристика, којима ћемо у довољној мери моћи да опишемо посматрану случајну промењиву. Као и код једнодимензионалних случајних промењивих, и овде је потребно прво одредити *Очекиване вредности* случајних промењивих и њихових функција.

Очекивана вредност

Очекивана вредност неке функције случајне променљиве дефинише се као и код једнодимензионалних променљивих, преко вредности те функције и закона вероватноћа.

Тако је, на пример, код дводимензионалне случајне променљиве (X,Y) *са законом вероватноћа*

Y / **X**	y_1	y_2	...
x_1	p_{11}	p_{12}	...
x_2	p_{21}	p_{22}	...
⋮	⋮	⋮	⋮⋮⋮

очекивана вредност функције $g(X,Y)$ *једнака*

$$E\{g(X,Y)\} = \sum_i \sum_j g(x_i, y_j)\, p_{ij} \tag{6.26}$$

За непрекидне случајне променљиве (X,Y), *са законом вероватноћа* $f(x,y)$, *очекивана вредност функције* $g(X,Y)$ *дата је интегралом*

$$E\{g(X,Y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dxdy \tag{6.27}$$

Посматрајмо, сад, *k*-димензионалну случајну променљиву $(X_1, X_2, ..., X_k)$. Нека је то променљива непрекидног типа са законом вероватноћа.

$$\left[x_{1i_1}, x_{2i_2}, ..., x_{ki_k};\ p_{i_1\, i_2 \ldots i_k} \right]$$

при чему је

$$i_1 \in \{1,2,...\}, ..., i_k \in \{1,2,...\}$$

Очекивана вредност функције $g(X_1, X_2, ..., X_k)$ *једнака је вишеструком збиру*

$$E\{g(X_1, X_2, ..., X_k)\} = \sum_{i_1} \sum_{i_2} ... \sum_{i_k} g\left(x_{1i_1}, x_{2i_2}, ..., x_{ki_k}\right) p_{i_1 i_2 ... i_k}$$

Ако је случајна промењива $(X_1, X_2, ..., X_k)$. *непрекидног типа са законом вероватноћа* $f(x_1, x_2, ..., x_k)$, *тада је очекивана вредност функције* $g(X_1, X_2, ..., X_k)$ *једнака вишеструком интегралу*

$$E\{g(X_1, X_2, ..., X_k)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} g(x_1, x_2, ..., x_k) f(x_1, x_2, ..., x_k) dx_1 dx_2 ... dx_k$$

Посматрајмо две функције случајних променљивих

$$g(X_1, ..., X_k) \quad i \quad h(X_1, ..., X_k)$$

које имају очекиване вредности. Тада је очекивана вредност линеарне комбинације ових функција, једнака линеарној комбинацији њихових очекиваних вредности, тј.

$$E\{ag(X_1, ..., X_k) + bh(X_1, ..., X_k)\} =$$
$$= aE\{g(X_1, ..., X_k)\} + bE\{h(X_1, ..., X_k)\}$$

Специјално, очекивана вредност збира или разлике случајних променљивих једнака је збиру или разлици њихових очекиваних вредности.

Ако су компоненте случајне промењиве међусобно *независне, онда је очекивана вредност њиховог производа једнака производу њихових очекиваних вредности*, тј.

$$E\{X_1 X_2 ... X_k\} = E(X_1) \cdot E(X_2) \cdot ... \cdot E(X_k)$$

Код дводимензионалне случајне промењиве (X, Y) очекивана вредност збира, или разлике, једнака је збиру, или разлици, очекиваних

вредности. Ако су *X* и *Y* независне, тада је очекивана вредност њиховог производа једнака производу њихових очекиваних вредности.

Моменти случајних променљивих

У анализи случајних променљивих најчешће се користе очекиване вредности специјалних функција $g(X,Y)$, преко којих се дефинишу тзв. *Моменти*.

Нека су *r* и *s* два ненегативна цела броја.

Обични моменат, реда (r,s), *дводимензионалне случајне променљиве* (X,Y), *представља очекивану вредност функције*.

$$g(X,Y)=X^r \cdot Y^s$$

За непрекидну случајну променљиву, обичан моменат, реда (r,s) биће

$$m_{rs} = E\left\{X^r \cdot Y^s\right\} = \sum_i \sum_j x_i^r y_j^s p_{ij} \tag{6.28}$$

Ако је (X,Y) непрекидна случајна променљива, онда је њен обични момент реда (r,s) дат интегралом

$$m_{rs} = E\left\{X^r \cdot Y^s\right\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s f(x,y)dxdy \tag{6.29}$$

Специјално, обични моменат, реда *(1,0)*, није ништа друго него очекивана вредност случајне променљиве *X*. Означавамо га са

$$m_x = m_{10} = E(X)$$

Слично је за променљиву *Y* обичан моменат, реда *(0,1)* је очекивана вредност случајне променљиве *Y*. Означићемо га са

$$m_y = m_{01} = E(Y)$$

Очекивану вредност за X можемо одредити из маргиналног закона вероватноћа. Она је дата збиром

$$m_x = \sum_i x_i \cdot p_i$$

за прекидну случајну промењиву, односно, интегралом

$$m_x = \int_{-\infty}^{\infty} x f_1(x)dx$$

за непрекидну промењиву X.

На сличан начин се може одредити и очекивана вредност промењиве Y. Биће то збир

$$m_y = \sum_j y_j \cdot p_{\bullet j}$$

односно, интеграл

$$m_y = \int_{-\infty}^{\infty} y f_2(y)dy$$

Тачка (m_x, m_y) у равни XOY, представља *тежиште расподеле* случајне промењиве (X,Y). У ствари, то је она тачка равни XOY око које се „најчешће" групишу поједине вредности промењиве (X,Y).

Другу класу момената случајних промењивих (X,Y) чине тзв. *централни моменти*.

Централни моменат, реда (r,s), *је очекивана вредност функције*

$$g(X,Y) = (X - m_x)^r (Y - m_y)^s$$

За прекидну случајну промењиву (X,Y), централни моменат реда (r,s) одређује се преко збира

$$\mu_{rs} = \sum_i \sum_j (x_i - m_x)^r (y_j - m_y)^s p_{ij} \qquad (6.30)$$

а за непрекидну случајну променљиву централни моменат, реда (r,s), дат је интегралом

$$\mu_{rs} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-m_x)(y-m_y)\, f(x,y)dxdy \qquad (6.31)$$

У пракси се најчешће користе централни моменти реда *(1,0), (0,1), (2,0), (0,2) i (1,1).*

Није тешко проверити да су моменти μ_{10} *i* μ_{01}, једнаки *0.*

Централни моменат реда *(2,0)* није ништа друго до *варијанса случајне променљиве X*. Означићемо га са

$$\sigma_x^2 = \mu_{20} = E(X-m_x)^2$$

Кад се израз у загради квадрира и искористе особине очекиване вредности, варијанса случајне променљиве *X* биће једнака

$$\sigma_x^2 = E\left(X^2\right) - m_x^2$$

односно,

$$\sigma_x^2 = m_{20} - m_x^2$$

при чему је m_{20}, обични моменат реда *(2,0).*

Варијанса случајне променљиве Y је у ствари централни моменат реда *(0,2),* тј.

$$\sigma_y^2 = \mu_{02} = E(Y-m_y)^2$$

Варијанса променљиве *Y,* такође, се може изразити преко обичних момената. Наиме, варијанса променљиве *Y* је једнака

$$\sigma_y^2 = m_{02} - m_y^2$$

при чему је m_{02} обични моменат реда *(0,2).*

У анализи вишедимензионалних променљивих, специјалну улогу има централни моменат реда (1,1). То је тзв. *коваријанса случајних променљивих.*

Коваријанса *случајних променљивих* (X,Y) *дефинише се преко очекиване вредности*

$$\sigma_{xy} = E\left\{(X - m_x)(Y - m_y)\right\}$$

а одређује се преко збира

$$\sigma_{xy} = \mu_{11} = \sum_i \sum_j (x_i - m_x)(y_j - m_y)p_{ij} \tag{6.32}$$

односно, преко интеграла

$$\sigma_{xy} = \mu_{11} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_x)(y - m_y)\, f(x,y)dxdy \tag{6.33}$$

Из (2.32) или (2.33) лако се добија да је коваријанса једнака

$$\sigma_{xy} = \sum_i \sum_j x_i y_j p_{ij} - m_x m_y$$

односно,

$$\sigma_{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x,y)dxdy - m_x m_y$$

тако да је

$$\sigma_{xy} = m_{11} - m_x \cdot m_y$$

при чему је m_{11} обични моменат реда *(1,1).*

Параметри σ_x^2, σ_y^2 *и* σ_{xy} представљају мере варијабилитета дводимензионалне случајне променљиве (X,Y).

Варијанса σ_x^2 је показатељ растурања око средине m_x, вредности промењиве X у правцу осе X, док је варијанса σ_y^2 показатељ растурања промењиве Y, око средине m_y, у праcу осе Y.

Коваријанса σ_{xy} садржи растурања X-а і Y-а појединачно, али исто тако и заједничка растурања промењивих X и Y.

Претпоставимо да су X и Y независне случајне промењиве. Тада је њихов заједнички закон вероватноћа једнак производу одговарајућих маргиналних закона вероватноћа, тј.

$$p_{ij} = p_{i\bullet} \cdot p_{\bullet j}\text{, односно, } f(x,y) = f_1(x) \cdot f_2(y)$$

па коваријансе имају вредности

$$\sigma_{xy} = \sum_i x_i p_{i\bullet} \sum_j y_j p_j - m_x m_y = 0$$

$$\sigma_{xy} = \int_{-\infty}^{\infty} x f_y(x)dx \int_{-\infty}^{\infty} y f_2(y)dy - m_x m_y = 0$$

Сад можемо констатовати следеће:

Ако су X и Y независне случајне промењиве, онда је њихова коваријанса једнака нули.

Са друге стране ако коваријанса σ_{xy} није једнака нули, онда то значи да су растурања промењивих X и Y на неки начин, међусобно повезана. Другим речима, значиће то да међу промењивим X и Y постоји извесна зависност.

Коефицијент корелације

Један од основних проблема анализе вишедимензионалних случајних промењивих је проблем мерења узајамне зависности појединих компоненти случајне промењиве.

Посматрајмо, прво, дводимензионалну случајну промењиву (X,Y). Претпоставимо да њен закон вероватноћа има вредности различите од нуле само у оним тачкама (x,y) за које је y једнозначно одређена функција од x и обрнуто. Другим речима, нека за сваку вредност $X=x$, постоји само једна вредност y, за коју је вероватноћа различита од 0. За такве случајеве је заједнички закон вероватноћа једнак маргиналним законима вероватноћа, тј.

$$f(x,y) = f_1(x) = f_2(y) \tag{6.34}$$

Тада кажемо да су X и Y потпуно (функционално) зависне случајне промењиве.

Претпоставимо сада да је заједнички закон вероватноћа случајне промењиве (X,Y) једнак производу маргиналних, тј.

$$f(x,y) = f_1(x)\,f_2(y) \tag{6.35}$$

У том случају, за сваку вредност $X=x$, вероватноћа појединих вредности промењиве Y одређује се из маргиналног закона вероватноћа $f_2(y)$ и не зависи уопште од вредности x. Другим речима, за сваку вредност x, промене случајне промењиве Y су потпуно независне од те вредности x.

Тада кажемо да су X и Y (стохастички) независне случајне промењиве.

У свим другим случајевима међусобног односа компоненти случајне промењиве (X,Y), кажемо да су X и Y међусобно корелиране промењиве. Другим речима:

Ако закон вероватноћа случајне промењиве (X,Y) нема особине

$$f(x,y) = f_1(x) = f_2(y)$$

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

кажемо да су X и Y међусобно корелиране случајне промењиве.

На пример, у *n* дана се мери обим падавина у два места *A* и *B*. Нека је *X* дневна количина падавина у месту *A*, а *Y* количина падавина у месту *B*. Ако су *A* и *B* веома близу један другом, тада су *X* и *Y* потпуно зависне променљиве, тј. свакој вредности *X* одговараће тачно једна вредност *Y*, и обрнуто.

Ако су *A* и *B* на супротним странама Земље, можемо очекивати да ће *X* и *Y* бити стохастички независне променљиве.

Међутим, када са *A* и *B* удаљени међусобно неколико стотина километара, неће се десити ниједан од претходних екстремних случајева. Тада ћемо посматрати извесну корелациону везу између *X* и *Y*.

Сада се поставља проблем утврђивања једног параметра, који зависи од закона вероватноћа и који представља меру корелационе зависности променљивих *X* и *Y*.

У статистичким испитивањима користи се веома велики број параметара као мера корелационих веза. Ипак, најчешће се користи тзв. ***коефицијент корелације***, који служи као мера линеарне зависности порменљивих *X* и *Y*.

Коефицијент корелације је дефинисан преко варијанси и коваријанси случајних променљивих *X* и *Y*. Видели смо да коваријанса σ_{xy} садржи у себи варијабилитет променљиве *X*, варијабилитет променљиве *Y* и узајамне промене *X* и *Y*.

Да би добили чист показатељ заједничких одступања променљивих *X* и *Y*, потребно је да у коваријанси елиминишемо онај део који је последица појединачних одступања променљиве *X* и променљиве *Y*.

Појединачна одступања променљивих *X* и *Y* меримо стандардним девијацијама σ_x *i* σ_y. Зато ћемо коваријансу поделити са σ_x *i* σ_y.

Тако ћемо добити израз за коефицијент корелације

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{6.36}$$

Овако дефинисани коефицијент корелације поседује следеће особине:

Особина 1. Претпоставимо да су X и Y независне променљиве. Тада је њихова коваријанса једнака нули, па имамо тврђење:

Ако су X и Y независне случајне променљиве, њихов коефицијент корелације ρ *једнак је нули.*

Поставља се, сад, обрнуто питање: **Ако је коефицијент корелације ρ, једнак нули, да ли су X и Y независне случајне променљиве?**

Нажалост, одговор је негативан. Наиме, може се десити да је коефицијент корелације једнак нули, али између променљивих X и Y постоји чак и функционална веза.

На пример, нека је закон вероватноће случајне променљиве (X,Y) дат функцијом

$$f(x,y) = \begin{cases} \dfrac{1}{R^2\pi} & \text{за } x^2 + y^2 \le R^2 \\ 0 & \text{за } x^2 + y^2 > R^2 \end{cases}$$

Тада је коваријанса дата изразом

$$\sigma_{xy} = \int_D \int xy \cdot \frac{1}{R^2\pi} dxdy$$

при чему је D област унутар круга $x^2 + y^2 = R^2$.

У првом и трећем квадранту XOY равни, подинтегрална функција је позитивна, а у другом и четвртом је негативна. Зато је $\sigma_{xy} = 0$, тако да је коефицијент корелације $\rho = 0$.

Међутим, ако је $X = 0$, онда се Y мења униформно у интервалу $(-R; R)$, а ако је $X = R$, Y може узети само вредност *0*, што значи да између X и Y постоји функционална зависност.

Особина 2. Коефицијент корелације има вредност +1 или -1, ако и само аку су X и Y међусобно линеарно зависне случајне промењиве.

Доказ. Прво ћемо претпоставити да су X и Y линеарно зависни. Нека је та зависност изражена функцијом

$$Y = aX + b \tag{6.37}$$

На основу особине очекиване вредности, имаћемо везу

$$m_y = am_x + b \tag{6.38}$$

Коваријанса је једнака

$$\sigma_{xy} = E\left\{(X - m_x)(Y - m_y)\right\}$$

Замењујући (6.37) и (6.38) у горњи израз добићемо

$$\sigma_{xy} = E\left\{(X - m_x)\left[a(X - m_x) + b - b\right]\right\}$$

односно,

$$\sigma_{xy} = aE\left\{(X - m_x)^2\right\} = a\sigma_x^2 \tag{6.39}$$

Са друге стране из (6.37) добијамо

$$X = \frac{1}{a}Y - \frac{b}{a}$$

тако да је коваријанса једнака

$$\sigma_{xy} = E\left\{\left[\frac{1}{a}(Y-m_y)-\frac{b}{a}+\frac{b}{a}\right](Y-m_y)\right\} = \frac{1}{a}E\left\{(Y-m_y)^2\right\}$$

односно,

$$\sigma_{xy} = \frac{1}{a}\sigma_y^2 \tag{6.40}$$

Множећи (6.39) и (6.40), добиће се

$$\sigma_{xy}^2 = \sigma_x^2 \cdot \sigma_y^2$$

односно,

$$\sigma_{xy} = \pm\sigma_x \cdot \sigma_y \tag{6.41}$$

Ако је $a>0$, онда је σ_{xy} позитивно, па је коефицијент корелације

$$\rho = \frac{\sigma_x \cdot \sigma_y}{\sigma_x \cdot \sigma_y} = 1$$

а ако је $a<0$, онда је σ_{xy} негативно, па је

$$\rho = \frac{\sigma_x \cdot \sigma_y}{\sigma_x \cdot \sigma_y} = -1$$

а то је требало доказати.

Сада ћемо претпоставити да коефицијент корелације има вредност

$$\rho = \pm 1$$

У том случају је

$$\sigma_{xy}^2 = \sigma_x^2 \cdot \sigma_y^2$$

односно,

$$\frac{\sigma_{xy}}{\sigma_x^2} = \frac{\sigma_y^2}{\sigma_{xy}}$$

Означимо горњи количник са λ. Добићемо две једначине

$$\sigma_{xy} - \lambda\sigma_x^2 = 0$$

$$\sigma_y^2 - \lambda\sigma_{xy}^2 = 0$$

а то се може изразити преко очекиваних вредности

$$E\left\{(X - m_x)(Y - m_y)\right\} - \lambda E(X - m_x)^2 = 0$$

$$E(Y - m_y)^2 - \lambda E\left\{(X - m_x)(Y - m_y)\right\} = 0$$

Користећи особину очекиване вредности, горње једначине се могу написати у облику

$$E\left\{(X - m_x)(Y - m_y) - \lambda\,(X - m_x)^2\right\} = 0$$

$$E\left\{(Y - m_y)^2 - \lambda\,(X - m_x)(Y - m_y)\right\} = 0$$

Ако прву једначину помножимо са $-\lambda$ и саберемо са другом, добићемо

$$E\left\{(Y - m_y)^2 - 2\lambda\,(X - m_x)(Y - m_y) + \lambda^2\,(X - m_x)^2\right\} = 0$$

односно,

$$E\left\{(Y - m_y) - \lambda(X - m_x)\right\}^2 = 0$$

Горњи израз представља очекивану вредност квадрата једне случајне променљиве. Та очекивана вредност биће једнака нули једино ако је

$$(Y - m_y) - \lambda\,(X - m_x) = 0$$

односно,

$$Y - m_y = \lambda\,(X - m_x)$$

а то значи да између Y и X постоји линерна веза. То је и требало доказати.

Особина 3. *Коефицијент корелација има вредност између -1 и +1.*

Доказ. Посматраћемо израз

$$\left\{t(X - m_x) + (Y - m_y)\right\}^2$$

дефинисан за свако реално *t.* Очекивана вредност овог израза је ненегативна, тј.

$$E\left\{t(X - m_x) + (Y - m_y)\right\}^2 \geq 0$$

Одавде се, после квадрирања, добија да је

$$t^2\sigma_x^2 + 2t\sigma_{xy} + \sigma_y^2 \geq 0$$

што представља квадратни трином са негативном дискриминантом,

$\sigma_{xy}^2 - \sigma_x^2\sigma_y^2 \leq 0$, односно, $\sigma_{xy}^2 \leq \sigma_x^2 \cdot \sigma_y^2$

одакле се добија да је

$$\rho^2 \leq 1$$

а то је и требало доказати.

У пракси се најчешће користи тзв. Нормална расподела. Видећемо касније да за њу важи следеће тврђење:

Кад (X,Y) *има Нормалну расподелу, тада су X и Y независне ако и само ако је њихов коефицијент корелације* $\rho = 0$.

Пример. Из скупа цифара *(1,2,3,4,5)*, на случајан начин се бирају две цифре. Нека је X број непарних бројева, а Y број парних бројева међу изабраним цифрама. Тада је закон вероватноћа (X,Y) дат табелом

X \ **Y**	0	1	2
0	0	0	$\frac{\binom{2}{2}}{\binom{5}{2}}$
1	0	$\frac{\binom{3}{1}\binom{2}{1}}{\binom{5}{2}}$	0
2	$\frac{\binom{3}{2}}{\binom{5}{2}}$	0	0

Варијансе ће имати вредности

$$\sigma_x^2 = \sigma_y^2 = \frac{9}{25}$$

а коваријанса је једнака

$$\sigma_{yx} = -\frac{9}{25}$$

па је коефицијент корелације $\rho = -1$. То значи да су X и Y линеарно зависни. Заиста, између X и Y постоји веза $X + Y = 2$.

Дисперсиона и корелациона матрица

Код k-димензионалне случајне променљиве $(X_1, X_2, \ldots, X_k)$, параметри се одређују на сличан начин као и код дводимензионалне променљиве.

Основни параметри преко којих се могу утврдити извесне особине k-димензионалне случајне променљиве, су следећи:

а) k средњих вредности или, краће, *средина*

$$m_1, m_2, \ldots, m_k$$

б) k варијанси

$$\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$$

в) $k(k-1)$ коваријанси

$$\sigma_{12}, \ldots, \sigma_{1k}, \sigma_{23}, \ldots, \sigma_{2k}, \ldots, \sigma_{(k-1)k}$$

Нека је дата k-димензионлна случајна променљива $(X_1, X_2, \ldots, X_k)$ прекидног типа. Нека је за *i-му* компоненту X_i маргинални закон вероватноћа

X_i	x_{1i}	x_{2i}	...
p_i	p_{1i}	p_{2i}	...

Очекивану вредност компоненте X_i означаваћемо са m_i. То ће бити

$$m_i = E(X_i) = \sum_j x_{ji} p_{ji}$$

Ако је $(X_1, X_2, \ldots, X_k)$ непрекидног типа, онда је очекивана вредност *i-те* компоненте дата интегралом

$$m_i = \int_{-\infty}^{\infty} x_i f_i(x_i) dx_i,$$

при чему је $f_i(x_i)$ одговарајући маргинални закон вероватноћа променљиве X_i.

Тачка *M*, са координатама

$$(m_1, m_2, ..., m_k)$$

представља ***тежиште расподеле*** случајне променљиве $(X_1, X_2, \ldots, X_k)$. То је та тачка у *k*-димензионалном простору око које се „најчешће" групишу поједине тачке целе расподеле посматране променљиве.

За сваку компоненту случајне променљиве $(X_1, X_2, \ldots, X_k)$, можемо одредити њену варијансу преко њеног маргиналног закона вероватноћа. За *i-му* компоненту то ће бити

$$\sigma_i^2 = E(X_i - m_i)^2$$

односно,

$$\sigma_i^2 = E(X_i^2) - m_i^2$$

Посматрајмо две компоненте случајне променљиве $(X_1, X_2, \ldots, X_k)$. *Нека су то* X_i *и* X_j. *Из заједничког закона вероватноћа, можемо одредити дводимензионални маргинални закон вероватноћа за променљиву* (X_i, X_j), *а из њега одредити коваријансу преко очекиване вредности*

$$\sigma_{ij} = E(X_i - m_i)(X_j - m_j)$$

односно,

$$\sigma_{ij} = E(X_i X_j) - m_i m_j$$

На тај начин добили смо укупно $k \cdot k = k^2$ варијанси и коваријанси. Њих можемо написати као елементе квадратне матрице

$$W = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \dots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22}^2 & \dots & \sigma_{2k} \\ \dots & \dots & \dots & \dots \\ \sigma_{k1} & \sigma_{k2} & \dots & \sigma_{kk}^2 \end{bmatrix}$$

Матрица W, назива се дисперсиона матрица случајне променљиве $\left(X_1, X_2, \dots, X_k\right)$.

Пошто је

$$E\left\{\left(X_i - m_i\right)\left(X_j - m_j\right)\right\} = E\left\{\left(X_j - m_j\right)\left(X_i - m_i\right)\right\}$$

онда је

$$\sigma_{ij} = \sigma_{ji}$$

што значи да је дисперсиона матрица симетрична и садржи укупно

$$\frac{k(k+1)}{2}$$

различитих елемената. На дијагонали дисперсионе матрице налазе се варијансе компоненти случајне променљиве $\left(X_1, X_2, \dots, X_k\right)$.

За сваке две компоненте X_i *i* X_j *k*-димензионалне случајне променљиве можемо одредити њихов коефицијент корелације. На основу дефиниције коефицијента корелације, то је количник

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \cdot \sigma_j}$$

Матрица, чији су елементи коефицијенти корелације, има облик

$$R=\begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1k} \\ - & 1 & \dots & \rho_{2k} \\ \cdot & \cdot & \dots & \cdot \\ - & - & \dots & \rho_{(k-1)k} \\ - & - & \dots & 1 \end{bmatrix}$$

и назива се корелациона матрица.

Означимо са D дијагоналну матрицу са стандардним девијацијама компоненти $X_1,\dots,X_k$, на дијагонали.

$$D=\begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & \sigma_k \end{bmatrix}$$

Тада се дисперсиона матрица W може изразити прењко корелационе матрице са

$$W = DRD \tag{6.42}$$

Исто тако се и корелациона матрица R може изразити преко дисперсионе матрице, јер је

$$R = D^{-1}WD^{-1} \tag{6.43}$$

при чему је D^{-1} инверзна матрица матрице D, односно дијагонална матрица код које су на дијагонали реципрочне вредности стандардних девијација случајних променљивих $X_1,\dots,X_k$..

Посматрајмо линеарну комбинацију компоненти случајне променљиве $\left(X_1,X_2,\dots,X_k\right)$ облика

$$t_1\left(X_1-m_1\right)+t_2\left(X_2-m_2\right)+\dots+t_k\left(X_k-m_k\right)$$

Очекивана вредност квадрата ове линеарне комбинације је ненегативан број, тј.

$$E\left\{\sum_{i=1}^{k} t_i \left(X_i - m_i\right)\right\}^2 \geq 0$$

Користећи особине очекиване вредности, после квадрирања добићемо

$$\sum_i \sum_j t_i t_j \sigma_{ij} \geq 0$$

а то представља квадратну форму чија матрица је дисперсиона матрица *W*. Ова квадратна форма је ненегативна једино ако је њена матрица позитивно семидефинитна. То значи да је одговарајућа детермината дисперсионе матрице ненегативна, тј.

$$\det W \geq 0$$

а из релације (2.41) следи да је, такође, детерминанта корелационе матрице ненегативна, тј.

$$\det R \geq 0$$

Ранг дисперсионе и корелационе матрице

Са σ_x^2, σ_y^2, σ_{xy} означили смо варијансе и коваријансе компоненти случајне промењиве (X,Y). Њена дисперсиона матрица је

$$W = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix},$$

а детерминанта је

$$\det W = \sigma_x^2 \sigma_y^2 - \sigma_{xy}^2 \geq 0.$$

У овом случају, ранг матрице *W* може бити *0, 1* или *2*.

Ако је ранг матрице $W = 0$, то значи да су варијансе компоненти X и Y једнаке *0*, тј. $\sigma_x^2 = \sigma_y^2 = 0$.

Тада су маргиналне расподеле обе компоненте такве да је цела маса концентрисана у једној тачки. У дводимензионалном простору је, дакле, цела маса концентрисана у једној тачки. Обрнуто, ако знамо да је цела маса расподеле концентрисана у једној тачки, онда то значи да су варијансе компоненти једнаке *0*, па из израза за детерминанту W следи да је

$$\det W = -\sigma_{xy}^2 \geq 0$$

што значи да је $\sigma_{xy} = 0$. Ранг дисперсионе матрице W је, у овом случају, једнак *0*.

Претпоставимо, сад, да је ранг матрице $W = 1$. Тада ће квадратна форма имати вредност *0* за неко $t = t_0$ и $u = u_0$, а то ће бити могуће једино ако је цела маса расподеле случајне променљиве (X, Y) распоређена дуж праве

$$t_0\left(X - m_x\right) + u_0\left(Y - m_y\right) = 0$$

Обрнуто, ако знамо да је цела маса распоређена дуж праве, онда то мора бити права облика

$$t_0\left(X - m_x\right) + u_0\left(Y - m_y\right) = 0$$

тако да је очекивана вредност њеног квадрата у тачки t_0, u_0, једнака

$$E\left\{\left(t\left(X - m_x\right) + u\left(Y - m_y\right)\right\}\right\} = t^2\sigma_x^2 + 2tu\sigma_{xy} + u^2\sigma_y^2 = 0$$

што значи да је ранг матрице $W = 1$.

На тај начин смо даказали следеће тврђење:

а) Ранг дисперсионе матрице $W=0$ ако, и само ако је укупна маса расподеле случајне променљиве (X,Y) концентрисана у једној тачки XOY равни.

б) Ранг дисперсионе матрице $W=1$ ако, и само ако је укупна маса расподеле случајне променљиве (X,Y) распоређена дуж једне праве, али не у једној тачки.

в) Ранг матрице $W=2$ онда и само онда кад не постоји права дуж које је концентрисана сва маса расподеле случајне променљиве (X,Y).

Наравно, ово тврђење се може уопштити и на *k*-димензионалну случајну променљиву $(X_1, X_2, ..., X_k)$.

Да би ранг дисперсионе матрице W за k-димензионалну случајну променљиву био мањи од k, потребно је, и довољно, да је цела маса расподеле концентрисана на линеарни подпростор димензије мање од k.

Дисперсиона матрица W k-димензионалне случајне променљиве $(X_1, X_2, \ldots, X_k)$, *има ранг једнак k онда, и само онда кад у простору* R_k *не постоји хиперраван која садржи целу масу расподеле случајне променљиве.*

И на крају, напоменимо да ова тврђења, која се односе на ранг дисперсионе матрице *W*, у потпуности важе и за ранг корелационе матрице *R*.

Пример. Нека је у серији од *100* производа $p_1\%$ прве класе, $p_2\%$ друге класе, $p_3\%$ треће и $p_4\%$ четврте класе. Ако на случајан начин бирамо десет производа, тада можемо посматрати четвородимензионалну случајну променљиву (X_1, X_2, X_3, X_4) код које X_i представља број поризвода *i-те* класе у узорку од *10* производа. Ова случајна променљива има закон вероватноћа

$$p(x_1, x_2, x_3, x_4) = \frac{10!}{x_1!\,x_2!\,x_3!\,x_4!}\, p_1^{x1}\, p_2^{x2}\, p_3^{x3}\, p_4^{x4}$$

при чему је $x_i \in \{0,1,2,\ldots,10\}$ за свако $i = 1,2,3,4$. Јасно је да ће бити $X_1 + X_2 + X_3 + X_4 = 10$. Зато ће ова случајна променљива имати дисперсиону матрицу ранга *3*. Очекиване вредности и варијансе компоненти су

$$E(X_1) = 10p_1; \quad E(X_2) = 10p_2; \quad E(X_3) = 10p_3; \quad E(X_4) = 10p_4;$$

$$\sigma_1^2 = 10\,p_1(1-p_1); \quad \sigma_2^2 = 10\,p_2(1-p_2);$$

$$\sigma_3^2 = 10\,p_3(1-p_3); \quad \sigma_4^2 = 10\,p_4(1-p_4);$$

Коваријансе ће имати вредности

$$\sigma_{ij} = -10\,p_i p_j$$

па је дисперсиона матрица једнака

$$W = 10\begin{bmatrix} p_1(1-p_1) & -p_1p_2 & -p_1p_3 & -p_1p_4 \\ -p_1p_2 & p_2(1-p_2) & -p_1p_3 & -p_2p_4 \\ -p_1p_3 & -p_2p_3 & p_3(1-p_3) & -p_3p_4 \\ -p_1p_4 & -p_2p_4 & -p_3p_4 & p_4(1-p_4) \end{bmatrix}$$

одакле је лако закључити да је $\det W = 0$, што значи да је ранг матрице W мањи од 4. Лако је проверити да је то матрица ранга *3*.

6.3. РЕГРЕСИОНА АНАЛИЗА

У свакој научној дисциплини основни проблем је утврђивање веза између променљивих величина. Те везе могу бити потпуно одређене. На пример, у физици се може утврдити тачна функционална зависност између удаљености објеката од земље и гравитационе силе, или између гаса у затвореној посуди и температуре.

Међутим, у биолошким и друштвеним наукама морамо се суочити са много компликованијом ситуацијом. Овде имамо далеко мање разлога да очекујемо откривање тачно одређене везе између променљивих величина. Зато су се, у овим научним дисциплинама, морала користити статистичка изучавања која мере просечне промене једне величине изазване променама друге величине. *Регресиона анализа* управо има за циљ да утврђује и мери везе таквог типа.

Регресионе криве

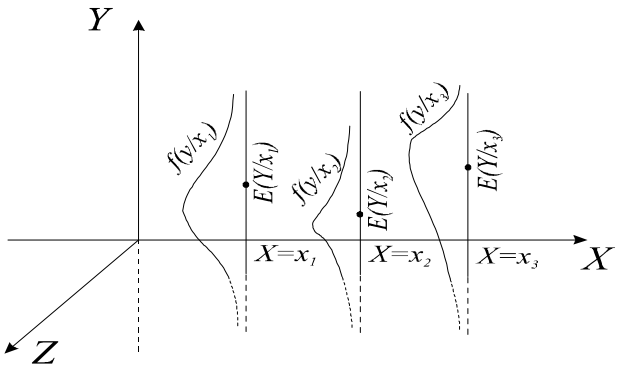
Прво ћемо посматрати дводимензионалну случајну променљиву *(X,Y)* са законом вероватноћа $f(x,y)$. Ако фиксирамо вредност случајне променљиве *X*, тј. ставимо да је *X=x*, онда ће случајна променљива *Y* узимати вредности из интервала $(-\infty;\infty)$ и имаће закон вероватноћа изражен преко условног закона вероватноћа

$$f(y/x)=\frac{f(x,y)}{f_1(x)}$$

као што је приказано на Сл. 6.6.

Можемо сматрати да је *x* „независна” променљива, тако да свакој вредности *x* одговара једна случајна променљива *Y* са својим условним законом вероватноћа $f(y/x)$.

Да би испитали промене случајне променљиве *Y* у зависности од промена *X*, можемо посматрати једну вредност случајне променљиве *Y* као „представника" целе условне расподеле за *Y*. Наравно, то може бити неки од параметара централне тенденције случајне променљиве *Y*, као на пример модус, медијана, аритметичка средина и слично.



Слика 6.6. Условни закони вероватноћа.

У пракси, се углавном, узима као „представник" очекивана вредност случајне променљиве *Y*. Тако ћемо, за сваку вредност *x*, добити одговарајућу очекивану вредност $E(Y/x)$, коју ћемо означити са

$$m_2(x) = E(Y/x)$$

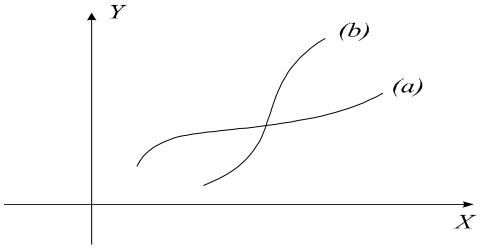
Скуп тачака

$$\left[x;\ y = m_2(x)\right] \tag{6.44}$$

у равни XOY, представљаће криву линију која се назива ***регресиона крива за случајну променљиву Y у односу на X.***

На основу облика регресионе криве могу се утврдити тачне промене средине (очекиване вредности) случајне променљиве *Y* у зависности од промена друге случајне променљиве *X*. То ће нам представљати полазну информацију у анализи промена саме случајне променљиве *Y*.

Регресиона крива, дата скупом тачака (6.44), приказана је на Слици 6.7. кривом (а).



Слика 6.7. Регресионе криве

А сад можемо посматрати Y као „независну" променљиву. Свакој вредности $Y=y$, одговараће случајна променљива X са условним законом вероватноћа

$$f\left(x / y\right) = \frac{f\left(x, y\right)}{f_2\left(y\right)}$$

Кад за сваку вредност y одредимо очекивану вредност X коју ћемо означити са

$$m_1\left(y\right) = E\left(X / y\right)$$

тада ће скуп тачака

$$\left[x = m_1(y); y\right] \tag{6.45}$$

представљати криву линију, која се назива ***регресиона крива за случајну промењиву X у односу на Y*** *(Слика 6.7. (b)).*

У општем случају, регресионе криве (6.44) и (6.45) неће се поклопити. Код неких модела расподеле регресионе криве (6.44) и (6.45) су праве линије.

У специјалном случају кад су промењиве X и Y међусобно независне, условни закон вероватноћа за промењиву Y своди се на маргинални закон вероватноћа

$$f\left(y/x\right)=f_2\left(y\right)$$

тако да је

$$E\left(Y/x\right)=m_2$$

У том случају регресиона крива је скуп тачака

$$\left[x;y=m_2\right]$$

а то су тачке на правој паралелној X оси, која сече осу Y у тачки $y=m_2$.

На сличан начин се може закључити да ће и друга регресина крива бити скуп тачака

$$\left[x=m_1;y\right]$$

а то је права линија паралелна Y оси, која сече осу X у тачки $x=m_1$.

Очигледно је да су регресионе криве, у овом случају праве које су нормалне једна на другу и које се секу у тежишту расподеле $\left(m_1,m_2\right)$.

У другом екстремном случају, кад су X *и* Y потпуно зависне случајне промењљиве очекивана вредност Y под условом да је $X=x$, мора се поклопити са вредношћу Y која одговара потпуној зависности. Аналоган закључак се може извести и за условну очекивану вредност X. Тако имамо следеће тврђење.

Ако су X и Y потпуно зависне случајне промењљиве, тада се обе регресионе криве поклапају међусобно и представљају криву која одговара вези промењљивих X и Y.

На сличан начин се могу дефинисати и регресионе криве и у случају кад је *(X,Y)* прекидна случајна промењљива.

За сваку вредност промењљиве $X=x_i$ и друга случајна промењљива Y узимаће вредности $y_1,y_2,...$ са условним законом вероватноћа

$$p_{j/i} = \frac{p_{ij}}{p_{i\bullet}}$$

Условну очекивану вредност означићемо са $m_2(x_i)$. То ће бити

$$m_2(x_i) = \sum_j y_j p_{j/i} = \frac{1}{p_i} \sum_j y_j \cdot p_{ij}$$

Тако ћемо добити скуп тачака

$$\left[x_i ; y_i = m_2(x_i) \right] \qquad i=1,2,... \tag{6.46}$$

којим ће бити представљена ***регресиона веза Y, у односу на X***.

При томе је $m_2(x_i)$ условна очекивана вредност *Y*, под условом да је $X = x_i$.

На сличан начин ће скуп тачака

$$\left[x_j = m_1(y_j), y_j \right] \qquad j=1,2,... \tag{6.47}$$

представљати ***регресиону везу X, у односу на Y.***

При томе је $m_1(y_j)$ условна очекивана вредност за *X* под условом да је $Y = y_j$.

У оба случаја, посматране тачке можемо спојити правим линијама и на тај начин добити регресионе криве прекидне случајне промењиве (X,Y). Регресионе криве имају једно веома важно оптимално својство на основу кога се могу сматрати као „најбоље могуће" оцене једне промењиве преко вредности друге промењиве.

Наиме, ако желимо да за поједине вредности *X*, предвидимо вредност *Y*, онда то можемо учинити тако што ћемо одредити неку функционалну везу облика

$$Y_p = g(X)$$

при чему смо означили са Y_p, вредност Y одређену из функције g, а на основу вредности X.

Наше предвиђање ће бити боље уколико је одступање појединих предвиђених вредности Y_p и правих вредности Y, што мање. Пошто одступање меримо варијансом, онда то значи да треба да минимизирамо очекивану вредност квадрата разлика правих и предвиђених вредности. Према томе, наше предвиђање је „најбоље" ако

$$E\left\{\left(Y-Y_p\right)\right\}^2 = E\left\{Y-g(X)\right\}^2 \tag{6.48}$$

има најмању могућу вредност.

Дакле, наш циљ би био да се, од свих функција $g(X)$, нађе она функција код које ће (6.48) имати минималну вредност. Израз (6.48) представља средње квадратно одступање правих и предвиђених вредности за Y, а *поступак којим се минимизира средње квадратно одступање, назива се* ***метод најмањих квадрата***.

Уверићемо се да регресиона крива

$$\left[x;\ y=m_2(x)\right]$$

минимизира средње квадратно одступање.

Нека је $f(x,y)$ закон вероватноћа за променљиву (X,Y). Он се може изразти преко условног закона вероватноћа

$$f(x,y) = f_1(x)\cdot f(y/x)$$

Средње квадратно одступање функције $g(X)$ биће

$$E\left\{Y-g(X)\right\}^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[y-g(x)\right]^2 f(x,y)dxdy =$$

$$= \int_{-\infty}^{\infty}\left\{f_1(x)\int_{-\infty}^{\infty}\left[y-g(x)\right]^2 f(y/x)dy\right\}dx$$

Пошто је $f_1(x)$, за свако *x*, ненегативно (као закони вероватноћа), онда ће средње квадратно одступање имати минимум за ону вредност за коју израз

$$\int_{-\infty}^{\infty}\left[y-g(x)\right]^2 f\left(y/x\right)dx \tag{6.49}$$

има минимум.

У средњој загради израза (6.49) можемо додати и одузети условну очекивану вредност променљиве *Y*, под условм да је *X*=*x*. Њу смо означавали са $m_2(x)$. После квадрирања и растављања на три интеграла добиће се да је

$$\int_{-\infty}^{\infty}\left[y-g\left(x\right)\right]^2 f\left(y/x\right)dy=\int_{-\infty}^{\infty}\left[y-m_2\left(x\right)\right]^2 f\left(y/x\right)dy+\left[m_2\left(x\right)-g\left(x\right)\right]^2$$

Одавде се види да ће средње квадратно одступање имати минимум кад функција $g(X)$ задовољава услов

$$\left[m_2\left(x\right)-g\left(x\right)\right]^2=0$$

односно, кад је

$$g\left(x\right)=m_2\left(x\right)$$

Минимална вредност посматраног средњег квадратног одступања биће једнака

$$\int_{-\infty}^{\infty}\left[y-m_2\left(y\right)\right]^2 f\left(y/x\right)dy$$

а то није ништа друго него условна варијанса променљиве *Y*, под условом да друга променљива *X* има вредност *x*.

Тако смо дошли до следећег закључка: средње квадратно одступање случајне променљиве Y и било које функције друге случајне променљиве X,

$$E\left\{Y-g\left(X\right)\right\}^2$$

имаће најмању вредност ако се за функцију $g(X)$ *узме регресиона крива*

$$y=m_2\left(X\right)$$

а та минимална вредност биће једнака условној варијанси $\sigma^2(Y/x)$.

Наравно, особину „најбоље могуће" оцене поседује и друга регресиона крива везана за случајну променљиву *X* у односу на *Y*.

Кад је (X,Y) прекидна случајна променљива, онда се, сличним поступком, може закључити да ће скуп тачака (6.46) и (6.47), који представљају регресионе криве, бити „најбољи" могући скуп тачака преко кога се могу изразити вредности једне променљиве у односу на промене друге случајне променљиве.

Регресионе хиперповршине

Појам регресионих кривих може се проширити и на *k*-димензионалне случајне променљиве

$$\left(X_1,X_2,\ldots,X_k\right)$$

Најчешће нас интересује какве су промене једне компоненте рецимо X_1, у зависности од промена осталих компоненти $X_2,X_3,\ldots,X_k$..

За сваку фиксну вредност компоненти

$$X_2=x_2,\ X_3=x_3,\ldots,X_k=x_k$$

можемо одредити условну расподелу за X_1, а затим из ње одредити очекивану вредност. Нека је то

$$m_1\left(x_2,\ldots,x_k\right)=E\left\{X_1/x_2,x_3,\ldots,x_k\right\}$$

Тада скуп тачака у простору

$$\left[x_1 = m_1\left(x_2,\ldots,x_k\right), x_2,\ldots,x_k\right]$$

образују хиперповршину коју ***називамо регресиона хипепровршина*** X_1 ***у односу на*** $X_2, X_3,\ldots, X_k$.

Регресионе хиперповршине имају исту особину као и регресионе криве. Средње квадратно одступање промењиве X_1 и било које функције промењивих $X_2,\ldots, X_k$, тј.

$$E\left\{X_1 - g\left(X_2, X_3,..., X_k\right)\right\}^2$$

има *минималну вредност* ако је функција $g\left(X_2,\ldots, X_k\right)$, у ствари, регресиона хиперповршина, тј. ако је

$$x_1 = g\left(x_2, x_3,\ldots, x_k\right) = m_1\left(x_2, x_3,\ldots, x_k\right)$$

Поред тога, регресионе хиперповршине поседују следећу особину. Посматрајмо случајну промењиву X_1 и било коју функцију

$$Z = g\left(X_2, X_3,\ldots, X_k\right)$$

преосталих компоненти $X_2, X_3,\ldots, X_k$, као другу случајну промењиву.

Тада ће коефицијент корелације између X_1 и Z имати највећу вредност ако је функција g регресиона хиперповршина тј. ако је

$$g\left(X_2, X_3,..., X_k\right) = m_1\left(X_2, X_3,..., X_k\right)$$

Напоменимо да је минимум средњег квадратног одступања

$$E\left\{X_1 - g\left(X_2,\ldots, X_k\right)\right\}^2$$

једнак условној варијанси промењиве X_1, под условом да остале компоненте имају вредности $x_2, x_3,\ldots, x_k$.

Средње квадратне регресионе праве

Често је регресионе криве тешко одредити, а и изрази којима су одређене најчешће су компликовани, па је њихова употреба ограничена. Зато је потребно да на неки начин утврдимо једноставније везе између променљивих.

Поступак утврђивања једноставнијих веза је следећи: уместо било какве функције

$$Y = g(X)$$

посматраћемо само линеарне функције

$$Y = \alpha X + \beta$$

па ћемо на основу неког критеријума изабрати „најбољу" линеарну везу.

Наравно, за критеријум избора „најбоље" везе послужиће нам средње квадратно одступање, тј. послужиће нам функција

$$F(\alpha, \beta) = E\{Y - (\alpha X + \beta)\}^2 \qquad (6.50)$$

која представља очекивану вредност квадрата одступања случајне променљиве Y и процењене њене вредности, преко X, линеарном везом $\alpha X + \beta$. Наш је циљ да одредимо α и β из функције (6.50) тако да $F(\alpha, \beta)$ има минималну вредност. Доказаћемо следеће тврђење:

Линеарна функција $y = \alpha X + \beta$, *која минимизира средње квадратно одступање* $F(\alpha, \beta)$, *је функција*

$$y - m_y = \rho \frac{\sigma_y}{\sigma_x}(x - m_x)$$

односно, функција

$$\frac{y-m_y}{\sigma_y}=\rho\frac{x-m_x}{\sigma_x} \tag{6.51}$$

при чему су m_x *и* m_y *очекиване вредности за X и Y, а* σ_x^2 *и* σ_y^2 *су њихове варијансе, док је* ρ *коефицијент корелације између X и Y.*

Доказ. Функцију $F(\alpha,\beta)$ написаћемо у облику

$$F(\alpha,\beta)=E\left\{(Y-m_y)-\alpha\,(X-m_x)+(m_y-\alpha\,m_x-\beta)\right\}^2$$

Квадрирајући израз у средњој загради и користећи особине очекиване вредности, функцију $F(\alpha,\beta)$ можемо изразити у облику

$$\begin{aligned}F(\alpha,\beta)&=E(Y-m_y)^2+\alpha^2E(X-m_x)^2+\\&+(m_y-\alpha\,m_x-\beta)^2-2\alpha\,E(Y-m_y)(X-m_x)\end{aligned} \tag{6.52}$$

јер је $E(Y-m_y)=0$ и $E(X-m_x)=0$.

Функција $F(\alpha,\beta)$ имаће екстремну вредност у оној тачки у којој су њени парцијални изводи једнаки нули. То ће бити тачка која задовољава једначине

$$\frac{\partial F}{\partial\alpha}=2\alpha\,E(X-m_x)^2-2(m_y-m_x-\beta)m_x-2E(X-m_x)(Y-m_y)=0$$

$$\frac{\partial F}{\partial\beta}=-2(m_y-\alpha\,m_x-\beta)=0$$

односно,

$$\alpha\left[E(X-m_x)^2+m_x^2\right]+\beta\,m_x=E(X-m_x)(Y-m_y)+m_xm_y \tag{6.53}$$

$$\alpha\,m_x+\beta=m_y$$

Решење система линеарних једначина (6.53) по α и β је следеће

$$\alpha = \frac{E\left(X - m_x\right)\left(Y - m_y\right)}{E\left(X - m_x\right)^2}$$

$$\beta = m_y - m_x \cdot \frac{E\left(X - m_x\right)\left(Y - m_y\right)}{E\left(X - m_x\right)^2}$$

Пошто је варијанса за X једнака

$$E\left(X - m_x\right)^2 = \sigma_x^2$$

варијанса за Y

$$E\left(Y - m_y\right)^2 = \sigma_y^2$$

а коваријанса између X и Y

$$E\left(X - m_x\right)\left(Y - m_y\right) = \sigma_{xy}$$

онда се израз за α и β може написати у облику

$$\alpha = \frac{\sigma_{xy}}{\sigma_x^2} \qquad \beta = m_y - m_x \frac{\sigma_{xy}}{\sigma_x^2} \tag{6.54}$$

Уврштавањем ових вредности за α и β у једначину праве

$$y = \alpha x + \beta$$

добиће се једначина

$$y = \frac{\sigma_{xy}}{\sigma_x^2} x + m_y - m_x \frac{\sigma_{xy}}{\sigma_x^2}$$

односно,

$$y - m_y = \rho \frac{\sigma_y}{\sigma_x}(x - m_x)$$

јер је коефицијент корелације

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Тиме је наведено тврђење доказано.

Минимална вредност средњег квадратног одступања једнака је

$$\min F(\alpha, \beta) = \sigma_y^2 (1 - \rho^2) \tag{6.55}$$

и назива се ***резидијумска варијанса****, а означава се са* $\sigma^2_{y \cdot x}$.

Права дата једначином

$$\frac{y - m_y}{\sigma_y} = \rho \frac{x - m_x}{\sigma_x}$$

назива се ***средња квадратна регресиона права за промењиву Y у односу на X.***

Из (6.55) види се да ће *резидијумска варијанса* бити једнака нули ако, и само ако је коефицијент корелације једнак ± 1, односно, ако и само ако између случајних промењивих *X* и *Y* постоји линеарна веза.

Наравно, овај поступак можемо применити и на одређивање линеарне везе за промењиву *X* у односу на *Y*. Минимизирањем функције

$$F(a,b) = E\{X - (aY + b)\}^2$$

добићемо средњу квадратну регресиону праву

$$\frac{y-m_y}{\sigma_y}=\frac{1}{\rho}\frac{x-m_x}{\sigma_x}$$

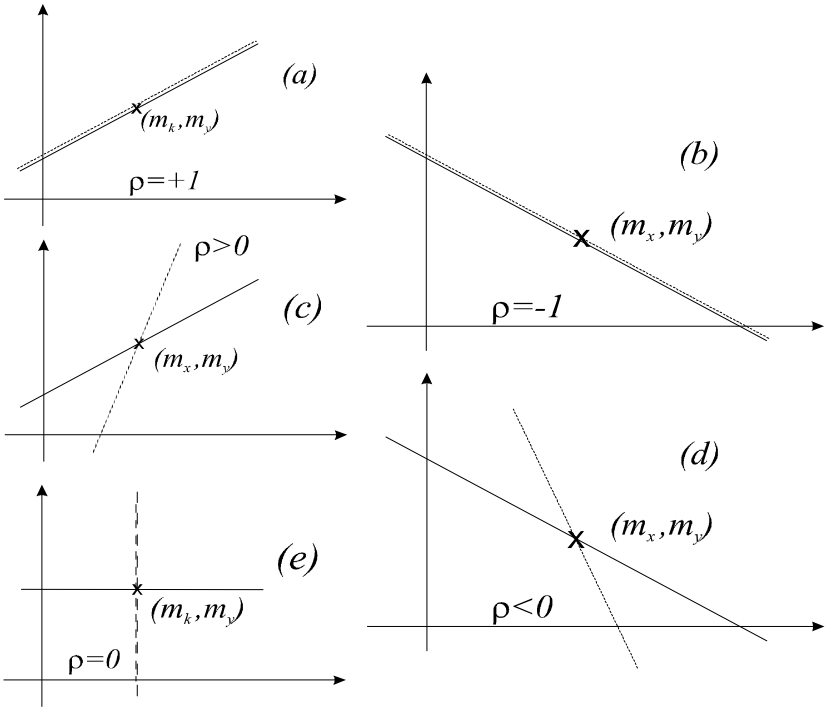
а резидијумска варијанса ће бити једнака

$$\min F(a,b)=\sigma_x^2\left(1-\rho^2\right)$$

и означићемо је са $\sigma_{x\cdot y}^2$.

Средње квадратне регресионе праве (6.51) и (6.56) у општем случају се не поклапају.

Поклопиће се онда, и само онда кад је $\rho=\pm 1$*, односно, кад постоји линеарна веза између X и Y (Види сл. 6.8.).*



Слика 6.8. Средње квадратне регресионе праве

Средње квадратне регресионе праве се секу у тежишту $\left(m_x, m_y\right)$. Кад је линеарна зависност између *X и Y* мања, онда ове праве одступају све више једна од друге (угао међу њима се повећава). Кад је коефицијент корелације једнак нули, средње квадратне регресионе праве су окомите једна на другу (Види слику. 6.8).

Видели смо да регресионе криве „најбоље" изражавају везу између променљивих *X* и *Y*, а средње квадратне регресионе праве, „најбоље" изражавају ту везу у класи линеарних веза између променљивих.

Поставља се питање да ли постоје такве расподеле случајних променљивих код којих ће се ове две регресионе везе поклопити. Одговор је потврдан. Једна од најчешће коришћених расподела у математичкој статистици, тзв. *Нормална расподела* има ту особину. Код ње су регресионе криве уствари праве које се поклапају са средњим квадратним регресионим правама.

Сада се може поставити питање утврђивања критеријума на основу кога ће се мерити ефикасност употреба средње квадратне регресионе праве у одређивању вредности случајне променљиве *Y* на основу изабраних вредности променљиве *X*.

Ту ефикасност меримо преко ***линеарног корелационог количника*** *дефинисаног изразом,*

$$n_{y\cdot x} = \frac{\sigma_{y\cdot x}^2}{\sigma_y^2} = 1 - \rho^2$$

који представља однос резидијумске варијансе Y у односу на X и варијансе Y.

Линеарни корелациони количник је једнак нули ако и само ако су *X* и *Y* линеарно зависне променљиве. Тада се, за сваку вредност *X*, може тачно одредити вредност *Y*. У другом крајњем случају, кад су *X* и *Y* некорелиране случајне променљиве, линеарни корелациони количник је једнак јединици и тада, познавање вредности за

променљиву X нам неће дати никакву информацију о вредности променљиве Y.

Средње квадратне регресионе хиперравни

Посматраћемо k-димензионалну случајну променљиву $(X_1, X_2, ..., X_k)$. Нека је дата линеарна функција

$$\beta_1 + \beta_2 X_2 + ... + \beta_k X_k$$

при чему су $\beta_1, \beta_2, ..., \beta_k$ непознати коефицијенти.

Одредићемо ове коефицијенте тако да линеарна функција на „најбољи" начин изражава везу између променљиве X_1 и преосталих компоненти $X_2, X_3, \ldots, X_k$.

Као критеријум избора узећемо средње квадратно одступање, и тражити његов минимум. Биће то функција

$$F(\beta_1, \beta_2, ..., \beta_k) = E\left\{X_1 - (\beta_1 + \beta_2 X_2 + ... + \beta_k X_k)\right\}^2$$

Коефицијенте $\beta_1, ..., \beta_k$, који минимизирају функцију F, добићемо као резултат решења система једначина.

$$\frac{\partial F}{\partial \beta_1} = \frac{\partial F}{\partial \beta_2} = ... = \frac{\partial F}{\partial \beta_k} = 0$$

Fункцију F можемо написати у подеснијем облику

$$F(\beta_1, ..., \beta_k) = E\left\{(X_1 - m_1) - \left[\beta_2 (X_2 - m_2) + ... + \beta_k (X_k - m_k)\right] - \right.$$
$$\left. - (\beta_1 - m_1 + \beta_2 m_2 + ... + \beta_k m_k)\right\}^2$$

Прва једначина

$$\frac{\partial F}{\partial \beta_1} = 0$$

своди се на једначину

$$\beta_1 - m_1 + \beta_2 m_2 + ... + \beta_k m_k = 0$$

Кад овај резултат заменимо у функцију *F*, добићемо да је

$$F(\beta_2,...,\beta_k) = E\left\{(X_1 - m_1) - \left[\beta_2(X_2 - m_2) + ... + \beta_k(X_k - m_k)\right]\right\}^2.$$

Изједначавањем парцијалних извода функције *F* по $\beta_2,...,\beta_k$ са нулом, добићемо систем од *(k-1)* једначина са *(k-1)* непознатих.

$$\begin{aligned} &\beta_2\ \sigma_{22} + ... + \beta_k\ \sigma_{k2} = \sigma_{21} \\ &\beta_2\ \sigma_{23} + ... + \beta_k\ \sigma_{k3} = \sigma_{31} \\ &..... \\ &\beta_2\ \sigma_{2k} + ... + \beta_k\ \sigma_{kk} = \sigma_{k1} \end{aligned} \qquad (6.57)$$

при томе су σ_{ij} *(i = 1,2,...,k; j = 1,2,...,k)* елементи дисперсионе матрице случајних промењивих.

Ако не постоји линеарна зависност између компоненти $X_2,..., X_k$, онда је детерминанта система (6.57) позитивна, а систем има једно решење.

Нека је то решење $\beta_2^*, \beta_3^*, ..., \beta_k^*$. *Тада је*

$$\beta_1^* = m_1 - \beta_2^* m_2 - ... - \beta_k^* m_k$$

а хиперраван

$$x_1 - m_1 = \sum_{k=2}^{k} \beta_k^* (x_k - m_k)$$

представља ***средњу квадратну регресиону хиперраван за*** X_1 ***у односу на*** X_2**,....,** X_k**.**

Минималну вредност функције $F(\beta_1,...,\beta_k)$, означићемо са $\sigma^2_{1\cdot 2...k}$ и звaћемо *резидијумска варијанса* компоненте X_1 у односу на преостале компоненте. Њена вредност је једнака

$$\sigma^2_{1\cdot 2...k} = \min F(\beta_1,...,\beta_k) = \frac{\det W}{\det W_{11}}$$

при чему је W дисперсиона матрица за $X_1, X_2, \ldots, X_k$, а W_{11} дисперсиона матрица за $X_2,..., X_k$.

6.4. ЗАДАЦИ

1. Одреди маргиналне и условне законе вероватноћа код дводимензионалне случајне променљиве (X,Y), чији закон вероватноћа је

X \ Y	0	1
-1	0,1	0,2
0	0,2	0,3
1	0,0	0,2

2. Истовремено се бацају две коцке. X представља број тачкица на првој, а Y број тачкица на другој коцки:
   а) Одреди вероватноћу да је $X + Y = 8$.
   б) Одреди условни закон вероватноћа за X, под условом да је $Y = 4$
   в) Да ли су X и Y независне?
3. На статистичком скупу од 100 елемената мерена су обележја X и Y. Добијени су следећи резултати:
   - пар (0.1: 100) појавио се код 15 елемената;
   - пар (0.1: 200) појавио се код 15 елемената;
   - пар (0.1: 350) појавио се код 5 елемената;
   - пар (0: 100) појавио се код 5 елемената;
   - пар (0: 200) појавио се код 25 елемената;
   - пар (0: 350) појавио се код 15 елемената;
   - пар (-0.2: 100) појавио се код 0 елемената;
   - пар (-0.2: 200) појавио се код 5 елемената;
   - пар (-0.2: 350) појавио се код 15 елемената;

   а) Одреди дводимензионалну случајну променљиву (X,Y) и њен закон вероватноћа;

б) Да ли су обележја X и Y независна?

в) Одреди условни закон вероватноћа за X, под условом да је $Y = 200$.

4. Расподела броја отказа машина у првој смени (X) и броја отказа у другој смени (Y), дата је табелом

X \ Y	0	1	2
0	0,10	0,20	0,20
1	0,04	0,08	0,08
2	0,06	0,12	0,12

а) Да ли су откази у сменама независни?

б) Ако је број отказа у првој смени 1, каква ће бити расподела броја отказа у другој смени?

5. Машина производи дневно 5 производа. Нека је X број неисправних производа, а Y број застоја машине. Расподела за (X,Y) је следећа

X \ Y	0	1	2	3	4	5
0	0,15	0,09	0,08	0,07	0,06	0,05
1	0,15	0,05	0,04	0,03	0,02	0,01
2	0,10	0,05	0,02	0,01	0,01	0,01

а) Нађи закон вероватноћа броја неисправних производа;

б) Одреди вероватноће да ће бити 2 неисправна производа, ако се зна да није било отказа машине;

в) Испитај независност X и Y.

6. Продавац продаје два стана по цени 1.000.000 и 3.000.000. У току дана први стан се може продати са вероватноћом 0.6, а

други са вероватноћом 0.2. Ако је X број продатих станова, а Y зарада, како изгледа расподела за (X,Y)?

а) Да ли су то независне случајне промењиве?

7. Маргинални закон вероватноћа случајне промењиве X дат је функцијом

$$f_1(x) = \begin{cases} 2e^{-2x} & \text{за x} > 0 \\ 0 & \text{за } x \le 0, \end{cases}$$

а условни закон вероватноћа случајне промењиве Y дат је функцијом

$$f(y/x) = \begin{cases} xe^{-xy} & \text{за } y > 0, x > 0 \\ 0 & \text{за } y \le 0. \end{cases}$$

Наћи маргинални закон вероватноћа за случајну промењиву Y.

8. Ако се, из групе од 8 исправних и 2 неисправна производа, бирају на случајан начин 4 производа, одреди закон вероватноћа за (X,Y), при чему је X број неисправних производа у изабраној групи, а Y број исправних.

9. Из скупа бројева (1,2,3,4,5) бира се на случајан начин један број. Нека је то X. У другом бирању се бира један од бројева 1,2,...,X. Нека је то Y. Одреди закон вероватноћа за (X,Y).

10. Случајна промењива X је једна од цифара (0,1,2,...,9), а Y је дато на следећи начин.

$$Y = \begin{cases} 0 & \text{ако је X паран број;} \\ 1 & \text{ако је X непаран број.} \end{cases}$$

Испитај зависност промењивих X и Y.

11. Величина X представља случајно изабрану цифру из скупа цифара $\{0,1,2,\ldots,9\}$, а Y представља случајни број из скупа бројева

$$\{X,\ X+1,\ X+2,\ X+3,\ X+4\}$$

Одреди закон вероватноће случајне променљиве (X,Y) параметре и коефицијент корелације.

12. Случајни догађаји $A_1, A_2, \ldots, A_k$ у једном експерименту се међусобно искључују. Вероватноће реализације сваког од ових догађаја једнаке су $p_1, p_2, \ldots, p_k$, респективно. При томе је $p_1 + p_2 + \ldots + p_k = 1$.

    Кад се експеримент понови n пута, можемо посматрати случајне променљиве $X_1, X_2, \ldots, X_k$, које представљају број понављања експеримента у којима су се реализовали догађаји $A_1, A_2, \ldots, A_k$. Одреди закон вероватноће за k-димензионалну случајну променљиву $(X_1, X_2, \ldots, X_k)$, а затим нађи корелациону матрицу.

13. Из цифара $\{0,1,\ldots,9\}$ бира се случајно једна цифра. Променљиве X_1, X_2, X_3 су одређене на следећи начин

$$X_1 = \begin{cases} 0 & \text{ако је изабран паран број} \\ 1 & \text{ако је изабран непаран број} \end{cases}$$

$$X_2 = \begin{cases} 0 & \text{ако изабрани број није дељив са 3} \\ 1 & \text{ако је изабрани број дељив са 3} \end{cases}$$

$$X_3 = \begin{cases} 0 & \text{ако је изабрани број мањи од 5} \\ 1 & \text{ако изабрани број није мањи од 5} \end{cases}$$

    Одреди корелациону матрицу за X_1, X_2, X_3. Објасни добијене вредности коефицијената корелације.

14. Два стрелца гађају, независно један од другог, мету по три пута. Вероватноће погодака су 0,5, за првог стрелца и 0,4 за другог. X

представља број погодака првог стрелца, а Y број погодака другог стрелца. Одреди расподелу за (X, Y).

15. Нека су $X_1, X_2, \ldots, X_k$ независне случајне промењиве са срединама једнаким 0 и варијансама једнаким 1. Покажи да је

$$P\left\{\sum_{i=1}^{k} x_2^2 \geq \lambda_k\right\} \leq \frac{1}{\lambda}$$

16. Ако су коефицијенти корелације код к-димензионалне случајне промењиве једнаки међусобно и имају вредност ρ, тада је

$$-\frac{1}{k-1} \leq \rho \leq 1$$

Докажи да је то тачно.

17. Нека су $(X_1, X_2, \ldots, X_m)$ независне случајне промењиве са варијансом 1, а $(Y_1, Y_2, \ldots, Y_n)$, такође независне са варијансом 1. Нека је коваријанса између X_i и Y_j

$$Cov(X_i, Y_j) = \rho$$

за $i = 1,2,\ldots,m,\ j = 1,2,\ldots,n$.

Покажи да је кефицијент корелације између $U = X_1 + \ldots + X_m$ и $V = Y_1 + \ldots + Y_n$

$$\rho\ (U,V) = \sqrt{mn}\rho$$

18. Одреди коефицијент корелације између У и В из задатка 17, кад је

$$Cov(X_i, X_j) = \rho_1$$

за $i = 1,2,\ldots,m;$ и $j = 1,2,\ldots,m,$ а,

$$Cov(Y_i, Y_j) = \rho_2$$

за $i = 1,2,\ldots,m;$ и $j = 1,2,...,m.$

19. Нека су $X_1, X_2, \ldots, X_k$ независне случајне промењиве са Нормалном расподелом

$$X_i : N\left(m_i; \sigma_i^2\right)$$

за $i = 1,2,...,k$. Напиши заједнички закон вероватноћа за k-димензионалну случајну промењиву $\left(X_1, X_2, \ldots, X_k\right)$.

20. Функција генератриса дводимензионалне случајне промењиве (X,Y) дефинисана је као очекивана вредност функције

$$e^{tX+uY}$$

за свако т, $u \in (-\infty;\infty)$, тј.

$$g(t,u) = E\left\{e^{tX+uY}\right\}$$

а) Напиши израз за $g(t,u)$ кад је (X,Y) прекидна и непрекидна случајна промењива.

б) Ако су X и Y независне, покажи да је

$$g(t,u) = g_1(t) \cdot g_2(u)$$

при чему су g_1 и g_2 функције генератрисе промењивих X и Y.

в) Одреди функцију генератрису за X+Y, ако су X и Y независне.

21. Закон вероватноћа случајне промењиве (X,Y) дат је функцијом

а) Одреди регресионе криве.

б) Одреди средње квадратне регресионе праве.

в) Анализирај функцију која представља разлику између условне варијансе $\sigma^2(y/x)$ и резидијумске варијансе $\sigma^2_{y\cdot x}$.

г) Анализирај разлику $\sigma^2(Y/x) - \sigma^2_{x\cdot y}$.

22. Ако промењиву (X,Y) заменимо са линеарном функцијом

$$U = aX + b$$
$$V = cY + d$$

покажи да ће коефицијент корелације између U и V бити онај исти као између X и Y. Како ће изгледати средње квадратне регресионе праве за случајну промењиву (U,V)?

23. Изрази резидијумску варијансу и корелациони количник преко корелационе матрице случајне промењиве $\left(X_1, X_2, \ldots, X_k\right)$.

24. Одреди средњу квадратну регресиону хиперраван за X_1 у односу на, $X_2, \ldots, X_k$, ако је

$$\rho_{ij} = \rho$$

за свако $i = 1,2,\ldots,k$, $j = 1,2,\ldots,k$.

Колика ће бити резидијумска варијанса $\sigma^2_{1\cdot 2\cdot 3\ldots k}$?

25. Ако су компонете $X_2, \ldots, X_k$ међусобно независне, а коефицјиент корелације

$$\rho_{ij} = \rho$$

за свако $j = 2,\ldots,k$, колика ће бити резидијумска варијанса $\sigma^2_{1\cdot 2\cdot 3\ldots k}$?

7. МОДЕЛИ РАСПОДЕЛА СЛУЧАЈНИХ ПРОМЕНЉИВИХ

На основу природе појаве коју анализирамо, често можемо претпоставити да расподела случајне променљиве X припада једној класи расподела, одређеној са једним или више непознатих параметара. Ако је та класа изражена функционалном везом

$$P(x;\theta_1,\theta_2,...\theta_k) \tag{7.1}$$

онда кажемо да функција (7.1) представља модел расподеле случајне променљиве X.

Модел расподеле (7.1) садржи све информације о посматраној појави која се изражава преко случајне променљиве X.

Постоје модели расподела који се најчешће користе и неки од њих ће бити детаљније разматрани у поглављима која следе. Биће наведени најчешће коришћени модели прекидних и непрекидних расподела случајних променљивих.

7.1. МОДЕЛИ ПРЕКИДНИХ РАСПОДЕЛА

Биномна расподела

Дефиниција 7.2: *Случајна променљива X има Биномну расподелу, ако може узети једну вредност k из низа ненегативних целих бројева*

$$\{0,1,2,...,n\}$$

са вероватноћом:

$$P_n(k) = P(X = k) = \binom{n}{k} p^k q^{n-k} \tag{7.2}$$

при чему је $0 < p < 1$, $p + q = 1$, а $\binom{n}{k} = \frac{n!}{(n-k)!k!}$ представља биномни коефицијент.

Чињеницу да случајна променљива X има Биномну расподелу означаваћемо са:

$$X : \mathrm{B}(n; p)$$

Закон вероватноћа Биномне расподеле дат је следећом табелом:

X	0	1	2	...	n
p(x)	$\binom{n}{0} p^0 q^{n-0}$	$\binom{n}{1} pq^{n-1}$	$\binom{n}{2} p^2 q^{n-2}$	...	$\binom{n}{n} p^n q^0$

Бројеви n и p су параметри модела.

На основу биномног обрасца, лако је проверити да је збир биномних вероватноћа једнак јединици. Заиста,

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

Функција генератриса Биномне расподеле је:

$$g(t) = \sum_{k=0}^{n} e^{kt}\binom{n}{k}p^k q^{n-k} = \sum_{k=0}^{n}\binom{n}{k}\left(pe^t\right)^k q^{n-k} = \left(\mathrm{q} + \mathrm{pe}^{\mathrm{t}}\right)^n \qquad (7.3)$$

Извод функције $g(t)$ у тачки $t = 0$, представља очекивану вредност:

$$E(X) = np \qquad (7.4)$$

Преко другог извода функције $g(t)$ у тачки $t = 0$ добија се други обични моменат, па се лако израчунава вредност варијансе:

$$\sigma^2 = npq \qquad (7.5)$$

Може се проверити да је коефицијент асиметрије:

$$\beta_1 = \frac{q-p}{\sqrt{npq}} \qquad (7.6)$$

а коефицијент спљоштености:

$$\beta_2 = 3 + \frac{1-6pq}{npq} \qquad (7.7)$$

Из (7.6) и (7.7), следи да Биномна расподела тежи симетричној расподели кад $n \to \infty$. Исто тако, она тежи расподели са „нормалном" спљоштеношћу.

Специјално, кад је $p = q = 0{,}5$, Биномна расподела је симетрична расподела за све вредности n.

Биномна расподела се користи у ***Бернулијевој (Bernulli) шеми*** независних понављања експеримента, обрађеној у поглављу 4.2.

Одиграла је изузетно важну улогу у теорији вероватноће, њеној примени и у развоју статистике. Вероватноће случајних догађаја везаних за Бернулијеву шему експеримената могу се одредити у оквиру Биномне расподеле.

Пуасонова расподела

Пуасонова расподела названа је по француском математичару чије име је Simeon D. Poisson (1781-1840). Она описује вероватноћу да ће случајни догађај наступити у одређеном временском или просторном интервалу под одређеним условима, при чему је вероватноћа догађаја веома мала, а број покушаја (временски или просторни интервал) веома велики. Посматрани догађај се, дакле, ретко реализује. Пуасоновом расподелом се описује, на пример, број грешака по страници књиге, број жртава аутомобилских удеса у току једног месеца у неком граду и слично.

Дефиниција 7.3: *Случајна променљива X има Пуасонову расподелу ако може узети вредност k из низа ненегативних целих бројева*

$$\{0, 1, 2, 3, \ldots\}$$

и то са вероватноћом

$$P(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

при чему је $\lambda > 0$*, реални број и представља параметар расподеле.*

Чињеницу да случајна променљива има Пуасонову расподелу са параметром λ означавамо са:

$$X : \mathcal{P}(\lambda)$$

Закон вероватноћа дат је табелом:

X	0	1	2	...	k	...
p(x)	$\frac{\lambda^0}{0!}e^{-\lambda}$	$\frac{\lambda^1}{1!}e^{-\lambda}$	$\frac{\lambda^2}{2!}e^{-\lambda}$	...	$\frac{\lambda^k}{k!}e^{-\lambda}$	...

Зна се да је

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

Зато је збир вероватноћа код Пуасонове расподеле једнак:

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

Очекивана вредност случајне променљиве која има Пуасонову расподелу је:

$$m = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \cdot \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \tag{7.8}$$

Други обичnі моменат је једнак:

$$m_2 = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \left[k(k-1) + k\right] \frac{\lambda^k}{k!} =$$

$$= e^{-\lambda} \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda = \lambda^2 + \lambda$$

Варијанса случајне променљиве X је, према томе:

$$\sigma^2 = m_2 - m^2 = \lambda \tag{7.9}$$

$$\mu_3 = \lambda$$

$$\mu_4 = 3\lambda^2 + \lambda$$

па је коефицијент асиметрије једнак:

$$\beta_1 = \frac{1}{\lambda}$$

а коефицијент спљоштености

$$\beta_2 = 3 + \frac{1}{\lambda}$$

Геометријска расподела

Посматрајмо статистички експеримент чији су могући исходи случаијни догађај A, чија је вероватноћа $P(A) = p$, и догађај $\overline{A}$ са вероватноћом $P(\overline{A}) = 1 - p = q$. Нека је у сваком (независном) понављању експеримента вероватноћа реализације догађаја A иста и једнака p. Док се у Бернулијевој шеми тражила вероватноћа да се у n понављања експеримента догађај A реализује тачно k пута, сада се посматра случајна променљива X која представља број понављања описаног експеримента до прве реализације догађаја A. Овако дефинисана случајна променљива X може да узме вредности 1, 2, 3, ... Вероватноћа да случајна променљива X узме вредност x дата је изразом:

$$p(x) = P(X = x) = q^{x-1} p \qquad \text{за } x = 1, 2, 3... \qquad (7.10)$$

и представља вероватноћу да се у првих $x - 1$ понављања експеримента није реализовао посматрани догађај A, а да се реализовао у x-том понављању.

Формулом (7.10) дефинисан је закон вероватноћа Геометријске расподеле. Овај назив расподеле проистекао је из чињенице што вероватноће ове расподеле чине геометријски низ. Коришћењем

обрасца за збир бесконачног опадајућег геометријског низа показује се да је збир свих вероватноћа једнак јединици:

$$\sum_{x=1}^{\infty} p(x) = p\sum_{x=1}^{\infty} q^{x-1} = p\frac{1}{1-q} = \frac{p}{p} = 1$$

Функција генератриса Геометријске расподеле је:

$$g(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} pq^{x-1} = \frac{pe^t}{1-qe^t}$$

Математичко очекивање и варијанса су:

$$m = \frac{1}{p}; \qquad \sigma^2 = \frac{q}{p^2}$$

Вероватноћа да се до прве реализације догађаја A експеримент понови највише k пута износи:

$$p(X \le k) = p_1 + p_2 + \ldots + p_k = p\sum_{x=1}^{k} q^{x-1} = 1 - q^k$$

Овај резултат може се добити и нешто другачије: ако се експеримент понавља више од k пута, то значи да се догађај A није реализовао у првих k понављања експеримента. Овом случају одговара вероватноћа q^k, па је тада вероватноћа супротног догађаја једнака $1-q^k$.

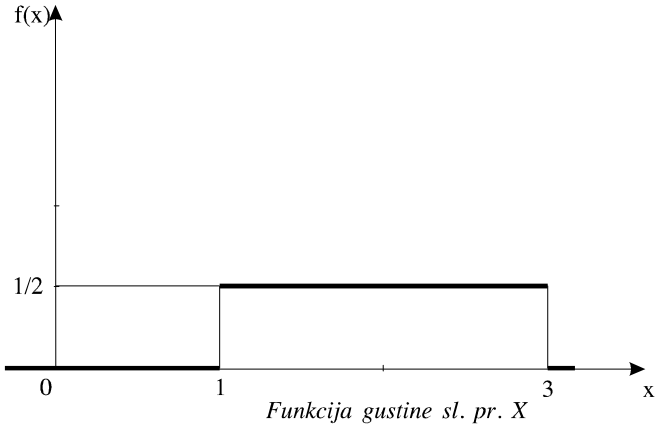
7.2. МОДЕЛИ НЕПРЕКИДНИХ РАСПОДЕЛА

Униформна расподела

Дефиниција 7.17: *Случајна променљива X има* ***Униформну расподелу*** *ако је њен закон вероватноћа дат функцијом:*

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha}, & \alpha<x<\beta\\ 0, & \text{за остале } x\end{cases}$$

Закон вероватноћа (функција густине) за случајну променљиву која има Униформну расподелу са параметрима $\alpha=1$ и $\beta=3$ дат је на слици:

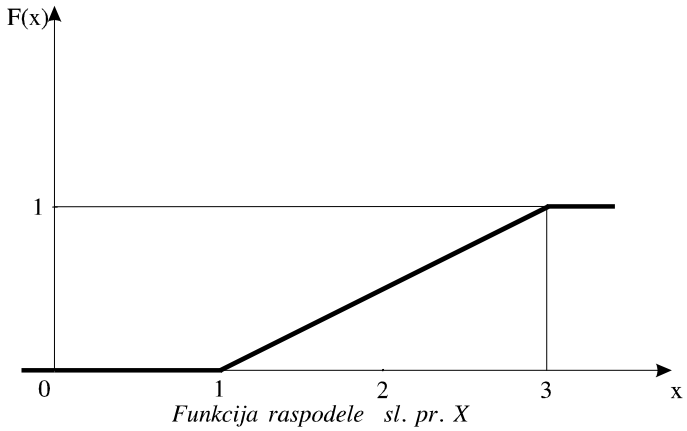


Funkcija gustine sl. pr. X

Функција расподеле за случајну променљиву X која има Униформну расподелу са параметрима α и β је:

$$F(x) = P\{X \le x\} = \begin{cases} 0 & ; \quad x < \alpha \\ \dfrac{x-\alpha}{\beta-\alpha} & ; \quad \alpha \le x \le \beta \\ 1 & ; \quad x > \beta \end{cases}$$

За $\alpha=1$ и $\beta=3$, график ове функције је приказан на слици:



Funkcija raspodele sl. pr. X

Карактеристике ове расподеле су следеће:

Математичко очекивање:

$$E(X) = \frac{\alpha+\beta}{2}$$

Варијанса:

$$\sigma^2 = \frac{(\beta-\alpha)^2}{12}$$

Оса симетрије закона вероватноћа случајне променљиве која има Униформну расподелу са параметрима α и β је:

$$x = \frac{\alpha+\beta}{2}$$

Нормална расподела

Приликом решавања проблема граничне функције Биномне расподеле, De Moivre је 1733. године открио Нормалну расподелу, што је остало доста незапажено све до радова C.F.Gauss-a 1809. године и P.S.Laplace-a 1812. године. Ова два аутора су развили теорију ***грешака опсервација*** и дошли до Нормалне расподеле. Laplace је доказао тзв. ***централну граничну теорему*** и дао одговоре на различита питања из теорије вероватноће.

Под њиховим утицајем се дуго сматрало да се статистичке расподеле, практично, свих врста података приближавају Нормалној расподели као идеалном облику и да је само потребно обезбедити довољно велики број тачних опсервација. Одступања сваке случајне варијабле од њене очекиване вредности посматране су као „грешке" и, на основу „закона грешака", описане су Нормалном расподелом. Показало се да је то јесте тачно у многим физичким, астрономским, демографским и биолошким мерењима, а такође и у техничким и економским наукама. Ова чињеница омогућава да се, у оквиру статистичког закључивања и доношења одлука базираних на статистичком закључивању, велики део метода базира на моделима који су генерисани Нормалном расподелом.

Дефиниција 7.5: *Случајна променљива X има* ***Нормалну расподелу*** *ако је њен закон вероватноћа дат функцијом*

$$f(x)=\frac{1}{b\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2b^2}} \quad \text{за } x\in(-\infty,\infty) \tag{7.11}$$

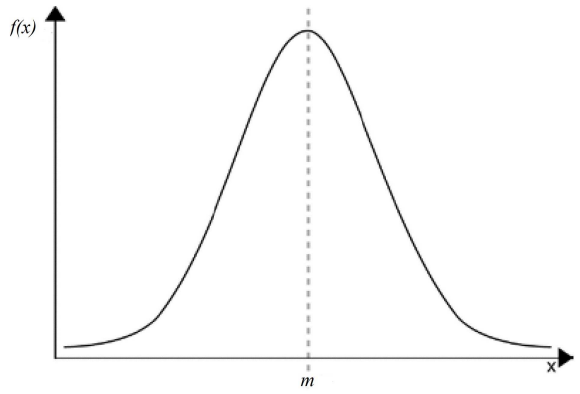
Вредности a и b^2 су параметри модела, a је било који реални број, а b^2 је позитивни реални број. Нормалну расподелу означаваћемо са:

$$X: \quad N\left(a,b^2\right)$$

Функција (7.11) је симетрична у односу на праву $x = a$. Има максимум у тачки $x = a$. Максимална вредност функције је једнака:

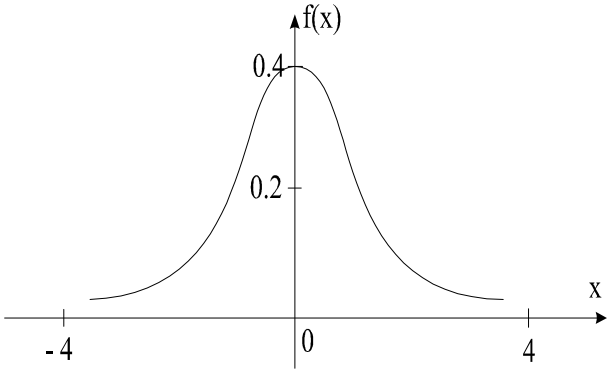
$$f(a) = \frac{1}{b\sqrt{2\pi}}$$

Графички приказ функције (7.11), дат је на слици:



Слика 7.1. Нормална расподела

Графички приказ функције (7.11), за *a=0*, дат је на слици:



Слика 7.2. Стандардизована Нормална расподела

Очекивана вредност случајне променљиве X која има Нормалну расподелу одређује се преко интеграла:

$$m = E(X) = \frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-a)^2}{2b^2}} dx$$

Увођењем смене $y = \frac{x-a}{b}$ и парцијалном интеграцијом, добија се да је

$$m = a \tag{7.12}$$

Варијансу случајне променљиве X са Нормалном расподелом израчунава се преко интеграла:

$$\sigma^2 = E(X-m)^2 = \frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2b^2}} dx$$

Увођењем смене $y = \frac{x-a}{b}$, и парцијалном интеграцијом, добија се да је

$$\sigma^2 = b^2 \tag{7.13}$$

Из (7.12) и (7.13) види се да су a и b^2, очекивана вредност и варијанса случајне променљиве X. Зато се најчешће за цлучајну променљиву са Нормалном расподелом и користи ознака:

$$X: \quad N(m, \sigma^2)$$

а њен закон вероватноћа дат је у облику:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Специјално, када је $m=0$ *и* $\sigma^2=1$, *случајна променљива са законом вероватноћа датим тзв. Gauss-овом кривом:*

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

има ***Стандардизовану нормалну расподелу*** *чије вредности су дате у табелама или у статистичким пакетима програма.*

Вредности функције расподеле $F(x)$, за стандардизовану Нормалну расподелу, дате су у табелама и означене су са $\Phi(x)$.

За одређивање вероватноће да ће се вредности случајне променљиве *X*, која има Нормалну расподелу са параметрима *m* и σ^2, наћи у интервалу (x_1, x_2) најчешће се користе вредности функције стандардизоване Нормалне расподеле дате у табелама. Наиме, из једначине

$$P\{x_1 < X \le x_2\}=\frac{1}{\sigma\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx$$

после смене

$$y=\frac{x-m}{\sigma}$$

добиће се да је

$$P\{x_1 < X \le x_2\}=\frac{1}{\sqrt{2\pi}}\int_{\frac{x_1-m}{\sigma}}^{\frac{x_2-m}{\sigma}} e^{-\frac{x^2}{2}}dx=\Phi\left(\frac{x_2-m}{\sigma}\right)-\Phi\left(\frac{x_1-m}{\sigma}\right)$$

На основу наведеног се може закључити следеће: ако важи

X: N(m,σ^2*),* тада $Y=\frac{X-m}{\sigma}$: *N(0,1).* Наведена трансформација се често назива *стандардизацијом* случајне променљиве *X*.

Функција генератриса Нормалне расподеле је функција

$$g(t) = e^{tm+\frac{t^2\sigma^2}{2}}$$

Израчунавањем обичних момената преко функције $g(t)$ и одговарјућих централних момената, лако је проверити да је

$$\beta_1 = 0$$
$$\beta_2 = 3$$

из чега се закључује да је Нормална расподела симетрична и да има нормалну спљоштеност.

Хи-квадрат расподела (χ^2)

Хи-квадрат расподела дефинише се као расподела суме квадрата независних случајних променљивих које имају Нормалну расподелу, са математичким очекивањем 0 и стандрадном девијацијом 1. Ова расподела је важна у статистици јер описује расподелу вероватноћа које се користе у неколико најчешће коришћених статистичких процедура.

Дефиниција 7.7: *Случајна променљива X има Хи-квадрат расподелу са n степени слободе ако је њен закон вероватноћа дат функцијом*

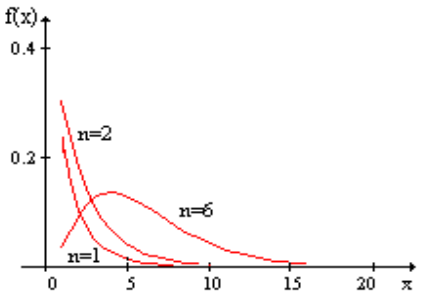
$$k_n(x) = \begin{cases} c_n x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{за x > 0} \\ 0, & \text{за x} \le 0 \end{cases}$$

при чему је n природан број, а константа c_n је дата са

$$c_n = \frac{1}{\int_0^{\infty} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx} = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$$

где је $\Gamma(p) = \int_0^{\infty} t^{p-1} e^{-t} dt, \quad p > 0$ *је Гама функција или Ојлерова функција друге врсте.*

Закони вероватноћа за *Хи*-квадрат расподелу и за неке вредности *n*, приказани су на слици:



Слика 7.3. Хи - квадрат расподеле

Случајну променљиву која има *Хи*-квадрат расподелу са *n* степени слободе означаваћемо са

$$X : \chi_n^2$$

Функција расподеле ове случајне променљиве означава се са

$$K_n(z) = C_n \int_0^z x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx$$

Њене вредности су дате табелама за степене слободе од 1 до 30. За веће степене слободе могу се користити приближне вредности из табеле функције Нормалне расподеле.

Очекивана вредност случајне променљиве која има *Хи*-квадрат расподелу са *n* степени слободе је:

$$E(x) = n$$

а варијанса:

$$\sigma^2 = E(x-m)^2 = 2n$$

Обичан моменат реда *k* дат је изразом

$$m_k = n(n+2)(n+4)...(n+2k-2)$$

тако да је

$$\mu_3 = 8n$$
$$\mu_4 = 12n(n+4)$$

Коефицијент асиметрије има вредност

$$\beta_1 = \sqrt{\frac{8}{n}}$$

а коефицијент спљоштености је

$$\beta_2 = 3 + \frac{12}{n}$$

Из формуле за коефицијент асиметрије се види да *Хи*-квадрат расподела тежи симетричној расподели кад $n \to \infty$, а из формуле за коефицијент спљоштености се види да њена спљоштеност тежи нормалној.

Студентова расподела

Студентову расподелу дефинисао је W.S. Gosset у раду објављеном под псеудонимом „Студент”.

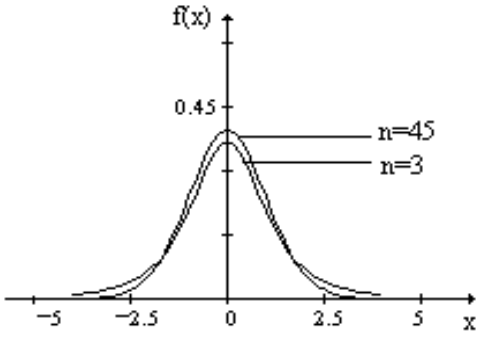
Дефиниција 7.8: *Случајна променљива X има* ***Студентову расподелу са n степени слободе*** *ако је њен закон вероватноћа дат функцијом*

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi \cdot n} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \qquad -\infty < x < \infty$$

Случајну променљиву која има Студентову расподелу са n степени слободе означаваћемо са

$$X: t_n$$

На слици су приказане функције $f_n(x)$, за $n = 3$ и $n = 45$:



Слика 7.4. Студентова расподела

Закон вероватноћа Студентове расподеле дат је симетричном функцијом са максималном вредношћу за $x = 0$. Очекивана вредност случајне промењиве која има Студентову расподелу је:

$$m = E(x) = 0$$

Обични и централни моменти су међусобно једнаки. Варијанса је једнака:

$$\sigma^2 = \frac{n}{n-2}, \qquad n>2$$

коефицијент асиметрије $\beta_1 = 0$, а коефицијент спљоштености је једнак

$$\beta_2 = 3 + \frac{6}{n-4}, \quad n>4$$

Вредности функције Студентове расподеле налазе се у табелама, а када $n \to \infty$, Студентова расподела тежи Стандардизованој нормалној расподели.

За $X: N(0,1)$ и $Y: \chi_n^2$, које су међусобно независне, важи:

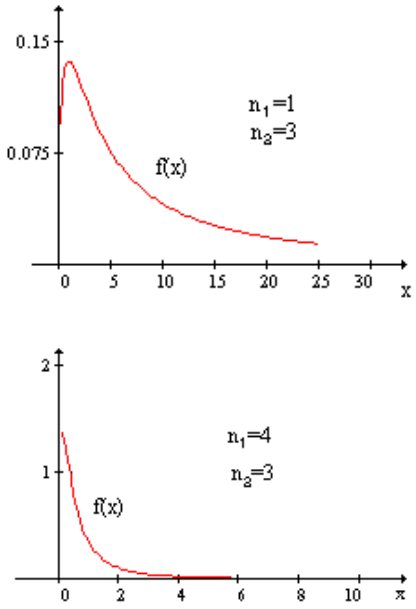
$$Z = \frac{X}{\sqrt{\frac{Y}{n}}} : \ t_n$$

F-расподела

<u>Дефиниција7.10</u>: *Случајна променљива X има* ***F-расподелу са n_1 и n_2 степени слободе****, ако је њен закон вероватноћа дат функцијом*

$$f_{n_1,n_2}(x)=\begin{cases}\dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right)\left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)}\cdot x^{\frac{n_1}{2}-1}\cdot\left(1+\dfrac{n_1 x}{n_2}\right)^{-\frac{n_1+n_2}{2}} & \text{за } x>0\\ 0 & \text{за } x\le 0\end{cases}$$

Графици закона вероватноћа *F*-расподеле за различите вредности параметара приказани су на следећим сликама:



Слика 7.5. F-расподела

Очекивана вредност *F*-расподеле једнака је

$$E(X)=\frac{n_2}{n_2-2}$$

а варијанса је

$$\sigma^2=\frac{2n_2^2\left(n_1+n_2-2\right)}{n_1\left(n_2-2\right)^2\left(n_2-4\right)}$$

При одређивању вероватноћа случајне променљиве са *F*-расподелом користимо табелу, која садржи вредности F_0 за које је функција расподеле једнака $1-\alpha=0,95$ и $1-\alpha=0,99$, а за различите степене слободе.

Ако је случајна променљива X дефинисана као количник

$$X=\frac{\frac{Y}{n_1}}{\frac{Z}{n_2}}=\frac{n_2Y}{n_1Z},$$

при чему су Y и Z независне случајне променљиве које имају Хи-квадрат расподеле са n_1 *и* n_2 *степени слободе, респективно. Тада X има* ***F-расподелу са*** $\mathbf{n_1}$ ***и*** $\mathbf{n_2}$ ***степени слободе****.*

Дводимензонална Нормална расподела

<u>Дефиниција 7.19</u>: *Дводимензионална случајна променљива (X,Y), непрекидног типа, чије вредности су све тачке равни X0Y, имаће* ***Нормалну расподелу*** *ако је закон вероватноћа дат функцијом:*

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\cdot e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-m_x)^2}{\sigma^2 x}-2\rho\frac{x-m_x}{\sigma_x}\cdot\frac{y-m_y}{\sigma_y}+\frac{(y-m_y)^2}{\sigma_y^2}\right\}} \qquad (7.24)$$

при чему су m_x и m_y очекиване вредности σ_x^2 и σ_y^2 варијансе, а ρ коефицијент корелације променљивих X и Y.

Тачка (m_x, m_y) представља тежиште расподеле. То је тачка у којој функција има максимум.

Запремина ограничена *X0Y* равни и површином $f(x,y)$ једнака је двоструком интегралу

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=1.$$

Потребно је одредити **маргиналне законе вероватноћа**. Маргинални закон вероватноћа за *X* једнак је интегралу

$$f_1(x)=\int_{-\infty}^{\infty} f(x,y)dy$$

Експонент израза (7.24) може се написати у облику

$$-\frac{1}{2}\frac{(x-m_x)^2}{\sigma_x^2}-\frac{1}{2(1-\rho^2)}\left[\frac{y-m_y}{\sigma_y}-\rho\frac{x-m_x}{\sigma_x}\right]^2$$

тако да је

$$f_1(x)=\frac{e^{-\frac{1}{2}\frac{(x-m_x)^2}{\sigma_x^2}}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{y-m_y}{\sigma_y}-\rho\frac{x-m_x}{\sigma_x}\right]^2}dy$$

Уводећи смену

$$t=\frac{1}{\sqrt{1-\rho^2}}\left[\frac{y-m_y}{\sigma_y}-\rho\frac{x-m_x}{\sigma_x}\right]$$

горњи интеграл се своди на

$$f_1(x)=\frac{e^{-\frac{1}{2}\frac{(x-m_x)^2}{\sigma_x^2}}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\sigma_y\sqrt{1-\rho^2}\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt \qquad (7.14)$$

односно,

$$f_1(x)=\frac{1}{\sigma_x\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-m_x)^2}{\sigma_x^2}}$$

јер је

$$\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt=\sqrt{2\pi}$$

На сличан начин би добили и маргинални закон вероватноћа за *Y*. То би била функција

$$f_2(y)=\frac{1}{\sigma_y\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(y-m_y)^2}{\sigma_y^2}} \qquad (7.15)$$

На основу (7.14) и (7.15) можемо тврдити:

Ако дводимензионална случајна променљива (X,Y) има Нормалну расподелу дату законом вероватноћа (7.24), тада и X и Y имају маргиналне Нормалне расподеле са параметрима (m_x, σ_x) *и* (m_y, σ_y).

За даљу анализу потребно је утврдити условне законе вероватноћа. Видели смо да је условни закон вероватноћа за случајну променљиву *X*, под условом да је $Y = y$, дат количником

$$f(x/y) = \frac{f(x,y)}{f_2(y)}$$

За случај Нормалне расподеле, овај условни закон вероватноћа имаће облик

$$f(x/y) = \frac{\sigma_y\sqrt{2\pi}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-m_x)^2}{\sigma^2{}_x} - 2\rho\frac{x-m_x}{\sigma_x}\cdot\frac{y-m_y}{\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right\} + \frac{1}{2}\frac{(y-m_y)^2}{\sigma_y^2}}$$

После скраћења и сређивања израза у експоненту добиће се

$$f(x/y) = \frac{1}{\sigma_x\sqrt{1-\rho^2}\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma_x^2(1-\rho^2)}\left[x-\left(m_x-\rho\frac{\sigma_x}{\sigma_y}(y-m_y)\right)\right]^2} \qquad (7.16)$$

Ако са $\sigma^2_{x/y}$ означимо условну варијансу за променљиву *X*, под условом да је $Y = y$, а са $m_1(y)$ условну очекивану вредност, онда из условног закона вероватноћа (7.16) следи да је

$$\sigma^2_{x/y} = \sigma_x^2\left(1-\rho^2\right)$$

$$m_1(y) = m_x + \rho\frac{\sigma_x}{\sigma_y}(y-m_y)$$

Сада се условни закон вероватноћа за промењиву X може написати у облику

$$f(x/y)=\frac{1}{\sigma_{x/y}\sqrt{2\pi}}e^{-\frac{1}{2}\frac{[x-m_1(y)]^2}{\sigma^2_{x/y}}}$$

Тако смо добили следећи резултат:

Условна расподела за случајну промењиву X, под условом да је $Y=y$ *, је Нормална расподела са условном средином*

$$m_1(y)=m_x+\rho\frac{\sigma_x}{\sigma_y}(y-m_y) \tag{7.17}$$

и условном варијансом

$$\sigma^2_{x/y}=\sigma^2_x(1-\rho^2) \tag{7.18}$$

На сличан начин се може закључити да ће условна расподела за случајну промењиву Y, под условом да је $X=x$, бити Нормална расподела са условном средином

$$m_2(x)=m_y+\rho\frac{\sigma_y}{\sigma_x}(x-m_x) \tag{7.19}$$

и условном варијансом

$$\sigma^2_{y/x}=\sigma^2_y(1-\rho^2) \tag{7.20}$$

Знамо од раније да је, у случају да су случајне промењиве X и Y међусобно независне, коефицијент корелације $\rho=0$.

А сад ћемо претпоставити да X и Y имају Нормалну расподелу и да је $\rho=0$. Онда је њихов заједнички закон вероватноћа функција

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}e^{-\frac{1}{2}\left\{\frac{(x-m_X)^2}{\sigma_x^2}+\frac{(y-m_y)^2}{\sigma_y^2}\right\}}$$

а то се може написати у облику

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

при чему су $f_1(x)$ и $f_2(y)$ маргинални закони вероватноћа за X и Y. То значи да су X и Y независне случајне променљиве. Можемо то исказати следећим тврђењем:

Ако (X,Y) има Нормалну расподелу, тада је за њихову независност потребно и довољно да је коефицијент корелације једнак нули.

Параметри дводимензионалне Нормалне расподеле

У функцији $f(x,y)$, која представља закон вероватноћа дводимензионалне случајне променљиве *(X,Y)*, појављују се следећи параметри: $m_x, m_y, \sigma_x^2, \sigma_y^2$ *i* ρ.

Маргинални закон вероватноћа за X је функција

$$f_1(x)=\frac{1}{\sigma_x\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-m_X)^2}{\sigma_x^2}}$$

одакле се може закључити да је m_x очекивана вредност случајне променљиве X, а σ_x^2 њена варијанса.

Из маргиналног закона вероватноћа $f_2(y)$ може се такође закључити да је m_y очекивана вредност Y, а σ_y^2 њена варијанса.

За основне анализе дводимензионалне Нормалне расподеле потребно је још утврдити њену коваријансу.

На основу дефиниције коваријансе имаћемо

$$\sigma_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x-m_x\right)\left(y-m_y\right)f\left(x,y\right)dxdy$$

што се, у случају Нормалне расподеле, своди на

$$\sigma_{xy} = C\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x-m_x\right)\left(y-m_y\right)e^{-\frac{1}{2\left(1-\rho^2\right)}\cdot Q}dxdy \qquad (7.21)$$

при чему смо, ради краткоће, означили са

$$C = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}},$$

$$Q = \frac{\left(x-m_x\right)^2}{\sigma_x^2} - 2\rho\frac{x-m_x}{\sigma_x}\cdot\frac{y-m_y}{\sigma_y} + \frac{\left(y-m_y\right)^2}{\sigma_y^2}$$

Да би решили интеграл (7.21), прво ћемо увести смену

$$\frac{x-m_x}{\sigma_x} = u \qquad \frac{y-m_y}{\sigma_y} = v$$

тако да је

$$\sigma_{xy} = \frac{\sigma_x\sigma_y}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}u\cdot v\cdot e^{-\frac{1}{2\left(1-\rho^2\right)}\left[u^2-2\rho\ uv+v^2\right]}du\,dv$$

Уводећи смену

$$\frac{u-\rho v}{\sqrt{1-\rho^2}} = t \qquad v = s$$

чији Јакобијан је

$$J^{-1} = \frac{1}{\sqrt{1-\rho^2}}$$

израз за коваријансу сводимо на интеграл

$$\sigma_{xy} = \frac{\sigma_x \sigma_y \rho}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} \int_{-\infty}^{\infty} s^2 e^{-\frac{s^2}{2}} ds\, dt$$

односно,

$$\sigma_{xy} = \rho \sigma_x \sigma_y ,$$

јер је

$$\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \int_{-\infty}^{\infty} s^2 e^{-\frac{s^2}{2}} ds = \sqrt{2\pi}$$

На основу дефиниције коефицијента корелације, види се да параметар ρ , који се појављује у закону вероватноће $f(x,y)$, представља коефицијент корелације.

Сад можемо резимирати:

Тежиште расподеле је у тачки

$$T\left(m_x, m_y\right)$$

а дисперсиона матрица је матрица

$$W = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

Инверзна матрица дисперсионе матрице W је матрица

$$W^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix}$$

Означимо са Q квадратну формулу

$$Q = \left(x - m_x, y - m_y\right) W^{-1} \begin{bmatrix} x - m_x \\ y - m_y \end{bmatrix}$$

Замењујући W^{-1} и множећи, добиће се да је

$$Q = \frac{1}{1-\rho^2}\left\{\frac{\left(x-m_x\right)^2}{\sigma_x^2} - 2\rho\frac{x-m_x}{\sigma_x}\cdot\frac{y-m_y}{\sigma_y} + \frac{\left(y-m_y\right)^2}{\sigma_y^2}\right\}$$

Детерминанта дисперсионе матрице W је

$$\det W = \sigma_x^2\sigma_y^2\left(1-\rho^2\right)$$

Сад се закон вероватноће дводимензионалне случајне променљиве (X,Y) може написати у облику

$$f\left(x,y\right) = \frac{1}{2\pi\cdot\det W}e^{-\frac{1}{2}Q}$$

Овај резултат указује на то да вишедимензионалну Нормалну расподелу треба проучавати преко матрице и одговарајућих квадратних форми.

Регресионе криве код Нормалне расподеле

Да би се одредила регресиона крива за Y у односу на X, треба из условног закона вероватноћа

$$f\left(y/x\right) = \frac{1}{\sigma_y\sqrt{1-\rho^2}\sqrt{2\pi}}\cdot e^{-\frac{1}{2\sigma_y^2\left(1-\rho^2\right)}\left[y-m_y-\rho\frac{\sigma_y}{\sigma_x}\left(x-m_x\right)^2\right]}$$

одредити очекивану вредност. Како је то закон вероватноћа Нормалне расподеле, онда је очекивана вредност једнака

$$m_2(x) = m_y + \rho\frac{\sigma_y}{\sigma_x}\left(x - m_x\right)$$

тако да је регресиона крива скуп тачака

$$\left[x;\ \ \frac{y-m_y}{\sigma_y} = \rho\frac{x-m_x}{\sigma_x}\right]$$

а то су тачке на правој која није ништа друго него ***средња квадратна регресиона права.***

На сличан начин би одредили и другу регресиону криву за променљиву *X* у односу на *Y*.

Ако (X,Y) *има Нормалну расподелу, тада се* ***регресионе криве*** *своде на средње квадрате регресионе праве.*

$$\frac{y-m_y}{\sigma_y} = \rho\cdot\frac{x-m_x}{\sigma_x} \text{ i}$$

$$\frac{y-m_y}{\sigma_y} = \frac{1}{\rho}\cdot\frac{x-m_x}{\sigma_x}$$

Поред тога резидијумске варијансе су једнаке

$$\sigma^2_{y/x} = \sigma^2_y\left(1-\rho^2\right)$$

$$\sigma^2_{x/y} = \sigma^2_x\left(1-\rho^2\right)$$

Види се да резидијумска варијанса за *Y*, под условом да је $X = x$, не зависи од вредности *x* и иста је за сваку вредност. Ако је $\rho = 0$, резидијумска варијанса за *Y* је једнака чистој варијанси променљиве *Y*, а то значи да нам познавање вредности за *X*, неће дати никакву информацију о вредности променљиве *Y*. У том случају *X и Y су стохастички независне.*

Са друге стране, ако је $\rho^2 = 1$, тада је резидијумска варијанса за Y једнака нули, а то значи да ћемо, на основу вредности X, моћи тачно да одредимо вредност Y. У том случају *између X и Y постоји линеарна зависност*.

Наравно, аналогни закључци на основу резидијумске варијансе, важе и за другу променљиву X, под условом да је $Y = y$.

8. РАСПОДЕЛЕ НЕКИХ ФУНКЦИЈА СЛУЧАЈНИХ ПРОМЕНЉИВИХ

Основу примене теорије вероватноће у решавању реалних проблема, представљају одређене функције случајних променљивх и њихове расподеле. Ми ћемо се овде, укратко, упознати са неким од тих функција и њиховим расподелама.

8.1. ЛИНЕАРНЕ ФУНКЦИЈЕ

Разматраћемо k-димензионалну случајну променљиву

$$X = (X_1, ..., X_k)$$

Линеарном функцијом случајних променљивих зовемо функцију дату у облику

$$L = a_1 X_1 + a_2 X_2 + ... + a_k X_k$$

при чему су $a_1, a_2, ..., a_k$ *, одређене константе.*

Означимо са

$$m_1,\ m_2, ..., m_k \tag{8.1}$$

очекиване вредности случајних променљивих $X_1, X_2, \ldots, X_k$, а са

$$W = \left[\sigma_{ij}\right] \qquad \begin{matrix} i = 1, 2, \ldots, k \\ j = 1, 2, \ldots, k \end{matrix}$$

њихову дисперсиону матрицу. При томе је

$$\sigma_{ij} = E\left[\left(X_i - m_i\right)\left(X_j - m_j\right)\right]$$

Прво ћемо одредити очекивану вредност и варијансу линеарне функције L дате са (8.1).

Кад је $(X_1, \ldots, X_k)$ прекидна случајна променљива са законом вероватноћа

$$(x_{1i_1}, x_{2i_2}, \ldots, x_{ki_k}); \qquad p_{i_1 i_2 \ldots i_k},$$

где је $i_1 = 1,2,\ldots;\ i_2 = 1,2,\ldots;\ i_k = 1,2,\ldots,$ онда је очекивана вредност линеарне функције

$$E(L) = \sum_{i_1=1} \ldots \sum_{i_k=1} \left[\sum_{j=1}^{k} a_j x_{ij}\right] p_{i_1 i_2 \ldots i_k} =$$

$$= a_1 \sum_{i_1=1} x_{1 i_1 - i_2 = 1} \left\{\sum \ldots \sum_{i_k=1} p_{i_1 i_2 \ldots i_k}\right\} +$$

$$+ \ldots + a_k \sum_{i_k=1} x_{k i_{k-i_2=1}} \left\{\sum \ldots \sum_{i_{k-1}=1} p_{i_1 i_2 \ldots i_k}\right\} = a_1 m_1 + \ldots + a_k m_k$$

односно,

$$E(L) = \sum_{i=1}^{k} a_i m_i \tag{8.2}$$

Кад је $(X_1, \ldots, X_k)$ непрекидна случајна променљива са законом вероватноћа

$$f(x_1, \ldots\ldots, x_k)$$

очекивана вредност линеарне функције ће бити

$$E(L)=\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}(\sum_{i=1}^{k}a_i x_i)\;\; f(x_1,\ldots,x_k)dx_1\ldots dx_k=\sum_{i=1}^{k}a_i m_i$$

Тако смо добили следећи резултат:

Очекивана вредност линеарне функције случајних променљивих $X_1, X_2, \ldots, X_k$ *је једнака истој линеарној функцији њихових очекиваних вредности.*

Варијанса линеарне функције L је једнака

$$\sigma^2(L)=E\{L-E(L)\}^2$$

односно,

$$\sigma^2(L)=E\left\{\sum_{i=1}^{k}a_i X_i-\sum_{i=1}^{k}a_i m_i\right\}^2=E\left\{\sum_{i=1}^{k}a_i(X_i-m_i)\right\}^2$$

Кад израз у загради квадрирамо и искористимо особине очекиване вредности добићемо да је

$$\sigma^2(L)=\sum_{i=1}^{k}a_i^2E(X_i-m_i)^2+\sum_{i\neq 1}^{k}\;\sum_{j\neq 1}^{k}a_i a_j E(X_i-m_i)(X_j-m_j)= \qquad (8.3)$$

$$=\sum_{i=1}^{k}a_i^2\sigma_i^2+\sum_{i\neq j}^{k}\;\sum_{i\neq j}^{k}a_{ij}\sigma_{ij}$$

Специјално, ако су $X_1, X_2, \ldots, X_k$ *независне случајне променљиве онда је варијанса линеарне функције једнака*

$$\sigma^2(L)=\sum_{i=1}^{k}a_i^2\sigma_i^2$$

тј. *варијанса линеарне функције променљивих, једнака је линеарној функцији њихових варијанси.*

Претпоставимо да су X и Y две независне случајне промењиве. Тада је њихов заједнички закон вероватноћа једнак производу маргиналних, тј.

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

Очекивана вредност производа две независне случајне промењиве биће једнака производу њихових очекиваних вредности, тј.

$$E(X \cdot Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x,y)dxdy =$$

$$\int_{-\infty}^{\infty} xf_1(x)dx \int_{-\infty}^{\infty} yf_2(y)dy = E(X) \cdot E(Y)$$

Лако се уверити да ће ова особина важити и за производ било које две функције независних случајних промењивих:

Очекивана вредност производа две функције независних случајних промењивих једнака је производу њихових очекиваних вредности, тј.

$$E\left[g(X)L(Y)\right] = E\left[g(X)\right] \cdot E\left[L(Y)\right] \tag{8.4}$$

Методом потпуне индукције се лако закључује да ова особина важи и за производ k-независних промењивих.

Резултат (8.4) применићемо на функцију генератрису. Нека су $g_1(t)$ i $g_2(t)$ функције генератрисе независних случајних промењивих X_1 i X_2 тј.

$$g_1(t) = E\left[e^{tX_1}\right]$$

$$g_2(t) = E\left[e^{tX_2}\right]$$

Тада је функција генератриса збира ових промењивих $X_1 + X_2$ једнака

$$g(t)=E\left[e^{(X_1+X_2)t}\right] = E\left[e^{X_1t}e^{X_2t}\right] = E\left[e^{X_1t}\right] \cdot E\left[e^{X_2t}\right] = g_1(t)g_2(t)$$

Потпуном индукцијом, овај се резултат може проширити и на *к* независних случајних променљивих.

Функција генератриса збира k независних случајних променљивих једнака је производу њихових функција генератрисе.

Посматрајмо случајну променљиву X_i која има Нормалну расподелу са параметрима m_i i σ_i^2. Функција генератриса за променљиву

$$Y = X_i - m_i$$

имаће облик

$$g_i(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty} e^{-\frac{y^2}{2\sigma_i^2}} dy \tag{8.5}$$

$$g_i(t) = e^{\frac{1}{2}t^2\sigma_i^2}$$

за $i = 1, 2, ..., k$.

Сад ћемо посматрати независне случајне променљиве $X_1, X_2, \ldots, X_k$ са Нормалним расподелама. Фунцкија генератриса за

$$Y = a_1(X_1 - m_1) + a_2(X_2 - m_2) + \ldots + a_k(X_k - m_k)$$

биће једнака производу функција генератриса (8.5)

$$g(t) = g_1(a_1t)g_2(a_2t)\ldots g_k(a_kt) = \exp\left\{\frac{1}{2}t^2\left[a_1\sigma_1^2 + a_2\sigma_2^2 + \ldots + a_k\sigma_k^2\right]\right\}$$

Очекивана вредност случајне променљиве Y је једнака нули, а њена варијанса је

$$\sigma^2 = a_1\sigma_1^2 + a_2\sigma_2^2 + \ldots + a_k\sigma_k^2$$

па се функција генератриса може написати у облику

$$g(t) = e^{\frac{1}{2}t^2\sigma^2}$$

одакле следи тврђење: ***Случајна променљива Y има Нормалну расподелу***.

Линеарна функција $L = a_1 X_1 + ... + a_k X_k$, случајних промнељивих $X_1, X_2, ..., X_k$ има очекивану вредност $E(L) = a_1 m_1 + ... + a_k m_k$, и варијансу $\sigma^2(L) = a_1^2 \sigma_1^2 + ... + a_k^2 \sigma_k^2$.

Случајна променљива L - $E(L)$ се своди на функцију

$$L - E(L) = a_1 (X_1 - m_1) + ... + a_k (X_k - m_k),$$

Ако су $X_1, X_2, ..., X_k$, независне случајне променљиве са Нормалним расподелама, онда смо видели да ће њена функција генератриса имати облик

$$g(t) = e^{\frac{1}{2} t^2 \sigma^2 (L)}$$

што значи:

Линеарна функција

$$L = a_1 X_1 + ... + a_k X_k$$

k независних случајних променљивих са Нормалном расподелом, имаће такође Нормалну расподелу са очекиваном вредношћу

$$E(L) = a_1 m_1 + ... + a_k m_k$$

и варијансом

$$\sigma^2(L) = a_1^2 \sigma_1^2 + ... + a_k^2 \sigma_k^2$$

где смо са $m_1, .., m_k$ *и* $\sigma_1^2, .., \sigma_k^2$ *означили очекиване вредности и варијансе случајних променљивих* $X_1, X_2, ..., X_k$.

8.2. КВАДРАТНА ФУНКЦИЈА

Познато је да је *Hi*-квадрат расподела одређена законом вероватноћа

$$k_n(x) = C_n x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad \text{za } x \geq 0$$

при чему је C_n константа, тако одређена да је интеграл

$$\int_0^{\infty} k_n(x)dx = 1$$

Функција генератиса, за овакву случајну променљиву је функција

$$g(t) = E(e^{tX}) = C_n \int_0^{\infty} e^{tX} x^{\frac{n}{2}-1} e^{\frac{x}{2}-1} dx =$$

$$= C_n \int_0^{\infty} x^{\frac{n}{2}-1} e^{\frac{(1-2t)x}{2}} dx$$

Уводећи смену

$$y = (1-2t)x$$

посматрани интеграл се своди на

$$g(t) = (1-2t)^{-\frac{n}{2}} C_n \int_0^{\infty} y^{-\frac{n}{2}-1} e^{-\frac{y}{2}} dy$$

односно,

$$g(t) = (1-2t)^{-\frac{n}{2}}$$

јер је

$$C_n \int_0^{\infty} y^{-\frac{n}{2}-1} e^{-\frac{y}{2}} dy = 1$$

Дакле, функција генератриса случајне променљиве X која има Ні-квадрат расподелу са n степени слободе дата је функцијом

$$g(t) = (1-2t)^{-\frac{n}{2}} \tag{8.6}$$

А сад ћемо посматрати *n* независних случајних променљивих $X_1,\ldots,X_n$ које имају стандардизовану Нормалну расподелу *N(0;1)*. Са *X* ћемо означити случајну променљиву која представља збир квадрата случајних променљивих $X_1,\ldots,X_n$.

Кад су $X_1,X_2,\ldots,X_n$ *независне случајне променљиве са Нормалном расподелом N(0;1), онда случајна променљива*

$$X = X_1^2 + X_2^2 + \ldots + X_n^2$$

има Ні-квадрат расподелу са n степени слободе.

Доказ. Потражићемо функцију генератрису за случајну променљиву

$$Y = X_i^2$$

То ће бити функција

$$g_i(t) = E(e^{Yt}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{x^2 t} \cdot e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(1-2t)x^2}{2}} dx$$

Уводећи смену $(1-2t)x = y$ и користећи чињеницу да је интеграл

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1$$

лако се добије да је

$$g_i(t) = (1-2t)^{-\frac{1}{2}}$$

одакле се закључује: Случајна променљива $Y = X_i^2$ има Ні-квадрат расподелу са једним степеном слободе.

По претпоставци су $X_1, X_2, \ldots, X_n$ независне са Нормалном расподелом, а то значи да су и $X_1^2, X_2^2, \ldots, X_n^2$, такође независне променљиве. Функција генератриса њиховог збира биће једнака производу њихових функција генератриса. Зато је, за случајну променљиву

$$X = X_1^2 + X_2^2 + \ldots + X_n^2$$

функција генератриса једнак

$$g(t) = g_1(t) g_2(t) \ldots g_n(t) = \prod_{i=1}^{n} (1-2t)^{-\frac{1}{2}}$$

односно,

$$g(t) = (1-2t)^{-\frac{n}{2}}$$

а то је уствари функција генератриса *Hi*-квадрат расподеле. Тиме је доказано наведено тврђење.

Претпоставимо да две променљиве X и Y имају *Hi*-квадрат расподелу са n и m степени слободе и да су међусобно независне. *Тада њихов збир*

$$Z = X + Y$$

има такође Hi-квадрат расподелу са $(n+m)$ *степени слободе.*

Доказ. Непосредно из функције генератрисе за Z, следи наведено тврђење.

8.3. СТУДЕНТОВ КОЛИЧНИК

За случајну променљиву X кажемо да има Студентову t-расподелу, ако је њен закон вероватноћа дат функцијом

$$s_n(x) = C_n(1+\frac{x^2}{n})^{-\frac{n+1}{2}}, \quad \text{za} \quad x \in (-\infty;\ \infty)$$

Посматраћемо случајну променљиву Y са Нормалном расподелом $N(0;1)$ и случајну променљиву Z која има Hi-квадрат расподелу са n степени слободе. Нека су то независне случајне променљиве.

Тада случајна променљива

$$X = \sqrt{n}\frac{Y}{\sqrt{Z}}$$

има Студентову t-расподелу са n степени слободе.

Доказ. Закон вероватноћа за дводимензионалну случајну променљиву (x, z), дат је функцијом

$$f(y,z) = C\, y^{\frac{n}{2}-1} e^{-\frac{y}{2}} e^{-\frac{y^2}{2}} \tag{8.7}$$

Функција расподеле случајне променљиве X дефинисана је преко вероватноће

$$F(x) = P\{X \le x\};$$

За посматрани случај, то ће бити функција

$$F(x) = P\left\{\sqrt{n}\frac{Y}{\sqrt{Z}} \le x\right\} = P\left\{Y \le \frac{x\sqrt{Z}}{\sqrt{n}}\right\} = \iint f(y,z)dydz,$$

$$y \le \frac{x\sqrt{z}}{\sqrt{n}}$$

при чему y узима вредности из интервала $(-\infty;\ \infty)$, z из интервала $(0;\ \infty)$. Замењујући y горњи интервал функцију $f(y,z)$ дату са (8.7), добићемо

$$F(x)=C\int_0^{\infty}\left[y^{\frac{n}{2}-1}e^{-\frac{y}{2}}\int_{-\infty}^{\frac{x\sqrt{y}}{\sqrt{n}}}e^{-\frac{z^2}{2}}dz\right]dy$$

Уводећи смену

$$z=\frac{t\sqrt{y}}{\sqrt{n}},\quad dz=\sqrt{\frac{y}{n}}\,dt$$

добија се

$$F(x)=C\int_0^{\infty}\left[y^{\frac{n-1}{2}}e^{\frac{y}{2}}\int_{-\infty}^{x}e^{-\frac{t^2y}{2n}}dt\right]dy$$

односно

$$F(x)=C\int_{-\infty}^{x}\left[\int_0^{\infty}y^{\frac{n-1}{2}}e^{\frac{y}{2}(1+\frac{t^2}{n})}dy\right]dt$$

После смене

$$u=\frac{y}{2}(1+\frac{t^2}{n}),\quad du=\frac{1}{2}(1+\frac{t^2}{n})dy$$

добије се

$$F(x)=C_1\int_{-\infty}^{x}\left[(1+\frac{t^2}{n})^{-\frac{n-1}{2}-1}\int_0^{\infty}u^{\frac{n-1}{2}}e^{-u}du\right]dt$$

$$F(x)=C_n\int_{-\infty}^{x}(1+\frac{t^2}{2})^{-\frac{n+1}{2}-1}\,dt$$

јер интеграл

$$\int_0^{\infty} u^{\frac{n-1}{2}} d^{-u} du$$

има коначну вредност. Функција расподеле може се написати у облику

$$F(x) = C_n \int_{-\infty}^{x} (1+\frac{t^2}{2})^{-\frac{n+1}{2}} dt$$

а то није ништа друго него функција расподеле случајне променљиве која има Студентову *t*-расподелу са *n* степени слободе. Тиме је завршен доказ наведеног тврђења.

Нека су случајне променљиве

$$Y, X_1, X_2, \ldots, X_n$$

независне случајне променљиве са Нормалном расподелом *N(0;1)*.

Студентов количник *је случајна променљива*

$$X = \frac{Y}{\sqrt{\frac{1}{n}\left[X_1^2 + \ldots + X_n^2\right]}}$$

која има Студентову t-расподелу са n степени слободе.

8.4. FISHER – SNEDEKOR-ОВ F – КОЛИЧНИК

Случајна променљива X има F-расподелу са n_1 и n_2 степени слободе ако је њен закон вероватноће дат функцијом.

$$f_{n_1;n_2} = C_{n_1;n_2} x^{\frac{n_1}{2}-1} (n_2 + n_1 x)^{\frac{n_1+n_2}{2}-1}, \text{ за } x \in (0;\infty)$$

Претпоставимо да су случајне променљиве Y и Z независне и да имају *Hi*-квадрат расподелу са n_1 и n_2 степени слободе. Посматраћемо случајну променљиву као количник ове две променљиве

$$X = \frac{n_2}{n_1}\frac{Y}{Z}$$

Заједнички закон вероватноћа за (Y,Z) биће функција

$$f(y,z) = C\, y^{\frac{n_1}{2}-1} z^{\frac{n_2}{2}-1} e^{-\frac{y+z}{2}}$$

Уводећи смену

$$x = \frac{n_2}{n_1}\frac{y}{z}, \quad u = z,$$

у посматрани закон вероватноћа добиће се

$$g(x,u) = C_1 x^{\frac{n_1}{2}-1} u^{\frac{n_1+n_2}{2}} e^{-\frac{u}{2}\left(1+\frac{n_1}{n_2}x\right)}$$

који представља заједнички закон вероватноћа за X и U. Интегришући ову функцију по u у границама од нула до бесконачности, добићемо маргинални закон вероватноћа за променљиву X у облику

$$f_{n_1;n_2} = C_{n_1;n_2} x^{\frac{n_1}{2}-1} (n_2 + n_1 x)^{-\frac{n_1+n_2}{2}}, \text{ за } x \in (0;\infty)$$

тако да се може тврдити:

Ако су Y и Z две независне случајне променљиве са Hi-квадрат расподелама, онда количник

$$X = \frac{n_2}{n_1}\frac{Y}{Z}$$

има F-расподелу са n_1 *и* n_2 *степени слободе.*

Посматрајмо две групе независних случајних променљивих

$$X_1, X_2, \ldots, X_{n_1}$$

$$Y_1, Y_2, \ldots, Y_{n_2}$$

Нека свака од њих има Нормалну расподелу *N(0;1).*

Тада количник

$$X = \frac{\frac{1}{n_1}(X_1^2 + \ldots + X_{n_1}^2)}{\frac{1}{n_2}(Y_1^2 + \ldots + Y_{n_2}^2)}$$

има F-расподелу са n_1 *и* n_2 *степени слободе.*

ЛИТЕРАТУРА

1. Andersosn, T. W. [1958] An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, Ltd., NewYork.
2. Bartlet S. [1933] On the Theory of Statistical Regresion, Proc. R. Soc., 53, Edingburgh
3. Blalock H. M. [1979] Social Statistics, McGraw-Hill, New York
4. Cochran, W. G. [1972] Sampling Techniques, Secondedition, WilleyEastern Private Limited, NewDelhi.
5. Cramer H. [1946] Mathematical Metods of Statistics, University Press, Princenton.
6. Cramer H. [1955] The Elements of Probability Theory and Some of its Applications, JohnWiley & Sons, Ltd., New York.
7. Cuthbertson, K., S. G. Hall, M. P. Taylor [1992] Applied Econometrics Techniques, Phillip Allan, University of Michigan Press.
8. Dillon, W. R., M. Goldstein [1984] Multivariate Analysis: Methods and Applications, JohnWiley & Sons, Ltd., New York.
9. Draper N. R., Smith H. [1996] Applied Regression Analysis, J. Wiley., New York
10. Dugue D. [1973] Appliation des propriete de la limite an sens du calcul des probabilites a l'etude de diverses questions d' estimation, Journ. De l'Ec. Polytechn., Paris.
11. Enders, W. [1995] Applied Econometric Time Series, A User's Guide, John Wiley & Sons, Ltd., New York.

12. Feller W. [1968] An Introduction to Probability Theory and its Applications, Vol 1, J. Wiley & Sons, Ltd., New York.
13. Fisher A. R. [1954] Statistical Methods for Research Workers, Oliver & Bojd, Edinburg.
14. Gibbons J. D. [1992] Nonparametric Statistical Inference, Marcel Dekker.
15. Gnedenko V. B. [1969] Kurs teorii veroathnosteii, Moskva.
16. Gosset S. W. [1908] "Student": The Probable Error of a Mean, Biometrika, V. 6.
17. Grais, B. [1983] Exercices de statistique descrptive, Dunod, Paris.
18. Grđić, G., M. Eremić, D. Mladenović [1980] Ekonomsko-statistička analiza, Ekonomski fakultet, Beograd.
19. Griffiiths, W., R. Carter Hill, G. Judge [1993] Learning and Practicing Econometrics, JohnWiley & SonsLtd., New York.
20. Group of Authors [2006] e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook
21. Hartwig F, B. E. Dearing [1985] Exploratory Data Analysis, SagePublications, Beverly Hills.
22. Harvey, A. [1990] The Econometric Analysis of Time Series, 2nd ed., MIT Press.
23. Ivanović B. [1959] Teorija korelacije sa primenom u ekonomskim istraživanjima, Beograd.
24. Ivanović B., [1973] Teorijska statistika, Naučna knjiga, Beograd.
25. Ivković, Z [1980] Matematička statistika, Naučna knjiga, Beograd.
26. Ivković, Z [1988] Verovatnoća i matematička statistika, Naučna knjiga, Beograd
27. Johnston, J. [1988] Econometrics Methods, McGrawHill.
28. Juran M. J. [1962]Quality Control Handbook, New York

29. Kish L. [1965] Survey Sampling, JohnWiley andSons Inc., New York, London, Sydney.
30. Kish, L. [1987] Statistical Design for Research, John Wiley & Sons, New York.
31. Klaemer H. C. [1982]Kappa Coefficient, In: Encyclopedic ofStatistical Sciences, S. Kotz andN. L. Jonson, John Willyand Sons, New York.
32. Klainbaum D. G., Kupper L. L., Mogensferin H. [1982] Epidemiological Research Principles and Quantitative Methods, Lifetime Learnings Publishing.
33. Kovačić, Z. [1995] Analiza vremenskih serija, Ekonomski fakultet, Beograd.
34. Kraft H. C., Eeden VanC. [1968] A Nonparametric Introduction to Statistics, New York.
35. Krčevinac, S. [1993] Ekonometrijske metode, skripta, Fakultet organizacionih nauka, Beograd.
36. Lehmann L. E. [1959] Testing Statistical Hypotheses, New York.
37. Mann, P. S. [1995] Statistics for Business and Economics, JohnWiley & Sons, Ltd.
38. Markland, A. [1988] Quantitative Methods for Management Decisions, JohnWiley & Sons, Ltd.
39. Markus, G. B. [1984] Analyzing Panel Data, Sage Publications, Bevery Hills.
40. Merkle, M. [2001] Verovatnoća i statistika, Akademska misao, Beograd
41. Mills C. F. [1956] Introduction to Statistics, New York.
42. Milošević, V. [1983] Teorijska statistika, Naučna knjiga, Beograd.
43. Milošević, V. [2000] Teorija verovatnoće, Zavod za udžbenike, 2000.

44. Mises VonR. Mathematical Theory of Probability and Statistics, NewYork.
45. Mladenović, P. [2002] Verovatnoća i statistika, Matematički fakultet, 2002.
46. Norušis M. J. [1993] SPSS for Windows Base System User's Guide, SPSS Inc., Chicago
47. Norusis M.J. [1990] SPSS/PC+ v4.0 Advanced Statistica, SPSS Inc, Chicago.
48. Norušis, M.J. [1990] SPSS/PC+ 4.0 Base Manual, SPSS Inc., Chicago.
49. Ostrom, C. W., Jr. [1978] Time Series Anaylis: Regression Tecniques, Sage Univeristy Paper serieson Quantitative Applications inthe SocialSciences, SageUniveristy Paper, Beverly Hills/London.
50. Pearson K. [1900] On A Criterian That A System Of Deviations From The Probable In The Case Of A Correlated System Of Variables Is Such That It Can Ve Reasonable Supposed To Have Arisen In Random Sampling, Phil. Mag. V. 50.
51. Pindyck, R. S., D. L. Rubinfeld [1976] Econometric Models and Ecnonomic Forecasts, McGraw-Hill, New York.
52. Prohorov V. A., Rozanov A. [1973] Teoria veroathnosteii, Osnoviponotio, Predillnie teoremi, Slucaiinie processi, Moskva.
53. Rao R. C. [1968] Linear Statistical Inference and its Applications, New York, Prevod "Nauka", Moskva.
54. Searle S. R. [1971]Linear Models, John Wiley & Sons, New York
55. Sharma, S. [1996] Applied Multivariate techniques, JohnWiley & Sons, Ltd., New York
56. Siegel S. [1956] Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill, New York.

57. Storm R. [1970] Teoria verovathnosteii matematičeskaa statistika statističeskii kontrola kačestva, Prevodsa nemačkog, Moskva.
58. Theil H. [1971] Principals of Econometrics, JohnWiley & Sons, Ltd., New York.
59. Vuković N. [1978 i 1981] Verovatnoća i statistika, Beograd.
60. Vuković N. [1987] Statistička analiza, Naučna knjiga, Beograd.
61. Vuković, N. [2000 i 2002] PC Verovatnoća i statistika, FON, Beograd
62. Wilks S. [1967] Matematičeskaa statistika, Prevodsa engleskog, Moskva.
63. Wonnacott, R.J., T. H. Wonnacott [1979] Econometrics, 2 nd Edition, John Wiley & Sons, Ltd., NewYork
64. Zacks S. [1971] The Theory of Statistical Inference, NewYork, Moskva.
65. Žarković, S. S. [1976] Upotreba statističkih podataka, Savezni zavodza statistiku, Beograd.
66. Zečević, T., M. Kovačević, M. Kovačević [1991] Teorija uzoraka i planiranje eksperimenata, Ekonomski fakultet, Beograd.