Съдържание

Теория на вероятностите

I . Вероятност

1. Случайни събития

Ще въведем основните понятия с които създаваме математическия модел, описващ случайността.

Случаен експеримент или случаен опит наричаме такъв експеримент, изходът от който не е предварително известен, не е детерминиран, не може да се предвиди. Например, хвърляне на монета.

Естествено, може да допуснем, че изходът е неопределен само защото ние не разполагаме с достатъчна информация за опита. Ако знаем, силата на хвърляне, тежестта на монетата, височината на падане, въздушните течения и т.н. ще предвидим крайния резултат. На практика много рядко познаваме всички фактори, които влият на една система, а моделът базиран на случайни събития доста точно описва действителността. Освен това съществуват чисто случайни явления поне на ниво елементарни частици, и те се пренасят в макро света в който живеем. Без да навлизаме дълбоко във философски спорове, ще приемем, че светът е недетерминиран и в него съществува случайността.

Резултата от един случаен експеримент наричаме **елементарно събитие** - ω (омега).

Внимание!
Важни понятия.

Всевъзможните изходи от случайния експеримент наричаме **пространство на елементарните събития**. Означението, което ще използваме е $\Omega = \{\omega_1, \omega_2, \dots \omega_n\}$. След изпълнението на случаен експеримент се сбъдва едно единствено елементарно събитие ω_i.

Подмножествата на Ω наричаме **събития** и обикновено ги бележим с гловните букви - $A, B \subset \Omega$. Казваме, че събитието A се е изпълнило или сбъднало, ако се е изпълнило някое елементарно събитие, което му принадлежи. По този начин събитието се определя чрез възможните изходи от експеримента, които го съставляват.

По-долу ще дадем някои примери за пространства на елементарните събития.

Пример - Зар

При хвърляне на правилен зар $\mathbf{\Omega} = \{\omega_1, \omega_2, \ldots, \omega_6\} = \{1, 2, \ldots, 6\}$. Случайни събития са, например:
$A = \{$ пада се четно число$\} = \{2, 4, 6\}$
$B = \{$ пада се повече от две$\} = \{3, 4, 5, 6\}$

Пример - Два зара

При хвърляне на два различими зара пространството на елементарните събития е: $\mathbf{\Omega} = \{(i, j),\ i, j = 1, \ldots, 6\}$ или изписано подробно:

$$\mathbf{\Omega} = \left\{ \begin{array}{lllll} (1,1) & (1,2) & (1,3) & \ldots & (1,6) \\ (2,1) & (2,2) & (2,3) & \ldots & (2,6) \\ \ldots & & & & \\ (6,1) & (6,2) & (6,3) & \ldots & (6,6) \end{array} \right\}$$

Случайни събития са:
$A = \{$ Сумата от точките е $4\} = \{(1,3), (2,2), (3,1)\}$
$B = \{$Точките върху двата зара съвпадат$\} = \{(1,1), \ldots, (6,6)\}$

Пример - Време на живот

Ако изходът от екперимента е измерване на времето на безотказна работа, наречено още "време на живот" на дадено устройство, то тогава пространството на елементарните събития се описва така $\mathbf{\Omega} = \{x : 0 \leq x < \infty\}$, x е в часове. Случайно събитие е:
$A = \{$ Устройството работи поне 6 часа$\} = \{x > 6\}$

Доколкото събитията са множества, правилата за работа със събития съвпадат с операциите над множества:

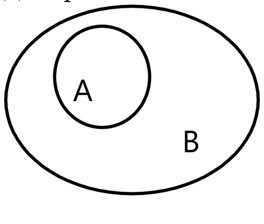
❧ $A \cap B$ сечение на събитията A и B наричаме събитие , което се сбъдва, когато се сбъдне елементарно събитие принадлежащо и на двете, т.е. при **едновременното сбъдване** на събитията A и B. В теория на вероятностите знакът за сечение често се изпуска. Пишем просто AB.

❧ $A \cup B$ обединение на събитията A и B наричаме събитие , което се сбъдва, когато се **сбъдва поне едно** от събитията A или B.

❧ $\bar{A}$ допълнение на събитието A (или противоположно събитие) наричаме събитие, което се сбъдва, когато **не се сбъдва** събитието A.

❧ $A \subset B$ винаги когато се изпълнява A се изпълнява и B, т.е. от A **следва** B.

Събитията често се представят чрез така наречените диаграми на Вен.



$A \subset B$

Ще запишем и основните закони при операции със събития. Доказателства на тези закони са тривиални, извършват се в стил характерен за теория на множествата. Взима се едно елементарнто събитие принадлежащо на лявата страна на равенството и се доказва, че то принадлежи и на дясната страна, и обратно. Затова ще докажем само законите на Де Морган.

Закони за работа със събития:

- Комутативен

$$A \cup B = B \cup A \qquad \text{и} \qquad A \cap B = B \cap A$$

- Асоциативен

$$A \cup (B \cup C) = (A \cup B) \cup C \qquad \text{и} \qquad A \cap (B \cap C) = (A \cap B) \cap C$$

- Дистрибутивен

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \qquad \text{и} \qquad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

- Закон на Де Морган

$$\overline{A \cup B} = \bar{A} \cap \bar{B} \qquad \text{и} \qquad \overline{A \cap B} = \bar{A} \cup \bar{B}$$

Док.Нека $\omega \in \overline{A \cup B}$, т.е. ω не принадлежи на обединението на A и B. Но това означава, че елементарното събитието ω не принадлежи нито на A, нито на B, защото в противен случай би принадлежало на обединението им. ω не принадлежи на A, означава че $\omega \in \bar{A}$, аналогично $\omega \in \bar{B}$ и следователно $\omega \in \bar{A} \cap \bar{B}$. Аналогично се доказва и другият закон. ✓

2. Класическа вероятност

Нека разглеждаме прост експеримент, при който има няколко възможни изхода и те се случват с една и съща честота. Например, хвърляме правилна монета или зар. Естествено, при хвърляне на монета ще наблюдаваме падането на "ези"приблизително толкова често, колкото и падането на "тура". Тогава, интуитивно най-подходящо би било да дефинираме вероятността за падане на "ези"като $1/2$, т.е. средно при половината случаи на хвърляне на монета ще се изпълнява това събитие. Аналогично при хвърляне на зар, шестица ще се пада в около една шеста от всички случаи, затова естествено би било да дефинираме вероятността за $1/6$.

Изглежда първи в тази област е работил италианския математик Джироламо Кардано (1501-1575). Страстатта му към хазарта го мотивира да напише 'Книга за игрите на късмета', публикувана посмъртно.

Така достигаме до класическата дефиниция за вероятност. Наричаме я "класическа"защото погледнато исторически тя е възникнала първа.

За целите на този параграф ще смятаме, че Ω е крайно множество и всички изходи са равновероятни.

В урна има 5 бели и 3 черни топки. От урната се тегли една топка. Любопитно е, че значителен процент от хората смятат вероятността за изтегляне на бяла топка за 1/5. Ако и вие сте сред тях, нашия съвет е да затворите тази книга и да си потърсите друго поприще в живота.

Дефиниция - Класическа вероятност

Вероятност $P(A)$ на дадено събитие A, дефинираме като отношението на броя на благоприятните случай (тези при които A се изпълнява), към броя на всички възможни случай, т.е.

$$\mathbf{P}(A) = \frac{\text{брой благоприятни случай}}{\text{брой всички случай}} = \frac{\# A}{\# \Omega}$$

Ясно е, че така дефинираната вероятност е реално число, при това $0 \leq \mathbf{P}(A) \leq 1$. Прието е събитията за които $P(A) = 0$ да се наричат **невъзможни събития**, те никога не се сбъдват. Събитията с $P(A) = 1$ се наричат **сигурни** или **достоверни** събития, те пък винаги се изпълняват.

Пример - Два зара

В примера с двата различими зара лесно се пресмята вероятността на събитието

$B = \{\text{Точките върху двата зара съвпадат}\} = \{(1,1), \ldots, (6,6)\}$

Всички възможни случай при хвърлянето на заровете са 36, събитието се изпълнява в 6 от тях, тогава

$$\mathbf{P}(B) = \frac{6}{36}$$

Виж Зад.1.19

Макар моделът на класическата вероятност да изглежда прост, задачите които се пораждат от него далеч не винаги са тривиални. Понякога резултатът противоречи на интуицията и тогава говорим за парадокс. Някой велики математици, например Лайбниц и Даламбер, също са се заблуждавали решавайки задачи свързани с класическа вероятност.

1663г.

Ще разгледаме един известен парадокс. При хвърляне на правилен зар всяко едно от числата $1, 2, \ldots, 6$ се пада с равна вероятност. Ако хвърляме два зара, сумата е число между 2 и 12. Сума 9 се получава по два начина - $9 = 3 + 6 = 4 + 5$. Сума 10 също се получава по два начина - $10 = 4 + 6 = 5 + 5$. Защо тогава при извършване на опита сума 9 се пада по-често отколкото сума 10?

Грешката идва от начина, по който въвеждаме елементарните събития. При хвърляне на два зара сума 9 се получава не при две елементарни събития, а при четири $(3,6), (6,3), (4,5)$ и $(5,4)$. Докато за сума 10 елементарните събития са три $(4,6), (6,4)$ и $(5,5)$. Трябва да гледаме на заровете като на различни, те са физически различни и се държат като различни, дори и ние да не ги различаваме. Ако в случая със сума 10, работим само с елементарните събития $(4,6)$ и $(5,5)$ е нарушено основното изискване елемнтарните събития да са равновероятни, тъй като $(4,6)$ се пада два пъти по-често от $(5,5)$.

Аналагично при хвърляне на три зара, сума 9 и 10 също се получават по равен брой начини и в двата случая шест:

$$9 = 1+2+6 = 1+3+5 = 1+4+4 = 2+2+5 = 2+3+4 = 3+3+3,$$
$$10 = 1+3+6 = 1+4+5 = 2+2+6 = 2+3+5 = 2+4+4 = 3+3+4,$$

но отново честотата на падане не е еднаква. Съобразете коя сума е по вероятна.

Дефиницията за класическа вероятност може да се разшири и за случай, в които множеството $\mathbf{\Omega}$ не е крайно, а изброимо. Например, ако се опитваме да пресметнем вероятността да се падне четно число при случаен избор измежду множеството на естествените числа. Интуитивно ясно е, че би следвало тази вероятост да е 0.5, тъй като половината естествени числа са четни.

Естествени са числата: $1, 2, 3, \ldots$

За решаването на подобни проблеми първоначално $\mathbf{\Omega}$ се ограничава да съдържа само N елемента. Пресмята се класическата вероятност на търсеното събитие A, тази вероятност разбира се ще зависи от N, т.е. пресмятаме $\mathbf{P}_N(A)$. След което се извършва граничен преход по $N \to \infty$

$$\mathbf{P}(A) = \lim_{N\to\infty} \mathbf{P}_N(A) = \lim_{N\to\infty} \frac{\#A_N}{\#\mathbf{\Omega}_N},$$

стига разбира се тази граница да съществува.

3. Геометрична вероятност

Класическата вероятност се използва когато се разглеждат елементарни опити, като хвърляне на зар, теглене на карта или избор на топка от урна, т.е. тогава когато възможните събития са краен или най-много изброим брой. Но този модел е неприложим, ако разглеждаме по-сложни събития, който се описват с реални числа. Например, случаен момент от време, или случайно място в пространството. Затова се налага дефинирането на вероятност в по общ случай.

Когато множеството от събития $\mathbf{\Omega}$ е неизброимо, разбира се не може да се говори за брой на елементарните събития. Тогава, за дефиниране на вероятност вместо „брой“ се използва “мярка”- μ, като под мярка разбираме евклидовата мярка. Така, ако множеството е едномерно мярката е дължината, при двумерни множества използваме лицето, в тримерни обема и т.н

Дефиниция - Геометрична вероятност

Нека $\mathbf{\Omega}$ е някакъв геометричен обект. Тогава вероятността на събитието $A \subset \mathbf{\Omega}$ се дефинира като

$$\mathbf{P}(A) = \frac{\mu(A)}{\mu(\Omega)}$$

За да бъде моделът на геометричната вероятност смислен би следвало да поискаме, също както при класическа вероятност, елементарните събития да са с еднаква честота. В случая, това би означавало вероятността да попаднем във всяка точка на пространството да е една и съща, т.е. вероятността да попаднем в някое множество да зависи само от мярката на множеството, а не от неговото положение, форма и т.н.

Пример

По график автобус се движи на интервали от 15 минути. Ще пресметнем вероятността да чакаме пристигането на автобуса по-малко от 5 минути.

Ние пристагаме на спирката в случаен момент, спрямо графика на автобуса. Нека с X означим времето за чакане. Ясно е, че X е реално число в интервала $[0, 15]$, при това вероятността да вземе коя да е стойност в този интервал е една и съща. Тогава $\boldsymbol{\Omega} = \{X : 0 \leq X \leq 15\}$ и събитието търсено в задачата е $A = \{X < 5\}$.

$\boldsymbol{\Omega}$ е едномерно множество, следователно мярката μ е дължината.

$$\mathbf{P}(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{5}{15}$$

В равнината са прекарани успоредни линии на разстояние L една от друга. Върху равнината се пуска игла с дължина k, като $k < L$. Каква е вероятността иглата да застъпи някоя от линиите?

В началото трябва да въведем подходящ математически модел описващ опита. Положението на иглата в равнината се определя от нейното изместване спрямо линиите, както и от ъгъла, който сключва с тях.

Нека X е острият ъгъл, който иглата сключва с линиите. Ясно е, че $0 \leq X \leq \pi/2$.

Нека Y е разстоянието от средата на иглата до по-близката линия. Тъй като разстоянието между линиите е L, то Y не може да бъде по-голямо от $L/2$.

Така получаваме двумерно пространство на елементарните събития.

$$\boldsymbol{\Omega} \;:\; \left| \begin{array}{l} 0 \leq X \leq \pi/2 \\ 0 \leq Y \leq L/2 \end{array} \right.$$

Мярката, която ще използваме за намиране на вероятността е лицето.

Сега трябва да установим условията при които иглата ще застъпва линия. Ако със Z означим отклонението на края на иглата спрямо нейната среда. То при $Z < Y$ иглата няма да достига до по-близката линия, т.е. няма да имаме пресичане. Докато при $Z > Y$ иглата ще настъпи линия.

Z лесно може да се изрази чрез дължината на иглата и ъгъла и спрямо линиите. Наистина $Z = k/2 \sin X$. Тогава събитието A, че иглата настъпва линия може да се запише като

$$A \;:\; Y < k/2 \sin X$$

Съгласно дефиницията за геометрична вероятност

$$\mathbf{P}(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{\text{лицето на защрихованата част}}{\text{лицето на правоъгълника}} =$$

$$= \frac{\int_0^{\pi/2} \frac{k}{2} \sin x \, dx}{\frac{\pi}{2} \frac{L}{2}} = \frac{\frac{k}{2}}{\frac{\pi L}{4}} = \frac{2k}{\pi L}$$

Тази задача дава възможност за експериментално пресмятане на π. Ако извършваме опита многократно и записваме броя T на застъпване на линия и броя на опитите N, то $\frac{T}{N} \to \mathbf{P}(A)$. Следователно за $N \to \infty$

$$\pi = \frac{2kN}{TL}$$

Подобни методи на пресмятане се наричат Монте Карло.

4. Аксиоматично дефиниране на вероятността

Моделът на класическата вероятност се използва когато се разглеждат елементарни опити, като хвърляне на зар, теглене на карта или избор на топка от урна. Но този модел е неприложим, ако елементарните събития не са равновероятни. Например, хвърля се зар, който поради нарушен баланс по-често пада на едната си страна. Затова се налага дефинирането на вероятност в по общ случай.

В 1900г Хилберт формулира все още нерешените проблеми в математиката. Един от тях е аксиоматично изграждане на Теория на вероятностите.
Задачата е изпълнена от Колмогоров в XX век. Той създава съвременната абстрактна теория на вероятностите, като предлага аксиоматика базирана на теория на мярката и лебеговия интеграл.

Добре е ако можем да регистрираме като резултат от експеримента всяко елементарно събитие. Не винаги е така.

Пример

Хвърляме две еднакви и симетрични монети. Всяка монета има две възможности E за ези и T за тура с равни вероятности. Множеството на елементарните събития се състои от четирите варианта за двете монети $\Omega = \{(E,E),(E,T),(T,E),(T,T)\}$. Събитията, които можем да наблюдаваме са три. Двете събития $\{(E,E)\}$ и $\{(T,T)\}$ са от по едно елементарно събитие и са с вероятност по $\frac{1}{4}$. Третото $\{(E,T),(T,E)\}$ е от две елементарни събития, които наблюдателят не може да различи. Вероятността му е $\frac{1}{2}$. Трите събития изчерпват всички възможности и образуват разделяне.

Също така, ако Ω е неизброимо множество, например интервал върху реалната права, няма да е възможно да се въведе вероятност за всички събития, т.е за всички подмножества. Това е известен проблем в теория на мярката.

Поради тези причини, вероятност се въвежда не върху всички събития, а само върху един специален клас, който ще дефинираме по-долу.

Най-простата възможна σ-алгебра е $\mathcal{A} = \{\emptyset, \Omega\}$. Наричаме я тривиална, тя е безинтересна.
σ-алгебрата породена от едно единствено събитие е $\mathcal{A} = \{\emptyset, A, \overline{A}, \Omega\}$. Не е трудно да се съобрази, че това наистина е σ-алгебра, т.е. всевъзможните сечения, обединения и допълнения на тези събития водят до същите събития.

Дефиниция - Сигма алгебра

Нека $\mathbf{\Omega}$ е произволно множество и нека $\mathcal{A}$ е съвкупност от подмножества на $\mathbf{\Omega}$. Казваме, че $\mathcal{A}$ е **σ-алгебра**, ако са изпълнени условията:

- ❧ $\emptyset \in \mathcal{A}$ и $\Omega \in \mathcal{A}$
- ❧ Ако $A \in \mathcal{A}$, то $\overline{A} \in \mathcal{A}$
- ❧ Ако $A_1, A_2, \ldots \in \mathcal{A}$, то $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ и $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$

Разбира се тази дефиниция е преопределена, не е необходимо да се включват едновременно изброимите сечения и обединения. Ако използваме само обединенията и допълнителните събития, то по закона на Де Морган ще получим сеченията, както и обратно.

Сигма алгебрата е клас, затворен относно операциите: допълнение, обединение и сечение. Това ни гарантира, че ако знаем вероятността на някои събития то ще знаем и вероятността на техните сечения, обединения и т.н.

Прието е функциите, които на множество съпоставят число да се наричат мерки, така че вероятността е неотрицателна, нормирана, адитивна, непрекъсната **мярка**.

Дефиниция - Вероятност

Вероятността $\mathbf{P}$ е функция, дефинирана върху сигма алгебрата $\mathcal{A}$ от подмножества на $\mathbf{\Omega}$, която удовлетворява аксиомите:

- ❧ **Неотрицателност**. $\mathbf{P}(A) \geq 0$,
- ❧ **Нормираност**. $\mathbf{P}(\Omega) = 1$
- ❧ **Адитивност**. Ако $AB = \emptyset$, то $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$
- ❧ **Непрекъснатост**. За всяка монотонно намаляваща редица $A_1 \supset A_2 \supset \ldots A_n \supset \ldots$, клоняща към празното множество - $\emptyset$ е изпълнено $\lim\limits_{n\to\infty} \mathbf{P}(A_n) = 0$.

Всички споменати по-горе събития принадлежат на $\mathcal{A}$.

Тройката $(\Omega, \mathcal{A}, P)$ наричаме **вероятностно пространство**. Това всъщност е математическия модел на някакъв случаен експеримент. Вероятността може да бъде дефинирана по различни начини, стремежът ни е да изберем модел, който най-добре описва действителността. Класическата вероятност, която въведохме порано отговаря на аксиомите за вероятност, но тя е само една от възможностите, далеч не единствената.

Пример

Хвърляме монета, тогава $\mathbf{\Omega} = \{E, T\}$.

Можем да приемем модела на класическа вероятност, т.е. че монетата е правилна $\mathbf{P}(E) = \mathbf{P}(T) = 0.5$. Това обаче не следва от аксиомите, а е наш избор.

Можем да изберем модел с неправилна монета, при който тура се пада по-често $\mathbf{P}(E) = 0.4, \mathbf{P}(T) = 0.6$. Аксиомите отново ще бъдат удовлетворени.

Така получаваме различни вероятностни пространства.

Сега ще докажем основните свойства на вероятността:

❧ $\mathbf{P}(\emptyset) = 0$

Док.$\emptyset \cap \emptyset = \emptyset$ и $\emptyset \cup \emptyset = \emptyset$, следователно

$$\mathbf{P}(\emptyset) = \mathbf{P}(\emptyset \cup \emptyset) = \mathbf{P}(\emptyset) + \mathbf{P}(\emptyset)$$

❧ $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$

Док. $A\,\overline{A} = \emptyset$ и също така $A \cup \overline{A} = \mathbf{\Omega}$. Тогава

$$1 = \mathbf{P}(\mathbf{\Omega}) = \mathbf{P}(A \cup \overline{A}) = \mathbf{P}(A) + \mathbf{P}(\overline{A})$$

❧ Ако $B \subset A$, то $\mathbf{P}(B) \le \mathbf{P}(A)$

Док. За произволни събития A и B е изпълнено:

$$A = A\mathbf{\Omega} = A(B \cup \overline{B}) = AB \cup A\overline{B}$$

Тогава от адитивността следва

$$\mathbf{P}(A) = \mathbf{P}(AB) + \mathbf{P}(A\overline{B}) \qquad \text{(I .1)}$$

От друга страна, ако $B \subset A$, то $AB = B$ и $\mathbf{P}(AB) = \mathbf{P}(B)$.

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

В последното равенство използвахме, че $\mathbf{P}(A\,\overline{B}) \ge 0$.

❧ $0 \le \mathbf{P}(A) \le 1$

Док. От $A \subset \mathbf{\Omega}$ следва $\mathbf{P}(A) \le \mathbf{P}(\mathbf{\Omega}) = 1$

❧ $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$

Док. Отново ще използваме представянето $A = AB \cup A\overline{B}$, както и аналогичното $B = AB \cup \overline{A}B$. Тогава

$$A \cup B = (AB \cup A\overline{B}) \cup (AB \cup \overline{A}B) = AB \cup A\overline{B} \cup \overline{A}B$$

При това събитията от дясната страна на равенството нямат сечение. Така, ако приложим адитивността два пъти последователно, ще получим:

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

Към това равенство ще прибавим и извадим $\mathbf{P}(AB)$

$$\mathbf{P}(A \cup B) = \left[\mathbf{P}(AB) + \mathbf{P}(A\overline{B})\right] + \left[\mathbf{P}(\overline{A}B) + \mathbf{P}(AB)\right] - \mathbf{P}(AB) =$$

$$= \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$$

❧ Принцип за включване и изключване

$$\mathbf{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_i \mathbf{P}(A_i) - \sum_{i,j} \mathbf{P}(A_i A_j) + \ldots + (-1)^{n+1} \mathbf{P}(A_1 A_2 \ldots A_n)$$

Док. Доказателството се извършва по индукция. Ще допуснем че равенството е изпълнено за $n-1$ на брой събития и ще го докажем за n. Съгласно предишното свойство

$$\mathbf{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = \mathbf{P}((A_1 \cup A_2 \cup \ldots \cup A_{n-1}) \cup A_n) = \qquad (\text{I }.2)$$

$$= \mathbf{P}(A_1 \cup A_2 \cup \ldots \cup A_{n-1}) + \mathbf{P}(A_n) - \mathbf{P}((A_1 \cup A_2 \cup \ldots \cup A_{n-1}) \cap A_n)$$

Прилагаме индукционното предположение за първото и третото събираемо

$$\mathbf{P}(A_1 \cup A_2 \cup \ldots \cup A_{n-1}) = \sum_{1 \le i \le n-1} \mathbf{P}(A_i) - \sum_{1 \le i,j \le n-1} \mathbf{P}(A_i A_j) + \sum_{1 \le i,j,k \le n-1} \mathbf{P}(A_i A_j A_k) - \ldots$$

$$\mathbf{P}((A_1 \cup A_2 \cup \ldots \cup A_{n-1}) \cap A_n) = \mathbf{P}(A_1 A_n \cup A_2 A_n \cup \ldots \cup A_{n-1} A_n) =$$
$$= \sum_{1 \le i \le n-1} \mathbf{P}(A_i A_n) - \sum_{1 \le i,j \le n-1} \mathbf{P}(A_i A_j A_n) + \sum_{1 \le i,j,k \le n-1} \mathbf{P}(A_i A_j A_k A_n) + \ldots$$

Ще заместим последните равенства в (I .2)

$$\mathbf{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{1 \le i \le n-1} \mathbf{P}(A_i) + \mathbf{P}(A_n) - \sum_{1 \le i,j \le n-1} \mathbf{P}(A_i A_j) - \sum_{1 \le i \le n-1} \mathbf{P}(A_i A_n) +$$

$$+ \sum_{1 \le i,j,k \le n-1} \mathbf{P}(A_i A_j A_k) + \sum_{1 \le i,j \le n-1} \mathbf{P}(A_i A_j A_n) - \ldots + (-1)^{n+1} \mathbf{P}(A_1 A_2 \ldots A_n)$$

Ще разгледаме поотделно всички единични, двойни, тройни и т.н. суми. Ясно е, че

$$\sum_{1 \le i \le n-1} \mathbf{P}(A_i) + \mathbf{P}(A_n) = \sum_{1 \le i \le n} \mathbf{P}(A_i)$$

Двойните суми са две. Първата съдържа всички възможни двойки сечения на първите $n-1$ събития. Втората сума се състои от всички сечения на събитието A_n с някое от първите $n-1$ събития. Ако съберем тези две суми, ще получим всички възможни двойни сечения изобщо, т.е.

$$\sum_{1 \le i,j \le n-1} \mathbf{P}(A_i A_j) + \sum_{1 \le i \le n-1} \mathbf{P}(A_i A_n) = \sum_{1 \le i,j \le n} \mathbf{P}(A_i A_j)$$

Аналогично се работи със сумите от тройни сечения и т.н. С това доказателството е завършено.

Част за напреднали ☙

Съществува и алтернативен начин за дефиниране на вероятността, като аксиомите за адитивност и непрекъснатост се заменят с една единствена аксиома - така наречената изброима адитивност.

Алтернативна дефиниция - Вероятност

Вероятността $\mathbf{P}$ е функция, която удовлетворява аксиомите:

❧ **Неотрицателност**. $\mathbf{P}(A) \geq 0$, за всяко събитие $A \in \mathcal{A}$,

❧ **Нормираност**. $\mathbf{P}(\Omega) = 1$

❧ **Изброима адитивност**. Ако $A_1, A_2, \dots \in \mathcal{A}$ и за $\forall i, j$: $A_i A_j = \emptyset$, то

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$$

Твърдение

Дефиницията за вероятност (стр.16) и алтернативната дефиниция за вероятност са еквивалентни.

Док.Първите две аксиоми „Неотрицателност” и „Нормираност” съвпадат, затова е достатъчно да докажем, че аксиомите „Адитивност” и „Непрекъснатост” са изпълнени тогава и само тогава когато е в сила „Изброима адитивност”.

5. Условна вероятност и независимост на събития

Нека A и B са произволни събития, като поставяме единствено изискването B да не е невъзможно събитие, т.е. $\mathbf{P}(B) \neq 0$. Често възниква необходимост да се пресметне вероятността на събитието A, след като сме наблюдавали сбъдването на B. Прието е тази вероятност да се означава с $\mathbf{P}(A \,|\, B)$.

Пример

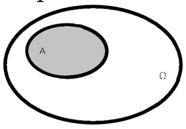
Ако наблюдаваме случайното теглене на карта от тесте с 32 карти може да забележим, че изтеглената карта е “картинка” (вале, дама или поп), без да обърнем внимание коя точно е картата. Ако при това положение се опитаме да намерим вероятността да е дама, това което търсим всъщност е условна вероятност. Нека да означим с A събитието - “изтеглена е дама”, а с B - “изтеглена е картинка”. Ние се опитваме да пресметнем вероятността на A, при условие че се е сбъднало B. Ако разсъждаваме в термините на класическа вероятност, то търсената вероятност ще е дроб, на която в знаменателя ще са всички начини за изтегляне на картинка, а в числителя тези от тях, при които е изтеглена дама, т.е.

$$\mathbf{P}(A|B) = \frac{\text{брой на дамите}}{\text{брой на картинките}} = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)} = \frac{4}{12}$$

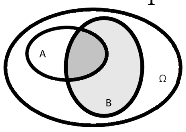
Тази вероятност явно е различна от вероятността да изтеглим дама.

Това ни дава основание да въведем понятието условна вероятност по следния начин.

Вероятност



Условна вероятност



Дефиниция - Условна вероятност

Нека $\mathbf{P}(B) \neq 0$. Вероятността да се изпълни A, ако знаем че се е изпълнило B се дефинира като:

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)}$$

Условната вероятност отговаря на аксиомите въведени за вероятността. Наистина:

❧ Неотрицателност

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)} \geq 0$$

❧ Нормираност

$$\mathbf{P}(\Omega \mid B) = \frac{\mathbf{P}(\Omega B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B)}{\mathbf{P}(B)} = 1$$

❧ Адитивност

Ако $A_1A_2 = \emptyset$, то $A_1B \cap A_2B = \emptyset$ и тогава:

$$\mathbf{P}(A_1 \cup A_2 \mid B) = \frac{\mathbf{P}(A_1B \cup A_2B)}{\mathbf{P}(B)} =$$

$$= \frac{\mathbf{P}(A_1B) + \mathbf{P}(A_2B)}{\mathbf{P}(B)} = \mathbf{P}(A_1 \mid B) + \mathbf{P}(A_2 \mid B)$$

❧ Непрекъснатост

Нека

$$A_1 \supset A_2 \supset \ldots A_n \supset \ldots \supset \emptyset$$

е монотонно намаляваща редица, същото е изпълнено и за редицата

$$A_1B \supset A_2B \supset \ldots A_nB \supset \ldots \supset \emptyset$$

Тогава $\lim\limits_{n \to \infty} \mathbf{P}(A_nB) = 0$ и следователно

$$\lim_{n \to \infty} \mathbf{P}(A_n \mid B) = \lim_{n \to \infty} \frac{\mathbf{P}(A_nB)}{\mathbf{P}(B)} = 0$$

След като са изпълнени аксиомите, то условната вероятност притежава и всички свойства характерни за вероятността изобщо, т.е. условната вероятност е вероятност.

Възможно е да се случи така, че сбъдването на събитието B да не влияе на вероятността за сбъдване на A, т.е. $\mathbf{P}(A \mid B) = \mathbf{P}(A)$. Съгласно дефиницията за условна вероятност тогава би трябвало:

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)} = \mathbf{P}(A)$$

Така достигаме до следната дефиниция:

Дефиниция - Независимост на събития

Казваме, че събитията A и B са независими тогава и само тогава, когато

$$\mathbf{P}(AB) = \mathbf{P}(A)\,\mathbf{P}(B)$$

За означаване на независимостта използваме следния знак $A \perp\!\!\!\perp B$.

Директно от дефиницията за независимост следва че, ако $A \perp\!\!\!\perp B$, то и $B \perp\!\!\!\perp A$.

Понятието независимост означава и независимост в обичайния смисъл на думата, т.е. едното събитие не влияе и не носи информация за другото. Например, ако се хвърлят два зара, какво ще се падне на единия и другия зар са независими събития.

Независимостта не винаги е очевидна. Понякога се налага да се направят конкретни пресмятания за да се установи дали са независими събитията. Ще се върнем към примера от началото на тази глава.

Пример

Тегли се карта от тесте от 32 карти. Дефинирани са събитията:

A - "изтеглена е дама"

B - "изтеглена е картинка"

C - "изтеглена е купа"

Оказва се, че събитията A и B са зависими, докато B и C са независими. Наистина.

Купите са 1/4 от всички карти, също така купите са 1/4 и от картинките. Ако знаем че е изтеглена картинка, това няма да промени вероятността да е изтеглена купа, тя пак е 1/4, т.е B не носи информация за C.

По съвсем различен начин стои въпросът с дамите. Дамите са 1/8 от всички карти, но само 1/3 от всички картинки. Така фактът, че е изтеглена картинка променя вероятността за изтеглена дама, т.е. A и B са зависими. Какво можете да кажете за независимостта на A и C?

Проверете това формално, като използвате дефиницията за независимост.

Пример - Различими или неразличими зарове

Хвърляме два еднакви правилни зара. Вероятността да се падне единица на всеки зар е $\frac{1}{6}$. Да намерим вероятността за $\{1, 1\}$. Ще разгледаме два случая.

Различими зарове	Неразличими зарове
{1,1} {1,2} {1,3}... {1,6} {2,1} {2,2} {2,3}... {2,6} ... {6,1} {6,2} {6,3}... {6,6}	{1,1} {2,1} {2,2} ... {6,1} {6,2} {6,3} ... {6,6}
Брой на всички случай 36. Вероятност на всеки случай $\frac{1}{36}$ $\mathbf{P}(\{1,1\}) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$ Заровете са независими.	Брой на всички случай 21. Вероятност на всеки случай $\frac{1}{21}$ $\mathbf{P}(\{1,1\}) = \frac{1}{21} \neq \frac{1}{6} \cdot \frac{1}{6}$ Заровете са зависими.

Ако приемем заровете за неразличими, ще се окаже, че те са зависими. Това е абсурд, няма как точките паднали се на единия зар да влияят на точките падащи се на другия. Затова, трябва винаги да смятаме заровете за различими, независимо дали ние ги различаваме или не, те са физически различни и се държат като различни.

Възниква следния въпрос, ако $A \perp\!\!\!\perp B$, то какво може да се каже за независимостта на $\overline{A}$ и $\overline{B}$.

Твърдение

Нека $A \perp\!\!\!\perp B$, тогава $A \perp\!\!\!\perp \overline{B}$, $\quad \overline{A} \perp\!\!\!\perp B \quad$ и $\quad \overline{A} \perp\!\!\!\perp \overline{B}$.

Док. От $A \perp\!\!\!\perp B$ следва $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$.

Виж равенство (I .1)

Знаем, че вероятността на произволно събитие може да се представи като следната сума $\mathbf{P}(A) = \mathbf{P}(AB) + \mathbf{P}(A\overline{B})$. Тогава

$$\mathbf{P}(A\overline{B}) = \mathbf{P}(A) - \mathbf{P}(AB) = \mathbf{P}(A) - \mathbf{P}(A)\mathbf{P}(B) =$$

$$= \mathbf{P}(A)(1 - \mathbf{P}(B)) = \mathbf{P}(A)\mathbf{P}(\overline{B})$$

От тук следва $A \perp\!\!\!\perp \overline{B}$. Аналогично се доказват и останалите равенства.

Лесно е да се съобрази, че празното събитие $\emptyset$ е независимо от всяко друго. Същото важи и за цялото събитие $\mathbf{\Omega}$.

За повече от две събития въвеждаме понятието независимост по аналогичен начин.

Дефиниция - Независимост в съвкупност

Казваме, че събитията $A_1, A_2, \ldots, A_n$ са **независими в съвкупност** (или просто независими), ако за всяко цяло число k, $2 \leq k \leq n$ и всеки набор от индекси $i_1, i_2, \ldots i_k$ е изпълнено

$$\mathbf{P}(A_{i_1}A_{i_2}\ldots A_{i_n}) = \mathbf{P}(A_{i_1})\mathbf{P}(A_{i_2})\ldots\mathbf{P}(A_{i_n})$$

Естествения въпрос, който следва е, ако събитията са независими две по две, т.е. всяко с всяко, то дали това означава , че те са независими в съвкупност. Отговорът се дава от следния контрапример

Пример

Хвърляме тетраедър, на който едната страна е бяла, другата зелена, третата червена, а върху четвъртата страна има и трите цвята. Дефинираме събития A , B и C на долната страна се пада, съответно бяло, зелено червено. Елементарно се пресмятат вероятностите

$$\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = \frac{1}{2}, \qquad \mathbf{P}(AB) = \mathbf{P}(BC) = \mathbf{P}(AC) = \frac{1}{4}$$

Това означава, че събитията са независими две по две. От друга страна $\mathbf{P}(ABC) = \frac{1}{4}$, следователно събитията не са независими в съвкупност.

Ако събитията $A_1, A_2, \ldots, A_n$ са независими, то вероятността да се изпълнят едновременно е

$$\mathbf{P}(A_1 A_2 \ldots A_n) = \mathbf{P}(A_1)\mathbf{P}(A_2) \ldots \mathbf{P}(A_n)$$

Вероятността за едновременно сбъдване на произволни събития се дава в следното твърдение.

Твърдение

$$\mathbf{P}(A_1 A_2 \ldots A_n) = \mathbf{P}(A_1)\ \mathbf{P}(A_2|A_1) \ldots \mathbf{P}(A_n|A_1 A_2 \ldots A_{n-1})$$

Док. Доказателството се провежда с индукция по n. За $n = 2$ директно от дефиницията на условна вероятност следва

$$\mathbf{P}(A_1 A_2) = \mathbf{P}(A_1)\ \mathbf{P}(A_2|A_1)$$

Разглеждаме $A_1 A_2 \ldots A_{n-1}$ като едно цяло събитие, тогава

$$\mathbf{P}(A_1 A_2 \ldots A_n) = \mathbf{P}((A_1 A_2 \ldots A_{n-1})A_n) = \mathbf{P}(A_1 \ldots A_{n-1})\ \mathbf{P}(A_n|A_1 \ldots A_{n-1}) =$$

Прилагаме индукционното предположение за $\mathbf{P}(A_1 \ldots A_{n-1})$

$$= \mathbf{P}(A_1)\ \mathbf{P}(A_2|A_1)\ \ldots \mathbf{P}(A_{n-1}|A_1 A_2 \ldots A_{n-2})\ \mathbf{P}(A_n|A_1 \ldots A_{n-1})$$

Пример

Какъв е най-малкият брой хора, които трябва да се изберат по случаен начин, така че вероятността рождените дни на поне двама от тях да съвпаднат да е по-голяма от 1/2?

В случая е по-удобно да пресметнем вероятността на противоположното събитие. Вероятността всички да са родени в различни дни, трябва да е по-малка от 1/2. Избираме хората последователно и за всеки човек дефинираме събитие

$A_k = \{$ рождения ден на к-тия човек е различен от предходните$\}$

$k = 1 \dots n$. Явно е, че търсим сечението на тези събития

$$\mathbf{P}(A_1 A_2 \dots A_n) = \mathbf{P}(A_1)\ \mathbf{P}(A_2|A_1)\ \mathbf{P}(A_3|A_1 A_2) \dots \mathbf{P}(A_n|A_1 A_2 \dots A_{n-1}) =$$

Ще приемем, че всички възможни рождени дни са 365 и при това са равновероятни. Трябва да определим n такова, че

$$= \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \dots \frac{366-n}{365} < \frac{1}{2}$$

От тук получаваме $n = 23$, което в известен смисъл противоречи на интуицията. Интересно е, че при $n = 70$ вероятността да има хора с един и същ рожден ден е 0.999.

6. Формула за пълната вероятност и формула на Бейс

Формулата за пълна вероятност ни дава възможност да описваме сложни модели с включени много събития зависещи едно от друго.

Нека събитията $H_1, H_2, \dots H_n$ са всички възможности при извършване на някакъв опит, т.е. нямат сечение и образуват покритие на Ω.

$$\forall i, j : \quad H_i H_j = \emptyset$$

$$\bigcup_{i=1}^{n} H_i = \Omega$$

Прието е такива множества да се наричат **пълна група от събития**. Също така ще ги наричаме **хипотези**. Ясно е, че за хипотезите е изпълнено:

$$\sum_{i=1}^{n} \mathbf{P}(H_i) = 1$$

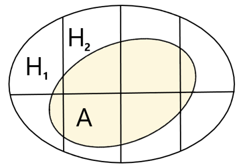
Нека A е произволно събитие зависещо по някакъв начин от хипотезите H_i . Пресмятането на вероятността на A често е по-лесно, ако самото A се разложи според хипотезите.

Твърдение - Формула за пълната вероятност

Ако $H_1, H_2, \dots H_n$ образуват пълна група от събития, то за произволно събитие A е изпълнено:

$$\mathbf{P}(A) = \sum_{i=1}^{n} \mathbf{P}(A|H_i)\, \mathbf{P}(H_i)$$

Док. Нека $H_1, H_2, \dots H_n$ са хипотези, тогава



$$A = A\,\Omega = AH_1 \cup AH_2 \cup \dots \cup AH_n$$

като събитията в това обединение са непресичащи се. От адитивността на вероятността следва

$$\mathbf{P}(A) = \mathbf{P}(AH_1 \cup AH_2 \cup \dots AH_n) = \sum_{i=1}^{n} \mathbf{P}(AH_i) = \sum_{i=1}^{n} \mathbf{P}(A|H_i)\, \mathbf{P}(H_i)$$

В последното равенство използвахме формулата за сечение на събития. ✓

Пример

Момче има в левия си джоб 1 монета от 1лв. и 2 от 2лв., а в десния 2 монети по 1лв. и 3 по 2лв. От случайно избран джоб момчето вади монета. Каква е вероятността тя да е от 1лв?

Момчето най-напред избира джоб, след което вади монета от него, затова ще формулираме хипотезите за джоба който е избрало момчето:

$H_1 = \{$избран е левия джоб$\}$, $H_2 = \{$ избран е десния джоб$\}$

Нека $A = \{$ извадена е монета от 1лв$\}$, т.е. A е събитието което търсим.

Ясно е че, $\mathbf{P}(H_1) = \mathbf{P}(H_2) = 1/2$. Остава да пресметнем вероятността да бъде извадена монета от 1лв, ако е избран левия, съотвотно десния джоб:

$$\mathbf{P}(A|H_1) = \frac{1}{3} \qquad \text{и} \qquad \mathbf{P}(A|H_2) = \frac{2}{5}$$

Сега, от формулата за пълната вероятност получаваме:

$$P(A) = \frac{1}{2}.\frac{1}{3} + \frac{1}{2}.\frac{2}{5} = \frac{11}{30}$$

Тази вероятност е различна от вероятността, която бихме получили, ако смятахме че от всички монети по случаен начин се избира една, т.е. ако всичките монети лежаха в един джоб. Тогава вероятността би била 3/8.

Формулата на Бейс в известен смисъл е обратна на формулата за пълната вероятност. Постановката на задачата е подобна. Отново имаме хипотези $H_1, H_2, \dots H_n$ и събитие A, което зависи от тях. В този случай обаче сме наблюдавали сбъдването на събитието A, и при това условие се опитваме да намерим вероятността на някоя от хипотезите.

Прието е вероятността за сбъдване на някоя хипотеза, изчислена преди провеждането на опита, т.е. $\mathbf{P}(H_k)$ да се нарича **априорна**. Вероятноста да се е била сбъднала някоя хипотеза, след като сме наблюдавали резултата от опита, т.е. $\mathbf{P}(H_k|A)$ се нарича **апостериорна**.

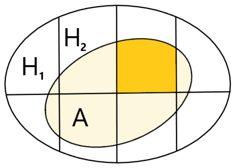
Апостериорните вероятности се изчисляват по следната формула.

Твърдение - Формула на Бейс

Ако $H_1, H_2, \dots H_n$ образуват пълна група от събития, а A е произволно събитие, то:

$$\mathbf{P}(H_k|A) = \frac{\mathbf{P}(A|H_k)\,\mathbf{P}(H_k)}{\sum_{i=1}^{n} \mathbf{P}(A|H_i)\,\mathbf{P}(H_i)}$$

Док. Съгласно дефиницията за условна вероятност



$$\mathbf{P}(H_k|A) = \frac{\mathbf{P}(H_k A)}{\mathbf{P}(A)}$$

За да завършим доказателството е достатъчно да приложим формулите за сечение в числителя и за пълна вероятност в знаменателя. ✓

Пример

Ще продължим предишния пример. Нека момчето е извадило монета от един лев. Каква е вероятността монетата да е била в левия джоб?

Знаем априорната вероятност момчето да избере левия джоб $\mathbf{P}(H_1) = \frac{1}{2}$. Това, което търсим е апостериорната вероятност $\mathbf{P}(H_1|A)$.

$$\mathbf{P}(H_1|A) = \frac{\mathbf{P}(A|H_1)\,\mathbf{P}(H_1)}{\mathbf{P}(A|H_1)\,\mathbf{P}(H_1) + \mathbf{P}(A|H_2)\,\mathbf{P}(H_2)} =$$

$$= \frac{\frac{1}{3}\frac{1}{2}}{\frac{1}{3}\frac{1}{2} + \frac{2}{5}\frac{1}{2}} = \frac{5}{11}$$

7. Допълнение - Комбинаторика

Пресмятането на класическа вероятност обикновено се свежда до изброяването на елементите на дадено множество. По-нататък ще припомним някой основни комбинаторни формули. Нека $\boldsymbol{\Omega} = \{\omega_1, \omega_2, \dots \omega_n\}$ е произволно множество.

Всяко подреждане на елементите на $\boldsymbol{\Omega}$ наричаме пермутация. Броят на всички пермутации означаваме

$$P_n = 1.2.3. \dots .n = n!$$

Ще разгледаме четири типа подмножества на $\boldsymbol{\Omega}$.

- **вариации без повторение** - нареденено подмножество с обем k, елементите не се повтарят

$$V_n^k = n.(n-1).(n-2). \dots .(n-(k-1)) = \frac{n!}{(n-k)!}$$

- **вариации с повторение** - наредено подмножество с обем k, елементите могат да се повтарят

$$\widetilde{V}_n^k = n.n.\ldots.n = n^k$$

- **комбинации без повторение** - ненаредено подмножество с обем k, елементите не се повтарят

$$C_n^k = \frac{n.(n-1).(n-2).\ldots.(n-(k-1))}{k!} = \binom{n}{k}$$

- **комбинации с повторение** - ненаредено подмножество с обем k, елементите могат да се повтарят

$$\widetilde{C}_n^k = \binom{n+k-1}{k}$$

Обобщение

аа

Задачи

Appendix - Комбинаторика

1.1 Разпределят се k различни частици в n различни клетки. Намерете броя на възможните начини на разпределяне, ако:
а) всяка клетка може да съдържа най-много една частица;
б) клетките могат да съдържат произволен брой частици.

1.2 Разпределят се k неразличими частици в n различни клетки. Намерете броя на възможните начини на разпределяне, ако:
а) всяка клетка може да съдържа най-много една частица;
б) клетките могат да съдържат произволен брой частици.

1.3 Десет души се нареждат в редица. Колко са подрежданията, при които три фиксирани лица се намират едно до друго.

1.4 Колко четирицифрени числа могат да се напишат от цифрите 1, 2, 3, 4 и 5, ако:
а) цифрите участват по веднъж;
б) допуска се повтаряне на цифри;
в) не се допуска повтаряне и числото е нечетно.

1.5 Група от 12 студенти трябва да изпрати при декана делегация от четирима свой представители. По колко начина може да се избере състава, ако:
а) няма ограничения за участие в нея;
б) студентите A и B не трябва да участват заедно;
в) студентите C и D могат да участват само заедно.

1.6 Пет различни топки се разпределят в три различни кутии A, B, C. Да се намери броя на всички различни разпределения, при които:
а) кутията A е празна;
б) само кутията A е празна;
в) точно една кутия е празна;
г) поне една кутия е празна;
д) няма празна кутия.

1.7 Нека $\boldsymbol{\Omega}$ е множеството на всички наредени $n-$торки с повторение на цифрите 1,2, и 3. Да се намери броят на елементите на $\boldsymbol{\Omega}$, които:
а) започват с 1;
б) съдържат точно к пъти цифрата 2;
в) съдържат точно к пъти цифрата 1, при което започват и завършват с 1;
г) са съставени от k_1 единици, k_2 двойки и k_3 тройки.

1.8 Всяка стена на всяко едно от сто кубчета е или червена, или синя, или зелена. Нека 80 кубчета имат поне една червена стена, 85 кубчета поне една синя, 75 кубчета поне една зелена. Какъв е най-малкият брой кубчета, които имат стени от трите цвята.

1.9 Дадено е множеството $\boldsymbol{\Omega} = \{a_1, \ldots, a_n, b_1, \ldots, b_k\}$. Колко са подмножествата на $\boldsymbol{\Omega}$, които съдържат поне един елемент a и поне един елемент b.

Класическа вероятност

1.10 Куб, на който всички страни са боядисани в различни цветове, е разрязан на 1000 еднакви кубчета. Да се определи вероятността случайно избрано кубче да има точно две боядисани страни.

1.11 Да се определи вероятността контролният номер на първата срещната лека кола:
а) да не съдържа еднакви цифри;
б) да има точно две еднакви цифри;
в) да има три еднакви цифри;
г) да има две двойки еднакви цифри;
д) да има една и съща сума от първите две и последните две цифри.

1.12 От десет лотарийни билета два са печеливши. Да се определи вероятността, между изтеглени по случаен начин пет билета:
а) точно един да бъде печеливш;
б) да има два печеливши;
в) да има поне един печеливш.

1.13 При игра на тото 6 от 49 да се пресметнат вероятностите за печалба на шестица, петица, четворка и тройка.

1.14 С цел намаляване броят на играните мачове, $2k$ отбора с жребий се разбиват на две групи. Да се определи вероятността двата най-силни отбора да са в различни групи.

1.15 Във влак с три вагона по случаен начин се качват седем пътника. Каква е вероятността в първият вагон да се качат четирима.

1.16 Група от n човека се нарежда в редица по случаен начин. Каква е вероятността между две фиксирани лица да има точно r човека.

1.17 Група от n човека се нарежда около кръгла маса. Каква е вероятността две фиксирани лица да се окажат едно до друго.

1.18 От урна, която съдържа топки с номера $1, 2, \ldots, n$, k пъти последователно се вади по една топка. Да се пресметне вероятността номерата на извадените топки, записани по реда на изваждането, да образуват растяща редица, ако:
а) извадката е без връщане;
б) извадката е с връщане.

1.19* На опашка за билети са се наредили 20 човека, десет от тях имат само банкнота от 10 лв., а останалите от 20 лв. Всеки един от тях иска да купи един билет от 10 лв. Каква е вероятността на никой от опашката да не му се наложи да чака за връщане на ресто, ако преди започване на продажбата в касата няма пари.

1.20 На избори се явяват двама кандидати. За първият са пуснати n гласа, а за вторият k $(n > k)$. Каква е вероятността през цялото време докато се броят резултатите, броят бюлетини за първият кандидат да е по-голям от броят на бюлетините за втория.

Геометрична вероятност

1.21 Даден е кръг с радиус R. Върху диаметъра по случаен начин е избрана точка А. През точка А е прекарана хорда перпендикулярна на диаметъра. Да се определи вероятността хордата да бъде по-къса от R.

1.22 На плоскост са прекарани два типа успоредни ивици, първите са с ширина 1 см, а вторите 2 см, разстоянието между ивиците е 1 см (плоскостта е на райета). Върху плоскостта се хвърля монета с диаметър 2 см. Нека A, B са събитията монетата застъпва първите съответно вторите ивици. Да се определи вероятността на: A, B, AB, A∪B, A|B.

1.23 Два парахода трябва да бъдат разтоварени на един и същи пристан. Всеки един от тях, независимо от другия, може да пристигне в кой да е момент на даден ден (24ч.). Каква е вероятността параходите да не се изчакват, ако за разтоварването на първия са необходими 6ч., а за втория 4ч.

1.24 Автобусите от линия A се движат на интервали от шест минути, а от линия B на четири минути, независимо от автобусите от линия A. Да се пресметне вероятността:

а) автобус A да дойде преди B;

б) пътник, дошъл в случаен момент на спирката, да чака не повече от две минути.

1.25 Дадена е отсечка с дължина K. По случаен начин се избират две други отсечки с дължина по-малка от K. Каква е вероятността от трите отсечки да може да се построи триъгълник.

1.26 Каква е вероятността от три избрани по случаен начин отсечки с дължина по-малка от K да може да се построи триъгълник.

1.27 Дадена е магнетофонна лента с дължина 100м. Върху всяка от двете страни на лентата, на случайно избрано място, е записано непрекъснато съобщение с дължина 20м. Каква е вероятността между 25 и 50м., считано от началото на лентата, да няма участък несъдържащ поне едно от двете съобщения?

1.28 Каква е вероятността сумата на две случайно избрани, положителни числа, всяко от които е по-малко от едно, да бъде по-малка от едно, а произведението им по-малко от 2/9.

1.29 В равнина са прекарани успоредни линии на разстояние L. Върху равнината се хвърля игла с дължина k, като $k < L$. Каква е вероятността иглата да застъпва някоя от линиите.

Условна вероятност

1.30 Вероятността стрелец да улучи мишена е 2/3, ако улучи той получава право на втори изстрел. Вероятността за улучване на двете мишени е 1/2. Каква е вероятността за улучване на втората мишена, ако стрелеца е получил право да стреля втори път.

1.31 Застрахователна компания води статистика за своите клиенти:

-всички клиенти посещават поне веднъж годишно лекар;

-60 % посещават повече от веднъж годишно лекар;

-17 % посещават хирург;

-15 % от тези, които посещават повече от веднъж годишно лекар, посещават хирург.

Каква е вероятността случайно избран клиент, който посещава само веднъж годишно лекар, да не е бил при хирург?

1.32 Да се определи вероятността, случайно избрано цяло положително число, да не се дели:

а) нито на две, нито на три;

б) на две или на три.

1.33 Хвърлят се два зара. Каква е вероятността сумата от падналите се точки да е по-малка от 8, ако се знае, че тя е нечетна? Независими ли са двете събития?

1.34 Около маса сядат 10 мъже и 10 жени. Каква е вероятността да няма лица от един пол едно до друго?

1.35 Какъв е най-малкият брой хора, които трябва да се изберат по случаен начин, така че вероятността рождените дни на поне двама от тях да съвпаднат да е по-голяма от 1/2?

1.36 Двама играчи последователно хвърлят монета, играта печели този, който първи хвърли герб. Да се намери вероятността за спечелване на играта за всеки от двамата играчи.

1.37 А получава информация и я предава на Б, той я предава на В, той на Г. Г съобщава получената информация. Известно е, че всеки от тях казва истина само в един от три случая. Ако излъжат двама човека, отново се получава истина. Каква е вероятността първият А да не е излъгал, ако е известно, че последният Г е съобщил истината.

1.38 Секретарка написала n писма, сложила ги в пликове и ги запечатала. Забравила кое писмо в кой плик е, но въпреки това написала отгоре n различни адреса и изпратила писмата. Да се определи вероятността:

а) всеки да получи своето писмо;

б) точно $n-1$ човека да получат своето писмо;

в) нито едно лице да не получи своето писмо.

II . Дискретни случайни величини

Основна тема изучавана в теорията на вероятностите са случайните величини и техните характерстики и свойства. Случайната величина е обект, който може да взима за стойности случайни реални числа с някаква вероятност. За означаване на сл.в. ще използваме главните латински букви - X, Y и т.н. Различните конкретни стойности на сл.в. X, ще означаваме с малките букви - $x_1, x_2, \ldots$.

Ще разглеждаме два типа случайни величини.

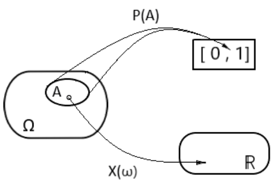
❧ Дискретни случайни величини, които могат да взимат само краен, или най-много изброим брой стойности. Такива например са: точките паднали се върху зар, броя опити необходим за да изтеглим асо от тесте карти, броят студенти прочели този учебник и т.н.

❧ Непрекъснати случайни величини, които взимат стойности в някое неизброимо множество. Такава например е външната температура в момента.

В тази глава ще разглеждаме дискретните случайни величини.

1. Определение, свойства

Нека $(\boldsymbol{\Omega}, \mathcal{A}, \mathbf{P})$ е вероятностно пространство. Да припомним $\boldsymbol{\Omega}$ е множеството на елементарните събития ω, който са резутата от някакъв прост опит. $\mathcal{A}$ е σ-алгебра съдържаща подмножества на $\boldsymbol{\Omega}$ и за тези подмножества е въведено понятието вероятност $\mathbf{P}$, като на всяко множество от $\mathcal{A}$ съпоставяме число от интервала $[0, 1]$. Вероятността $P(A)$ е изображение от множеството $\boldsymbol{\Omega}$ в този интервал.



Нека извършваме някакъв прост опит. При всяко сбъдване на елементарно събитие ω случайната величина X приема конкретна стойност, т.е случайната величина $X(\omega)$ е изображение от множеството $\boldsymbol{\Omega}$ върху множеството на реалните числа. Тогава, случайните величини са функции с аргумент елементарните събития. Правилно би било да пишем $X(\omega)$, в действителност аргументът се изпуска и се записва просто X, като се подразбира, че става дума за функция.

Преди да въведем формалната дефиниция на дискретните случайни величини ще дадем един прост пример.

Пример

Хвърляме два зара. Възможно е да търсим просто вероятността на някое събитие. Възможно е да изучаваме някакъв по сложен обект, свързан с опита, например:

X - сумата от точките върху заровете

Y - броя на падналите се шестици

Z - максимума на точките

При всеки резултат от хвърлянето на заровете X, Y и Z ще приемат някакави реални (в този частен случай цели) числа за стойности.

Ако изходът е

$\omega = \{1, 1\}$, то $X(\{1, 1\}) = 2$, $Y(\{1, 1\}) = 0$, $Z(\{1, 1\}) = 1$

$\omega = \{1, 2\}$, то $X(\{1, 2\}) = 3$, $Y(\{1, 2\}) = 0$, $Z(\{1, 2\}) = 2$

и т.н.

Множеството $\mathbf{\Omega}$ е произволно, върху него не се налагат никакви ограничения, така че то може да е с много сложна структура и съответно изображение от него в $\mathbb{R}$ да не е добре дефинирано. Затова се налага да се направи следната конструкция:

Нека H_j, $j = 1, 2, \ldots$ образуват разлагане (покритие) на $\mathbf{\Omega}$:

❧ $H_j \in \mathcal{A}$

❧ $\forall i, j : H_i H_J = \emptyset$

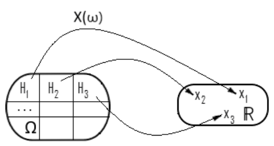
❧ $\bigcup_j H_j = \mathbf{\Omega}$

Ще дефинираме случайната величина X да взима различни стойности върху всяко от множествата H_j. За целта ще ни трябват индикаторите $\mathbb{I}_{H_j}$ на множествата. Да припомним:

$$\mathbb{I}_{H_j}(\omega) = \begin{cases} 1, & \omega \in H_j; \\ 0, & \omega \notin H_j. \end{cases}$$

Дефиниция - Дискретна случайна величина

Нека H_j, $j = 1, 2, \ldots$ е произволно, крайно или изброймо разлагане на $\mathbf{\Omega}$



Дискретна случайна величина наричаме:

$$X(\omega) = \sum_j x_j \, \mathbb{I}_{H_j}(\omega),$$

където $\mathbb{I}_{H_j}(\omega)$ е индикатора на множеството H_j, а x_j са произволни различни реални числа.

Тъй като H_j е разлагане на $\mathbf{\Omega}$, т.е. множествата H_j са непресичащи се, то ω принадлежи точно на едно множество H_j. Следователно в горната сума само един от индикаторите H_j е равен на едно, а останалите са нули, т.е. единствените стойности които сл.в. X приема са x_j, при това $\mathbf{P}(X(\omega) = x_j) = \mathbf{P}(H_j)$

Ние рядко използваме формалната дифиницията на случайна величина, по-често работим с нейното разпределение.

Дефиниция - Разпределение на дискретна сл.в.

Разпределение на дискретната случайна величина X, наричаме следната таблица:

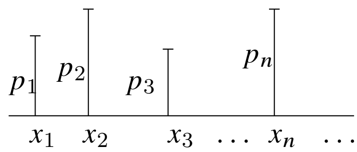
X	x_1	x_2	$\dots$	x_n	$\dots$
$\mathbf{P}$	p_1	p_2	$\dots$	p_n	$\dots$

където
x_i са стойностите на сл.в. те могат да бъдат краен или изброим брой;
$p_j = \mathbf{P}(X = x_j)$ са вероятностите с които сл.в. взема съответните стойности.

За да бъде X добре дефинирана е необходимо

$$\sum_j p_j = 1.$$

Също така "разпределение" много често наричаме и формула от която могат да бъдат пресметнати горните вероятности.



Графически ще представяме разпределението на случайните величини със схема подобна на хистограма. Като височината на j-тото стълбче съответства на вероятността p_j.

Пример

Нека хвърляме два зара и дефинираме сл.в. X -"сума от падналите се точки". Ще определим формалната дефиниция и разпределението на случайната величина.

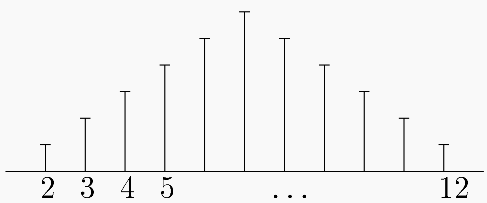
При хвърлянето на два зара $\boldsymbol{\Omega}$ се състои от 36 елементарни събития - $(1,1), (1,2), (2,1), \dots, (6,6)$, всяко от тях с вероятност $1/36$. Тогава
$X = X(\omega) = 2$, само ако $\omega = (1,1)$
$X = X(\omega) = 3$, ако $\omega = (1,2)$ или $\omega = (2,1)$
$X = 4$, ако $\omega = (1,3)$, или $\omega = (3,1)$, или $\omega = (2,2)$
и т.н. ...

Следователно формалната дефиниция на сл.в. X е

$$X(\omega) = 2\,\mathbb{I}_{(1,1)} + 3\,\mathbb{I}_{(1,2)\cup(2,1)} + 4\,\mathbb{I}_{(1,3)\cup(3,1)\cup(2,2)} + \dots + 12\,\mathbb{I}_{(6,6)}$$

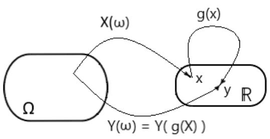
Ясно е, че разпределението е далеч по-удобно за работа:

X	2	3	4	$\dots$	7	$\dots$	12
$\mathbf{P}$	1/36	2/36	3/36	$\dots$	6/36	$\dots$	1/36



В практически задачи често се налага да изследваме не дадена случайна величина, а някаква функция зависеща от случайната величина. Това е атака наречената "Смяна на променливите".

Нека X е дискретна случайна величина, а $y = g(x)$ е произволна реална функция, не е трудно да се съобрази, че $Y = g(X)$ също е случайна величина.



Наистина, ако $X(\omega)$ е изображение от $\mathbf{\Omega}$ в $\mathbb{R}$, а $g(x)$ изображение от $\mathbb{R}$ в $\mathbb{R}$ то $Y(\omega) = g\left(X(\omega)\right)$ също е изображение от $\mathbf{\Omega}$ в $\mathbb{R}$.

Записано формално, ако

$$X(\omega) = \sum_j x_j \, \mathbb{I}_{H_j}(\omega),$$

то

$$Y(\omega) = \sum_j g(x_j) \, \mathbb{I}_{H_j}(\omega) = \sum_j y_j \, \mathbb{I}_{H_j}(\omega),$$

което явно е дефиниция на дискретна сл.в. Разбира се, ако за някои от стойностите $g(x_i) = g(x_j)$, т.е. ако функцията $g(X)$ не е еднозначно обратима, то съответните множества от разлагането се обединяват.

Смяната на променливите $Y = g(X)$ всъшност задава начин да се преизчислят стойностите на случайната величина. Разпределението на Y има вида:

Y	$g(x_1)$	$g(x_2)$	$\ldots$	$g(x_n)$	$\ldots$
P	p_1	p_2	$\ldots$	p_n	$\ldots$

Ако за някои i, j е изпълнено $g(x_i) = g(x_j)$, т.е. новите стойности съвпадат, то те се обединяват, а съответните им вероятности се събират.

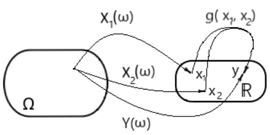
Пример

Нека X е случайната величина от предишния пример, т.е. сумата от точките при хвърлянето на два зара, а $Y = |X - 7|$. За разпределението на Y получаваме:

Y	0	1	2	$\ldots$	5
P	6/36	10/36	8/36	$\ldots$	2/36

По аналогичен начин може да се разсъждава, ако изследваме функция зависеща от няколко случайни величини, т.е. функция на много променливи. Ще разгледаме двумерния случай. Нека X_1 и X_2 са случайни величини, а $Y = g(X_1, X_2)$ е релна функция на две променливи. Тогава Y е случайна величина, защото

ако



$$X_1(\omega) = \sum_j x'_j \, \mathbb{I}_{H_j}(\omega) \qquad \text{и} \qquad X_2(\omega) = \sum_k x''_k \, \mathbb{I}_{T_k}(\omega)$$

то

$$Y(\omega) = \sum_j \sum_k g(x'_j, x''_k) \, \mathbb{I}_{H_j}(\omega)\mathbb{I}_{T_k}(\omega) = \sum_s y_s \, \mathbb{I}_{V_s}(\omega),$$

където означаваме $\mathbb{I}_{V_s}(\omega) = \mathbb{I}_{H_j}(\omega)\mathbb{I}_{T_k}(\omega) = \mathbb{I}_{H_j \cap T_k}(\omega)$ и $y_s = g(x'_j, x''_k)$.

Не е трудно да се съобрази, че ако H_j и T_k са покрития на Ω, то множеството от всички възможни техни сечения V_s също е покритие. И отново, ако имаме съвпадащи стойности на y_s за някои s, то съответните им множества от разлагането V_s се обединяват.

2. Функция на разпределение

В теория на вероятностите се използват различни функции, които носят в себе си информация за случайните величини и същевременно опростяват работата с тях.

Дефиниция - Функция на разпределение - Ф.Р.

Нека X е произволна сл.в. Функция на разпределение - $F(x)$ се дефинира с равенството:

$$F(x) = \mathbf{P}(X < x),$$

където x е произволно реално число.

Понякога като индекс се поставя самата случайна величина, т.е. записваме $F_X(x)$.

Използването на строго неравенство ($<$) е в известна степен произволно. В някои източници дефиницията е с нестрого неравенство ($\leq$) и получената Ф.р. има аналогични свойства.

Ще обърнем внимание, че в тази дефиниция не се поставя условие случайната величина да бъде дискретна, т.е. със същата формула се определя и $F_X(x)$ при непрекъснати случайни величини. Възможно е самата теория да бъде построена в обратен ред - първо да се въведе функцията на разпределение, а след това като “случайни величини” да се дефинират всички обекти, за които функцията $F_X(x)$ е определена, т.е. вероятността $\mathbf{P}(X < x)$ може да бъде пресметната за всяко реално x.

Доказва се, че тази дефиниция е еквивалентна на дефиницията за случайна величина, която ние въведохме по-горе.

Функцията на разпределение на дискретни случайни величини се използва рядко, но тя е изключително полезна при работа с непрекъснати сл.в. По нататък ще докажем редица нейни свойства. Сега ще начертаем типичната Ф.Р. на дискретна сл.в.

Функция на разпределение

Нека x е произволно реално число и разпределението на сл.в. X е следното:

X	x_1	x_2	$\dots$	x_n
$\mathbf{P}$	p_1	p_2	$\dots$	p_n

Ако $x \leq x_1$, то $\mathbf{P}(X < x) = 0$ и следователно $F(x) = 0$.
Ако $x_1 < x \leq x_2$, то явно $\mathbf{P}(X < x) = \mathbf{P}(X = x_1) = p_1$.
Ако $x_2 < x \leq x_3$, то $\mathbf{P}(X < x) = \mathbf{P}(X = x_1 \cup X = x_2) = p_1 + p_2$, и т.н.
Явно $F(x)$ е ненамаляваща, стъпаловидна функция със стой-

ности в интервала $[0, 1]$.

$$F(x) = \begin{cases} 0, & x \leq x_1 \\ p_1, & x_1 < x \leq x_2 \\ p_1 + p_2, & x_2 < x \leq x_3 \\ p_1 + p_2 + p_3, & x_3 < x \leq x_4 \\ \dots & \\ 1, & x > x_n \end{cases}$$

Свойства на функцията на разпределение.

❧ Нека $x_1 < x_2$, тогава $F(x_1) \leq F(x_2)$. Това свойство показва, че функцията на разпределение е ненамаляваща.

3. Математическо очакване

Често пъти се опитваме да опишем дадена случайна величина или съответното разпределение само с някои типични стойности. Искаме да използваме няколко константи, които да носят достатъчно информация и да описват поведението на случайната величина. За целта се въвеждат няколко важни числови характеристики на случайните величини.

Дефиниция - Математическо очакване - $\mathbf{E}\,X$

Математическо очакване на случайната величина X наричаме числото $\mathbf{E}\,X$ дефинирано по следния начин:

$$\mathbf{E}\,X = \sum_i x_i\,\mathbf{P}(X = x_i) = \sum_i x_i\,p_i. \tag{II .1}$$

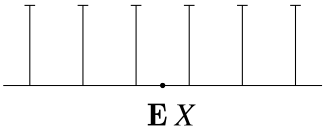
Ако сл.в. X има само краен брой стойности, то горната сума е крайна. Тогава математическото очакване $\mathbf{E}\,X$ задължително съществува и попада в интервала, в който се менят стойностите на случайната величина X.

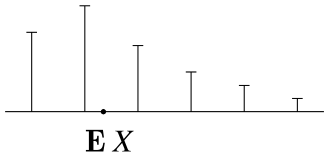
Ако стойностите на сл.в. X са изброим брой, то е възможно сумата да е разходяща. Тогава казваме, че не съществува математическо очакване.

По-общо, ако търсим очакването на функция от случайна величина $\mathbf{E}\,(g(X))$, съгласно формулата за смяна на променливите

$$\mathbf{E}\,(g(X)) = \sum_i g(x_i)\,\mathbf{P}(X = x_i) = \sum_i g(x_i)\,p_i. \tag{II .2}$$

Като синоним на „математическо очакване“ се използва и изразът „средна стойност“. Наистина, намирането на математическото очакване всъщност е пресмятане на средната стойност на случайната величина, като вероятностите p_i играят ролята на тегла. Физическия смисъл на математическото очакването е център на тежестта.





В частния случай, когато всички стойности се падат с една и съща вероятност, т.е. всички p_i са равни, математическото очакване е просто средното аритметично. В противен случай то се измества към по вероятните стойности.

Вероятностния смисъл на математическото очакване се дава от закона за големите числа, който ще бъде доказан по-късно. Съгласно този закон, ако измервате многократно стойностите на една случайна величина, средното аритметично на тези стойности клони към математическото очакване, т.е. средния резултат от много опити се дава от математическото очакване.

Пример

Ще разгледаме игра на рулетка, в която играчът залага 10 лв. само на едно число. Известно е, че числата в рулетката са от 1 до 36, като се добавя и 0 (в американската рулетка освен това има и 00). Ако играчът заложи на едно число и улучи, получава печалба 35 пъти по-голяма от залога. Нека сл.в X е чистата печалба на играча от една игра, т.е. включваме и залогът, който той е направил. Разпределението на X е следното:

X	-10	350
$\mathbf{P}$	$36/37$	$1/37$

Съгласно (II .1) за математическото очакване на X получаваме:

$$\mathbf{E}\,X = -10\,\frac{36}{37} + 350\,\frac{1}{37} = -\frac{10}{37} \approx -0,27\text{лв.}$$

Математическото очакване е отрицателно, което означава, че играчът губи средно по 0,27 лв. на игра. Съответно казиното печели средно 0,27 лв. на игра. Разбира се X е случайна величина и при малък брой игри е възможно играчът да спечели, както и да загуби, въпрос на късмет. При голям брой игри обаче, печалбата ще клони към математическото очакване. Това важи за казиното, в което на ден има хиляди игри. Така собственикът може да пресметне печалбата си, независимо че тя зависи от случайни фактори. Например, при 10000 игри от горния тип печалбата за казиното ще е приблизително 2700 лв.

Игри, в които очакваната печалба е нула ($\mathbf{E}\,X = 0$) се наричат справедливи игри (такава би била играта на рулетка, ако в нея няма нула). В справедлива игра, ако играчът играе достатъчно дълго, капиталът му ще се запази на началното ниво, той нито ще спечели, нито ще загуби.

Ще докажем някои по-важни свойства на математическото очакване.

❧ Ако $c = const$, то $\mathbf{E}\, c = c$.

Док.Константите могат да се разглеждат като случайни величини, които взимат една единствена стойност с вероятност единица $\mathbf{P}(X = c) = 1$, т.е. тяхното разпределение е от типа:

X	C
$\mathbf{P}$	1

Твърдението следва непосредствено от дефиницията (II .1). ✓

❧ $\mathbf{E}\,(cX) = c\,\mathbf{E}\,X$, където X е произволна сл.в., а $c = const$.

Док.

$$\mathbf{E}\,(cX) = \sum_i c x_i\, \mathbf{P}(X = x_i) = c \sum_i x_i\, \mathbf{P}(X = x_i) = c\,\mathbf{E}\,X$$

✓

❧ $\mathbf{E}\,(X+Y) = \mathbf{E}\,X + \mathbf{E}\,Y$, където X и Y са произволни сл.в.

Док.

$$\mathbf{E}\,(X+Y) = \sum_i \sum_j (x_j + y_i)\, \mathbf{P}(X = x_j, Y = y_i) =$$

$$= \sum_i \sum_j x_j\, \mathbf{P}(X = x_j, Y = y_i) + \sum_i \sum_j y_i\, \mathbf{P}(X = x_j, Y = y_i) = \qquad \text{(II .3)}$$

$$= \sum_j x_j \sum_i \mathbf{P}(X = x_j, Y = y_i) + \sum_i y_i \sum_j \mathbf{P}(X = x_j, Y = y_i)$$

Събитията $Y = y_i, i = 1\ldots$ образуват пълна група от събития. Тогава съгласно формула за пълна вероятност

$$\sum_i \mathbf{P}(X = x_j, Y = y_i) = \mathbf{P}(X = x_j).$$

Аналогично се преработва и последната сума в (II .3). Така получаваме:

$$\mathbf{E}\,(X+Y) = \sum_j x_j\, \mathbf{P}(X = x_j) + \sum_i y_i\, \mathbf{P}(Y = y_i) = \mathbf{E}\,X + \mathbf{E}\,Y.$$

✓

Съгласно последните две свойства математическото очакване е линейно.

❧ Нека $X \perp\!\!\!\perp Y$ са независими сл.в., тогава $\mathbf{E}\,(XY) = \mathbf{E}\,X\,\mathbf{E}\,Y$.

Док.От независимостта на случайните величини следва, че за всеки i, j е изпълнено $\mathbf{P}(X = x_j, Y = y_i) = \mathbf{P}(X = x_j)\,\mathbf{P}(Y = y_i)$. Тогава

$$\mathbf{E}\,(X\,Y) = \sum_i \sum_j x_j\, y_i\, \mathbf{P}(X = x_j, Y = y_i) =$$

$$= \sum_i \sum_j x_j y_i \mathbf{P}(X = x_j)\mathbf{P}(Y = y_i) = \sum_i y_i \mathbf{P}(Y = y_i) \left(\sum_j x_j \ \mathbf{P}b(X = x_j) \right) =$$

$$= \sum_i y_i \ \mathbf{P}(Y = y_i)\mathbf{E}\, X = \mathbf{E}\, X\, \mathbf{E}\, Y.$$

✓

Математическото очакване дава възможност да се оцени вероятността случайната величина да вземе големи стойности. Тази оценка се дава в следващото твърдение.

Неравенство на Марков

За произволна неотрицателна случайна величина X и произволна константа A е изпълнено

$$\mathbf{P}(X > A) \leq \frac{\mathbf{E}\, X}{A}$$

Док.Ще извършим доказателството за дискретна сл. в. Нека да разгледаме стойностите на сл. в. X подредени по големина със съответните вероятности $p_i = \mathbf{P}(X = x_i)$, за $i = 1, 2, \ldots$, и нека да съществува константа A, такава че

$$x_1 < x_2 \ldots x_k < A \leq x_{k+1} < \ldots.$$

Тогава

$$\mathbf{E}\, X = \sum p_i x_i \geq \sum_{i=k+1} p_i x_i \geq \sum_{i=k+1} A p_i = A\mathbf{P}(X > A),$$

откъдето непосредствено следва твърдението на теоремата.

Неравенството на Марков рядко се използва за оценяване на вероятността в практически задачи, за целта разполагаме с по точни методи. В определени случай, когато $\mathbf{E}\, X$ е прекалено голямо (или A прекалено малко), е възможно дясната страна на неравенството да надскочи единица и тогава неравенството на Марков не носи никаква информация.

Все пак има интересно следствие от неравенството на Марков.
...

4. Дисперсия

Дефиниция - Дисперсия - $\mathbf{E}\, X$

Дисперсия на случайната величина X наричаме числото:

$$\mathbf{D}\, X = \mathbf{E}\ (X - \mathbf{E}\, X)^2$$

Дисперсията е мярка за разсейването на стойностите на една случайна величина около нейното математическо очакване.

Числото $\sqrt{\mathbf{D}\,X}$ наричаме стандартно отклонение.

В началото ще изведем по-лесен начин за пресмятане на дисперсията.

Твърдение

$$\mathbf{D}\,X = \mathbf{E}\,X^2 - (\mathbf{E}\,X)^2.$$

Док. Ще използваме формулата за съкратено умножение и ще приложим свойства **E2** и **E3** на математическото очакване. Тогава

$$\mathbf{D}\,X = \mathbf{E}\,(X - \mathbf{E}\,X)^2 = \mathbf{E}\left(X^2 - 2X\,\mathbf{E}\,X + (\mathbf{E}\,X)^2\right) =$$
$$= \mathbf{E}\,X^2 - \mathbf{E}\,(2X\,\mathbf{E}\,X) + \mathbf{E}\,((\mathbf{E}\,X)^2),$$

където сме използвали, че $\mathbf{E}\,X$ е число, т.е. константа, тогава $(\mathbf{E}\,X)^2$ също е константа и от свойство **E1** следва $\mathbf{E}\,((\mathbf{E}\,X)^2) = (\mathbf{E}\,X)^2$. Аналогично от свойство **E2** следва $\mathbf{E}\,(2X\,\mathbf{E}\,X) = 2(\mathbf{E}\,X)(\mathbf{E}\,X)$. По този начин получаваме:

$$\mathbf{D}\,X = \mathbf{E}\,X^2 - 2(\mathbf{E}\,X)(\mathbf{E}\,X) + (\mathbf{E}\,X)^2 = \mathbf{E}\,X^2 - (\mathbf{E}\,X)^2.$$

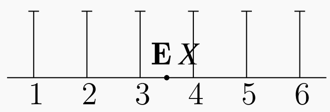
✓

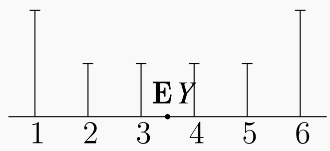
Пример

Хвърляме два зара. Първият е правилен, а за вторият вероятностите да се паднат едно и шест са по $1/4$, а вероятностите да се падне някоя от останалите цифри са $1/8$. Нека X и Y са точките паднали се съответно върху първия и втория зар. Ще пресметнем очакването и дисперсията на X и Y. Съгласно условието разпределениеята на X и Y са съответно:

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Y	1	2	3	4	5	6
P	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$





$$\mathbf{E}\,X = \frac{7}{2} \qquad \mathbf{E}\,Y = \frac{7}{2}$$

$$\mathbf{E}\,X^2 = \frac{1^2}{6} + \frac{2^2}{6} + \ldots + \frac{6^2}{6} = \frac{91}{6} \qquad \mathbf{E}\,Y^2 = \frac{1^2}{4} + \frac{2^2}{8} + \ldots + \frac{6^2}{4} = 16$$

$$\mathbf{D}\,X = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \qquad \mathbf{D}\,Y = 16 - \left(\frac{7}{2}\right)^2 = \frac{15}{4}$$

Математическото очакване в двата случая съвпада. Дисперсията във втория случай е по-голяма, тъй като има по-голяма вероятност случайната величина да е далеч от математическото си очакване, т.е. разсейването е по-голямо.

Ще докажем по-важните свойства на дисперсията.

❧ $\mathbf{D}\,X \geq 0$.

Док.Тъй като случайната величина $(X - \mathbf{E}\,X)^2 \geq 0$ то и математическото и очакване е неотрицателно, т.е. $\mathbf{D}\,X = \mathbf{E}\,(X - \mathbf{E}\,X)^2 \geq 0$. ✓

❧ $\mathbf{D}\,c = 0$, т.е. разсейването на константите е 0.

Док.

$$\mathbf{D}\,c = \mathbf{E}\,(c - \mathbf{E}\,c)^2 = \mathbf{E}\,(c - c)^2 = \mathbf{E}\,0 = 0.$$

✓

❧ $\mathbf{D}\,(cX) = c^2\mathbf{D}\,X$.

Док.

$$\mathbf{D}\,(cX) = \mathbf{E}\,(cX - \mathbf{E}\,(cX))^2 = \mathbf{E}\,(cX - c\mathbf{E}\,X)^2 =$$

$$= \mathbf{E}\left[c^2\,(X - \mathbf{E}\,X)^2\right] = c^2\mathbf{E}\,(X - \mathbf{E}\,X)^2 = c^2\mathbf{D}\,X$$

✓

❧ Нека $X \perp\!\!\!\perp Y$, тогава $\mathbf{D}\,(X + Y) = \mathbf{D}\,X + \mathbf{D}\,Y$.

Док.Ще използваме своиство **E3** на математическото очакване.

$$\mathbf{D}\,(X + Y) = \mathbf{E}\,[X + Y - \mathbf{E}\,(X + Y)]^2 = \mathbf{E}\,[(X - \mathbf{E}\,X) + (Y - \mathbf{E}\,Y)]^2 =$$

$$= \mathbf{E}\,(X - \mathbf{E}\,X)^2 + \mathbf{E}\,(Y - \mathbf{E}\,Y)^2 + 2\mathbf{E}\,[(X - \mathbf{E}\,X)(Y - \mathbf{E}\,Y)] =$$

$$= \mathbf{D}\,X + \mathbf{D}\,Y + 2\mathbf{E}\,[(X - \mathbf{E}\,X)(Y - \mathbf{E}\,Y)]$$

За да завършим доказателството е достатъчно да покажем, че последното събираемо е нула. Случайните величиниX и Y са независими и съгласно **E4** $\mathbf{E}\,(XY) = \mathbf{E}\,X\,\mathbf{E}\,Y$. Тогава:

$$\mathbf{E}\,[(X - \mathbf{E}\,X)(Y - \mathbf{E}\,Y)] = \mathbf{E}\,(XY - Y\mathbf{E}\,X - X\mathbf{E}\,Y + \mathbf{E}\,X\mathbf{E}\,Y) =$$

$$= \mathbf{E}\,(XY) - \mathbf{E}\,Y\mathbf{E}\,X - \mathbf{E}\,X\mathbf{E}\,Y + \mathbf{E}\,(\mathbf{E}\,X\mathbf{E}\,Y) = \mathbf{E}\,(XY) - \mathbf{E}\,Y\mathbf{E}\,X = 0.$$

✓

Неравенство на Чебишев

За произволна случайна величина X с крайни математическо очакване и дисперсия и произволно $\varepsilon > 0$ е изпълнено

$$\mathbf{P}(|X - \mathbf{E}\,X| \geq \varepsilon) \leq \frac{\mathbf{D}\,X}{\varepsilon^2},$$

или което е еквивалентно

$$\mathbf{P}(|X - \mathbf{E}\,X| < \varepsilon) \geq 1 - \frac{\mathbf{D}\,X}{\varepsilon^2}.$$

Доказателство: Ще извършим доказателството за дискретна сл. в.

Без ограничение на общността можем да считаме, че съществува $\varepsilon > 0$, такова че първите k стойности на сл. в. $|X - \mathbf{E}\,X|$ са по-малки от това $\varepsilon > 0$, а останалите стойности са не по-малки от нея. Тогава

$$\mathbf{P}(|X - \mathbf{E}\,X| < \varepsilon) = 1 - \sum_{i=k+1}^{\infty} p_i,$$

където $p_i = \mathbf{P}(X = x_i)$. За да намерим $\sum_{i=k+1}^{\infty} p_i$ ще разгледаме дисперсията на сл. в. $|X - \mathbf{E}\,X|$:

$$\mathbf{D}\,X = \sum_{i=1}^{\infty} p_i(x_i - \mathbf{E}\,X)^2 = \sum_{i=1}^{k} p_i(x_i - \mathbf{E}\,X)^2 + \sum_{i=k+1}^{\infty} p_i(x_i - \mathbf{E}\,X)^2 \geq \sum_{i=k+1}^{\infty} p_i(x_i - \mathbf{E}\,X)^2$$

$$= \varepsilon^2 \sum_{i=k+1}^{\infty} p_i,$$

откъдето

$$\sum_{i=k+1}^{\infty} p_i \leq \frac{\mathbf{D}\,X}{\varepsilon^2}.$$

Окончателно получаваме, че

$$\mathbf{P}(|X - \mathbf{E}\,X| < \varepsilon) \geq 1 - \frac{\mathbf{D}\,X}{\varepsilon^2}.$$

5. Пораждащи функции

Пораждащите функции улесняват пресмятането на вероятности, свързани с дискретните сл.в., както и на техните характеристиките.

П.Ф. се дефинират само за сл.в. със стойности цели числа. Съществуват функции с аналогични свойства, които се въвеждат за по-широк клас сл.в. Това са така наречените характеристични функции.

Дефиниция - Пораждаща функция (П.Ф.)

Нека X е случайна величина, чиито стойности са цели положителни числа. Пораждаща функция на X наричаме:

$$g_X(s) = \sum_{k=0}^{\infty} \mathbf{P}(X = k), \quad |s| \leq 1. \qquad \text{(II .4)}$$

Пораждаща функция на X е просто полином, в който пред k-тата степен на s стои вероятността $P(X = k)$. П.Ф. може да се изрази и в термините на математическо очакване, а именно:

$$g_X(s) = \mathbf{E}\,(s^X)$$

Пример

Ще пресметнем п.ф. на сл.в. X, която означава броя на точките, паднали се при хвърлянето на зар. Нейното разпределение ни е познато:

X	1	2	3	4	5	6
$\mathbf{P}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Тогава

$$g_X(s) = \frac{1}{6}s + \frac{1}{6}s^2 + \ldots + \frac{1}{6}s^6 = \frac{s}{6}\left(1 + s + \ldots + s^5\right) = \frac{s(1-s^6)}{6(1-s)}.$$

Ако случайната величина взима само краен брой стойности, то сумата е крайна и пораждащата функция е дефинирана за всяко s. Ако стойностите на сл.в. X са изброим брой, то е сигурно, че $g_X(1) = 1$, тъй като сумата от вероятностите е равна на единица. От тук следва, че поне за $|s| \le 1$ пораждащата функция със сигурност е сходяща, т.е. съществува. Това е достатъчно за нашите цели, така че по-нататък няма да разглеждаме въпроса за сходимостта на реда, чрез който се дефинират пораждащите функции.

Ще изведем, някои по-често използвани свойства на пораждащите функции.

❧ За пресмятане на математическото очакване чрез пораждаща функция се използва следната формула:

$\mathbf{E}\,X = g'_X(1)$.

Док.Ще диференцираме (II .4), след което ще положим $s = 1$.

$$g'_X(1) = \sum_{k=1}^{\infty} \mathbf{P}(X=k)\; k\; s^{k-1}\Big|_{s=1} = \sum_{k=1}^{\infty} k\,\mathbf{P}(X=k) = \mathbf{E}\,X.$$

✓

❧ Дисперсията на сл.в. също може да бъде пресметната чрез п.ф.: $\mathbf{D}\,X = g''_X(1) + g'_X(1) - \left(g'_X(1)\right)^2$.

Док.Ще пресметнем втората производна на $g_X(s)$:

$$g''_X(1) = \sum_{k=1}^{\infty} \mathbf{P}(X=k)\; k(k-1)\; s^{k-2}\Big|_{s=1} = \sum_{k=1}^{\infty} k(k-1)\mathbf{P}(X=k) = \mathbf{E}\,X(X-1).$$

В последното равенство използвахме формула (II .2). Сега като използваме

$$g''_X(1) = \mathbf{E}\;(X(X-1)) = \mathbf{E}\,X^2 - \mathbf{E}\,X,$$

следва:

$$\mathbf{E}\,X^2 = g''_X(1) + \mathbf{E}\,X = g''_X(1) + g'_X(1).$$

✓

❧ Пораждащите функциии могат да бъдат използвани за намирането на разпределението на суми от сл.в. Нека X и Y са независими случайни величини, тогава $g_{X+Y}(s) = g_X(s)g_Y(s)$.

Док.Ще образуваме произведението $g_X(s)g_Y(s)$ и ще пресметнем коефициента пред s^k :

$$g_X(s)g_Y(s) = \sum_{i=0}^{\infty} P(X = i)s^i \sum_{j=0}^{\infty} P(Y = j)s^j =$$

$$= P(X = 0)P(Y = 0)s^0 + [P(X = 0)P(Y = 1) + P(X = 1)P(Y = 0)]\, s^1 + \ldots$$

$$\ldots + \sum_{i=0}^{k} P(X = i)P(Y = k - i)s^k + \ldots$$

Ще запишем това равенство в затворен вид

$$g_X(s)g_Y(s) = \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} P(X = i)P(Y = k - i) \right) s^k. \tag{II .5}$$

Тъй като X и Y са независими сл.в., то

$$P(X = i)P(Y = k - i) = P(X = i, Y = k - i) = P(X = i, X + Y = k).$$

Ще използваме формулата за пълната вероятност, за да оценим вътрешната сума в (II .5):

$$\sum_{i=0}^{k} P(X = i)P(Y = k - i) = \sum_{i=0}^{k} P(X = i, X + Y = k) = P(X + Y = k).$$

Замествайки този резултат в (II .5) получаваме търсената пораждаща функция на случайната величина $X + Y$:

$$g_X(s)g_Y(s) = \sum_{k=0}^{\infty} P(X + Y = k)\, s^k = g_{X+Y}(s).$$

✓

По индукция този резултат се обобщава за повече от две независими случайни величини:

$$g_{X_1+X_2+\ldots X_n}(s) = g_{X_1}(s)\, g_{X_2}(s) \ldots g_{X_n}(s). \tag{II .6}$$

Хвърлят се 10 зара. Да се намери вероятността сумата от падналите се точки да бъде точно 19. Директното пресмятане на тази задача е твърде трудоемко. Това би означавало да се пресметнат всички начини числото 19 да се представи като сума на 10 цели числа в диапазона от 1 до 6. Затова ще използваме възможностите на апарата на пораждащите функции.

С X_i, $i = 1 \dots 10$ ще означим точките паднали се върху i-тия зар. В предишния пример изведохме пораждащата функция на тези случайни величини, а именно

$$g_{X_i}(s) = \frac{s(1-s^6)}{6(1-s)}.$$

Нека Y е сумата от падналите се точки, т.е. $Y = X_1 + X_2 + \dots + X_{10}$. Точките паднали се върху един зар по никакъв начин не влияят върху точките паднали се върху друг. Следователно, случайните величини X_i са независими. Тогава, съгласно равенство (II .6) пораждащата функция на Y е произведение от пораждащите функции на X_i, $i = 1 \dots 10$:

$$g_Y(s) = \prod_{i=1}^{10} g_{X_i}(s) = \prod_{i=1}^{10} \left(\frac{s(1-s^6)}{6(1-s)}\right) = \frac{s^{10}(1-s^6)^{10}}{6^{10}(1-s)^{10}}.$$

Съгласно дефиницията на пораждащата функция търсената вероятност ще бъде коефициента пред 19-тата степен на s в тази функция.

За да пресметнем коефициента на s^{19} ще развием тази функция по степените на s. Ще използваме формулата за бином на Нютон

$$(1-a)^n = \sum_{k=0}^{n} \binom{n}{k} (-a)^k$$

за да преобразуваме числителя. За знаменателя ще използваме формулата за отрицателен бином:

$$(1-a)^{-n} = \sum_{l=0}^{\infty} \binom{n+l-1}{l} a^l = \sum_{l=0}^{\infty} \binom{n+l-1}{n-1} a^l.$$

По този начин за пораждащата функция на Y получаваме:

$$g_Y(s) = \frac{s^{10}}{6^{10}} \left[\sum_{k=0}^{10} \binom{10}{k} (s^6)^k\right] \sum_{l=0}^{\infty} \binom{9+l}{l} s^l =$$

$$g_Y(s) = \frac{s^{10}}{6^{10}} \left[1 - \binom{10}{1} s^6 + \binom{10}{2} s^{12} + \dots\right] \sum_{l=0}^{\infty} \binom{9+l}{l} s^l =$$

Пред сумите стои s^{10}, следователно от произведението на двете суми трябва да получим s^9. Това може да стане само по два начина. Да вземем единица от първата сума и да я умножим с s^9 от втората сума. Или да вземем s^6 от първата и s^3 от втората сума. Останалите събираеми в първата сума са със степен равна или по-голяма от 12, тъй че няма как да се използват.

Окончателно, за търсената вероятност получаваме

$$\mathbf{P}(Y = 19) = coeff_{s^{19}}\left\{g_Y(s)\right\} = \frac{1}{6^{10}}\left[\binom{18}{9} - \binom{10}{1}\binom{12}{3}\right].$$

III . Двумерни дискретни случайни величини

Дефиниция - Съвместно разпределение

Съвместно разпределение на случайните величини X и Y, наричаме следната таблица:

$Y \backslash X$	x_1	x_2	$\dots$	x_n	$\dots$
y_1	$p_{1,1}$	$p_{1,2}$	$\dots$	$p_{1,n}$	$\dots$
y_2	$p_{2,1}$	$\dots$			$\dots$
$\dots$	$\dots$				
y_m	$p_{m,1}$	$p_{m,2}$	$\dots$	$p_{m,n}$	$\dots$
$\dots$	$\dots$				

където x_i и y_j са съответно стойностите на сл.в. X и Y и те могат да бъдат краен или най-много изброим брой;
$p_{j,i} = \mathbf{P}(X = x_i, Y = y_j)$ са вероятностите с които случайните величини вземат съответните стойности, при това

$$\sum_{i,j} p_{j,i} = 1$$

Пример

Хвърляме два зара. Нека случайната величина X е броят на шестиците, а Y броят на единиците паднали се върху заровете. Ще намерим съвместното разпределение на X и Y.

Ясно е, че X и Y могат да вземат като стойности числата 0, 1 и 2. Събитието $\{X = 0, Y = 0\}$ означава, че върху заровете не се е паднала нито една шестица или единица, тогава $\mathbf{P}(X = 0, Y = 0) = (4/6)^2$. Аналогично $\mathbf{P}(X = 1, Y = 0) = 2(1/6)4/6$ и т.н.

Съвместното разпределение на X и Y има вида:

$Y \backslash X$	0	1	2
0	$\frac{16}{36}$	$\frac{8}{36}$	$\frac{1}{36}$
1	$\frac{8}{36}$	$\frac{2}{36}$	0
2	$\frac{1}{36}$	0	0

✓

Ако разполагаме със съвместното разпределение на X и Y не е проблем да се пресметне разпределението само на едната случайна величина. Това разпределение се нарича *маргинално разпределение*. Намирането му става, съгласно формула за пълната вероятност, чрез сумиране по ред или по стълб, т.е.

$$\mathbf{P}(X = x_i) = \sum_j \mathbf{P}(X = x_i, Y = y_j), \qquad i = 1, 2, \ldots$$

Аналогично

$$\mathbf{P}(Y = y_j) = \sum_i \mathbf{P}(X = x_i, Y = y_j), \qquad j = 1, 2, \ldots$$

Ще намерим маргиналните разпределения на случайните величини от предходния пример. Маргиналните разпределения често се записват отдолу и отстрани на таблицата със съвместните разпределения

$Y \backslash X$	0	1	2	
0	$\frac{16}{36}$	$\frac{8}{36}$	$\frac{1}{36}$	$\frac{25}{36}$
1	$\frac{8}{36}$	$\frac{2}{36}$	0	$\frac{10}{36}$
2	$\frac{1}{36}$	0	0	$\frac{1}{36}$
	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$	

1. Ковариация. Корелация.

По-нататък ще въведем две понятия, които се използват като мярка за линейната зависимост между случайните величини X и Y.

Ковариация

Ковариация на случайните величини X и Y наричаме числото:

$$cov(X,Y) = \mathbf{E}\,[(X - \mathbf{E}\,X)(Y - \mathbf{E}\,Y)]. \qquad \text{(III .1)}$$

Ако $cov(X,Y) = 0$ казваме, че случайните величини са некорелирани.

За пресмятане на ковариацията обикновено се използва представянето дадено в следващото твърдение.

Твърдение

$$cov(X,Y) = \mathbf{E}\,(XY) - \mathbf{E}\,X\,\mathbf{E}\,Y. \qquad \text{(III .2)}$$

Док.Ще използваме линейността на математическото очакване.

$$cov(X,Y) = \mathbf{E}\left[(X-\mathbf{E}\,X)(Y-\mathbf{E}\,Y)\right] = \mathbf{E}\left[(XY-Y\mathbf{E}\,X-X\mathbf{E}\,Y+\mathbf{E}\,X\mathbf{E}\,Y)\right] =$$

$$= \mathbf{E}\,(XY) - \mathbf{E}\,(Y\,\mathbf{E}\,X) - \mathbf{E}\,(X\,\mathbf{E}\,Y) + \mathbf{E}\,(\mathbf{E}\,X\,\mathbf{E}\,Y),$$

където сме използвали, че $\mathbf{E}\,X$ е константа. Тогава от свойство **E2** следва $\mathbf{E}\,(Y\mathbf{E}\,X) = \mathbf{E}\,X\mathbf{E}\,Y$. Аналогично $\mathbf{E}\,(X\mathbf{E}\,Y) = \mathbf{E}\,X\mathbf{E}\,Y$. Съгласно **E1** математическото очакване на константа е самата константа, т.е. $\mathbf{E}\,(\mathbf{E}\,X\,\mathbf{E}\,Y) = \mathbf{E}\,X\,\mathbf{E}\,Y$. С това доказателстово е завършено. ◻

Твърдение

Ако случайните величини X и Y са независими, то

$$cov(X,Y) = 0.$$

Док.Твърдението следва директно от (III .2) и свойство **E4** на математическото очакване.

В общия случай обратното твърдение не е изпълнено. Възможно е $cov(X,Y) = 0$, но случайните величини да бъдат зависими.

Дефиниция - Корелация

Коефициент на корелация $\rho_{X,Y}$ на случайните величини X и Y наричаме:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{\mathbf{D}\,X}\sqrt{\mathbf{D}\,Y}}. \tag{III .3}$$

Твърдение

Коефициент на корелация $|\rho_{X,Y}| \leq 1$ за произволни случайни величини X и Y.

Док.Ще разгледаме следната случайна величина

$$\left[\frac{X-\mathbf{E}\,X}{\sqrt{\mathbf{D}\,X}} + \frac{Y-\mathbf{E}\,Y}{\sqrt{\mathbf{D}\,Y}}\right]^2 \geq 0 \tag{III .4}$$

Тя е неотрицателна, следователно и математическото и очакване е неотрицателно. Имаме следната верига:

$$0 \leq \mathbf{E}\left[\frac{X-\mathbf{E}\,X}{\sqrt{\mathbf{D}\,X}} + \frac{Y-\mathbf{E}\,Y}{\sqrt{\mathbf{D}\,Y}}\right]^2 = \tag{III .5}$$

$$= \frac{\mathbf{E}\,(X-\mathbf{E}\,X)^2}{\mathbf{D}\,X} + \frac{\mathbf{E}\,(Y-\mathbf{E}\,Y)^2}{\mathbf{D}\,Y} + 2\frac{\mathbf{E}\,(X-\mathbf{E}\,X)(Y-\mathbf{E}\,Y)}{\sqrt{\mathbf{D}\,X}\sqrt{\mathbf{D}\,Y}} =$$

$$= \frac{\mathbf{D}\,X}{\mathbf{D}\,X} + \frac{\mathbf{D}\,Y}{\mathbf{D}\,Y} + 2\rho_{X,Y} = 2 + 2\rho_{X,Y}.$$

В последното равенство приложихме дефинициите на дисперсия и корелация. Сега от $2 + 2\rho_{X,Y} \geq 0$ елементарно следва $\rho_{X,Y} \geq -1$.

Аналогично, за да се докаже неравенството $\rho_{X,Y} \leq 1$ се разглежда случайната величина

$$\left[\frac{X - \mathbf{E}\,X}{\sqrt{\mathbf{D}\,X}} - \frac{Y - \mathbf{E}\,Y}{\sqrt{\mathbf{D}\,Y}}\right]^2 \geq 0.$$

Твърдение

$|\rho_{X,Y}| = 1$ тогава и само тогава, когато случайните величини X и Y са линейно зависими.

Док.Нека X и Y са линейно зависими, т.е. съществуват константи a и b, такива че $X = aY + b$. Ще докажем че $|\rho_{X,Y}| = 1$. Наистина тогава

$$\mathbf{E}\,X = \mathbf{E}\,(aY + b) = a\,\mathbf{E}\,Y + b,$$
$$\mathbf{D}\,X = \mathbf{D}\,(aY + b) = \mathbf{D}\,(aY) + \mathbf{D}\,b = a^2\mathbf{D}\,Y.$$

Следователно

$$\rho_{X,Y} = \frac{\mathbf{E}\,[(X - \mathbf{E}\,X)(Y - \mathbf{E}\,Y)]}{\sqrt{\mathbf{D}\,X}\sqrt{\mathbf{D}\,Y}} = \frac{\mathbf{E}\,[(aY + b - (a\,\mathbf{E}\,Y + b))(Y - \mathbf{E}\,Y)]}{\sqrt{a^2\mathbf{D}\,Y}\sqrt{\mathbf{D}\,Y}} =$$

$$= \frac{\mathbf{E}\,[a(Y - \mathbf{E}\,Y))(Y - \mathbf{E}\,Y)]}{|a|\,\mathbf{D}\,Y} = \frac{a\,\mathbf{E}\,(Y - \mathbf{E}\,Y)^2}{|a|\,\mathbf{D}\,Y} = \frac{a}{|a|}$$

Последният израз е равен на ± 1 в зависимост от знака на a. С това твърдението е доказано в едната посока.

Забележка. Ако $a > 0$, т.е. когато едната случайна величина расте и другата расте, тогава е изпълнено $\rho_{X,Y} = 1$. Обратно Ако $a < 0$, т.е. когато едната случайна величина расте, а другата намалява $\rho_{X,Y} = -1$.

Нека сега $|\rho_{X,Y}| = 1$. Ще докажем, че случайните величини X и Y са линейно зависими.

За определеност ще приемем, че $\rho_{X,Y} = -1$. Отново ще разгледаме случайната величина (III .4). Съгласно равенства (III .5)

$$\mathbf{E}\left[\frac{X - \mathbf{E}\,X}{\sqrt{\mathbf{D}\,X}} + \frac{Y - \mathbf{E}\,Y}{\sqrt{\mathbf{D}\,Y}}\right]^2 = 2 + 2\rho_{X,Y} = 0$$

След като очакването на една неотрицателна случайна величина е 0. То и самата случайна величина е равна на нула, т.е.

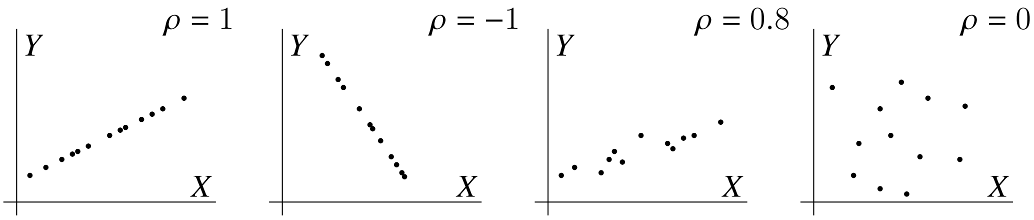
$$\frac{X - \mathbf{E}\,X}{\sqrt{\mathbf{D}\,X}} + \frac{Y - \mathbf{E}\,Y}{\sqrt{\mathbf{D}\,Y}} = 0$$

Ще напомним, че в това равенство $\mathbf{E}\,X$, $\mathbf{E}\,Y$, $\mathbf{D}\,X$ и $\mathbf{D}\,Y$ са константи. Можем да изразим X по следния начин

$$X = \left(-\frac{\sqrt{\mathbf{D}\,X}}{\sqrt{\mathbf{D}\,Y}}\right)Y + \frac{\sqrt{\mathbf{D}\,X}\,\mathbf{E}\,Y}{\sqrt{\mathbf{D}\,Y}} + \frac{\mathbf{E}\,X}{\sqrt{\mathbf{D}\,Y}},$$

което означава линейна зависимост между X и Y. ✓

На следващите схеми е показано влиянието на коефициента на корелация върху степента на линейна зависимост между случайните величини X и Y. Всяко наблюдение над случайните величини е представено като точка с координати (X, Y).



При $\rho = 1$ точките лежат върху растяща права.

При $\rho = -1$ правата е намаляваща.

При $\rho = 0.8$ точките са разположени в околност на права.

При $\rho = 0$ точките са разпръснати.

Не съществува връзка между коефициента на корелация ρ и наклона на правата. Знакът на ρ показва дали правата е растяща или намаляваща. А големината на ρ доколко силна е линейната зависимост. Обърнете внимание, че става дума само за линейна зависимост. Възможно е да съществува друг тип връзка между случайните величини, която няма как да се установи с коефициента на корелация.

IV . Дискретни разпределения

Дискретните случайни величини, които се срещат често в практически задачи, са добре изучени, класифицирани са, свойствата им са известни. В тази глава ще разгледаме някои от най-важните дискретни случайни величини.

1. Разпределение на Бернули - $X \in B(p)$

Това разпределение е кръстено на името на швейцарския математик Якоб Бернули. „Опит на Бернули“ наричаме опит, при който има само две възможности, наречени „успех“ с вероятност p или “неуспех’"с вероятност $q = 1 - p$. Стандартният пример е хвърляне на една монета. Съответно, случайната величина с разпределение на Бернули може да взима само две стойности - „1“ при успех и „0“ при неуспех, т.е. разпределението и има вида:

X	0	1
P	q	p

Елементарно се пресмятат $\mathbf{E}\,X = p$ и $\mathbf{D}\,X = pq$.

2. Биномно разпределение - $X \in Bi(n, p)$

Биномното разпределение може да се разглежда, като обобщение на разпределението на Бернули.

Извършваме последователни, независими бернулиеви опити, като вероятността за успех - p е една и съща при всеки опит, съответно вероятността за неуспех е q. Нека броят на опитите - n е предварително фиксиран. Случайната величина X равна на броя на успехите, наричаме биномно разпределена сл.в. Символично това се записва по следния начин: $X \in Bi(n, p)$.

$$\overbrace{\begin{array}{cccccccc} & * & * & & * & & & * \\ 0 & 1 & 1 & 0 & 1 & \dots & 0 & 1 \\ q & p & p & & \dots & & & \end{array}}^{X}$$

Ясно е, че стойностите на X, т.е. броят на успехите е цяло число в интервала от 0 до n. Ще пресметнем вероятността за точно k на брой успеха. Общо са проведени n опита, на k от които има успех. Съществуват C_n^k начина да изберем опитите, при които да има успех. Ако вероятността за успех при всеки опит е p, то вероятността за успех на k фиксирани опита ще бъде p^k, тъй като опитите са независими. Аналогично, вероятността на останалите $n-k$ опита да има неуспех ще бъде $(1-p)^{n-k}$. Така за вероятността за точно k успеха от n опита, получаваме:

$$\mathbf{P}(X=k) = C_n^k p^k (1-p)^{n-k} = \binom{n}{k} p^k q^{n-k}, \qquad \text{(IV .1)}$$
$$k = 0, 1, \ldots n.$$

Биномното разпределение е добре дефинирано, тъй като съгласно формулата за бинома на Нютон е изпълнено:

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

Оттук идва и самото наименование на разпределението.

Нека X е броят на шестиците паднали се при хвърлянето на три зара. Тогава $X \in Bi(3, \frac{1}{6})$. Непосредствено от формула (IV .1) се пресмята разпределението на случайната величина X :

X	0	1	2	3
$\mathbf{P}$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

За да пресметнем характеристиките на биномно разпределена случайна величина ще я представим като сума от Бернулиеви сл.в.

Нека с X_i, $i = 1, 2, \ldots, n$ означим успеха на i-тия опит, т.е. случайните величини X_i могат да вземат само две стойности 0, или 1. X_i и имат разпределение на Бернули, като $\mathbf{E}\, X_i = p$ и $\mathbf{D}\, X_i = pq$. Опитите са независими, следователно и случайните величини X_i, $i = 1, 2, \ldots, n$ са независими.

Ясно е, че $X = X_1 + X_2 + \ldots + X_n$. Тогава от свойствата на математическото очакване и дисперсията следва:

$$\mathbf{E}\, X = \mathbf{E}\,(X_1 + X_2 + \ldots + X_n) = \mathbf{E}\, X_1 + \mathbf{E}\, X_2 + \ldots + \mathbf{E}\, X_n = np,$$

$$\mathbf{D}\, X = \mathbf{D}\,(X_1 + X_2 + \ldots + X_n) = \mathbf{D}\, X_1 + \mathbf{D}\, X_2 + \ldots + \mathbf{D}\, X_n = npq.$$

Ще пресметнем и максималната стойност на вероятността, т.е. ще открием за кое $k = 0, 1, \ldots n$ вероятността $\mathbf{P}(X = k)$ достига максимум. Намирането на този максимум не е възможно по традиционния начин, познат от математическият анализ с намиране на първата производна, тъй като биномният коефициент в (IV .1) не може да бъде диференциран. Затова ще разгледаме отношението на вероятностите. Ако е изпълнено неравенството

$$\frac{\mathbf{P}(X = k)}{\mathbf{P}(X = k-1)} > 1,$$

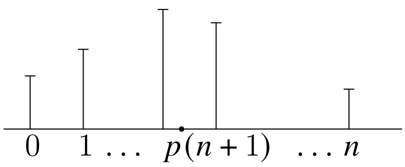
то вероятността $\mathbf{P}(X = k)$, разглеждана като функция по k, е растяща. По този начин ще определим интервалите на растене и намаляване на функцията. Съгласно (IV .1) горното неравенство е еквивалентно на:

$$1 < \frac{C_n^k p^k q^{n-k}}{C_n^{k-1} p^{k-1} q^{n-k+1}} = \frac{n!\,(k-1)!(n-k+1)!\,p}{k!(n-k)!\,n!\,q} = \frac{(n-k+1)\,p}{kq}.$$

Ще решим това неравенство спрямо k

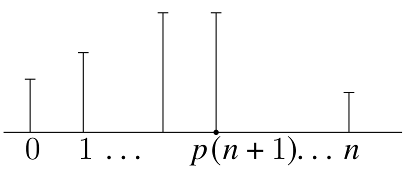
$$p(n+1) > (p+q)k = k. \tag{IV .2}$$

Тук p и $n+1$ са известни константи. Следователно за $k < p(n+1)$ вероятността $\mathbf{P}(X = k)$ е растяща по k. Аналогично, при $k > p(n+1)$ вероятността $\mathbf{P}(X = k)$ е намаляваща. Знаем, че k взима стойности $0, 1, \ldots n$. В началото вероятностите растат, достигат до някаква максимална стойност, след което започват да намаляват. Тогава, разпределението на случайната величина изглежда по следния начин:



Максималната стойност на вероятността се достига за най-голямото цяло число k, което е по-малко от $p(n+1)$, т.е. при k равно на цялата част на $p(n+1)$.

Ако числото $p(n+1)$ се окаже цяло, то неравенство (IV .2) се превръща в равенство за някое k. Тогава ще има две максимални стойности за вероятността $\mathbf{P}(X = k)$ и $\mathbf{P}(X = k-1)$, а разпределението има следния вид:



3. Геометрично разпределение - $X \in Ge(p)$

Отново ще разглеждаме схема на Бернули, т.е. извършваме n последователни независими опити с вероятност за успех на всеки опит p. Случайната величина X равна на броя на неуспехите до достигане на първи успех наричаме геометрично разпределена сл.в. Символично това се записва по следния начин: $X \in Ge(p)$. Броят на опитите не е ограничен, така че стойностите на X могат да варират от 0 до ∞.

Ще пресметнем вероятноста за точно k неуспеха до първия успех, т.е. $\mathbf{P}(X=k)$ за $k=0,1,2,\ldots$. Ясно е, че тази вероятност отговаря на събитието - „при първите k опита има неуспех, а на опит $k+1$ успех“:

$$\overbrace{0\;0\;0\;\ldots\;0}^{X}\;1\;\ldots$$

Опитите са независими, следователно $\mathbf{P}(X=k)=(1-p)^k p = q^k p$. Тук, както обикновено сме означили вероятността за неуспех с q.

За да се покаже, че това разпределение е добре дефинирано, е достатъчно да се приложи формулата за сума на безкрайна геометрична прогресия:

$$\sum_{k=0}^{\infty}\mathbf{P}(X=k)=\sum_{k=0}^{\infty}q^k p=\frac{p}{1-q}=1.$$

Директното пресмятане на математическото очакване и дисперсията на X изисква умения за сумиране на редове. Например,

$$\mathbf{E}\,X=\sum_{k=0}^{\infty}k\,q^k p.$$

Пораждащите функции дават възможност за значително опростяване на този процес. Пораждащата функция на X е сума на безкрайна геометрична прогресия:

$$g_X(s)=\sum_{k=0}^{\infty}\mathbf{P}(X=k)s^k=\sum_{k=0}^{\infty}q^k p s^k=p\sum_{k=0}^{\infty}(qs)^k=\frac{p}{1-qs}.$$

Съгласно свойство **g1**) математическото очакване на X е производната на пораждащата функция при $s=1$:

$$\mathbf{E}\,X=g_X'(1)=\left.\frac{pq}{(1-qs)^2}\right|_{s=1}=\frac{pq}{(1-q)^2}=\frac{q}{p}.$$

За да пресметнем дисперсията на X ще ни трябва втората производна на пораждащата функция:

$$g_X''(1)=\left.\left(\frac{pq}{(1-qs)^2}\right)'\right|_{s=1}=\frac{2pq^2}{(1-q)^3}=\frac{2q^2}{p^2}.$$

Сега, ще използваме свойство **g2**) на пораждащите функции:

$$\mathbf{D}\,X=g_X''(1)+g_X'(1)-\left(g_X'(1)\right)^2=\frac{2q^2}{p^2}+\frac{q}{p}-\frac{q^2}{p^2}=\frac{q(q+p)}{p^2}=\frac{q}{p^2}.$$

4. Отрицателно биномно разпределение - $X \in NB(r,p)$

5. Поасоново разпределение - $X \in Po(\lambda)$

Често се налага да се разглеждат модели, при които се извършват много независими опити, но вероятността за успех при всеки от тях е малка. Интерес представлява броят на успехите X. Тогава разглеждаме модел, в който случайната величина е биномно разпределена $X \in Bi(n,p)$, но $n \to \infty$, $p \to 0$. Това гранично разпределение е изведено от френския математик Симеон Поасон.

Казваме, че случайната величина X е Поасоново разпределена, ако тя взима целочислени стойности с вероятност, съответно

$$\mathbf{P}(X = k) = \frac{\lambda^k \, e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots \tag{IV .3}$$

където $\lambda > 0$ е константа.

Поасоновото разпределение се означава съкратено с $X \in Po(\lambda)$.

Разпределението е добре дефинирано, тъй като

$$\sum_{k=0}^{\infty} \frac{\lambda^k \, e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Поасоновото разпределение се използва за описване на редки събития. Типичен пример е заявки към сървър. Броят на компютрите в мрежата е голям, а вероятността конкретен компютър да потърси връзка е малка. Тогава броят на заявките е Поасоново разпределена случайна величина. Поасоново разпределени се оказват и броят на мутиращите клетки при рентгеново облъчване, броят на получените писма за определен период от време, головете по време на футболна среща и т.н.

Изобщо Поасоновото разпределение се използва, когато пресмятаме броя на сбъдванията X на събитие в определен интервал от време, ако сбъдването на събитието не зависи от времето изминало от сбъдването на предишното събитие, т.е. събитията са независими и освен това имаме предварителна информация за средния брой сбъдвания, т.е. знаем математическото очакване на X.

Следващата теорема дава условията, при които Поасановото разпределение може да се използва като апроксимация за биномното.

Теорема на Поасон. Нека сл.в. $X \in Bi(n, p_n)$, т.е.

$$\mathbf{P}(X = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k}. \qquad \text{(IV .4)}$$

Ако $n \to \infty$, $p_n \to 0$, така че $np_n \to \lambda > 0$, то за всяко фиксирано $k = 0, 1, 2, \ldots$ е изпълнено

$$\lim_{n\to\infty} \mathbf{P}(X = k) = \frac{\lambda^k\, e^{-\lambda}}{k!}.$$

Док.Най-напред ще преработим биномния коефициент, участващ в (IV .4)

$$\binom{n}{k} = \frac{n(n-1)\ldots(n-k+1)}{k!} = \frac{n^k(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{k-1}{n})}{k!}.$$

Ще запишем биномната вероятност $\mathbf{P}(X = k)$ от равенство (IV .4) по следния начин:

$$\mathbf{P}(X = k) = \frac{(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{k-1}{n})}{k!} n^k p_n^k (1 - p_n)^{n-k}. \qquad \text{(IV .5)}$$

Сега ще намерим границите при $n \to \infty$ на отделните множители в този израз.

За всяко $i = 1, 2, \ldots k - 1$ при $n \to \infty$ е изпълнено:

$$\lim_{n\to\infty} \left(1 - \frac{i}{n}\right) = 1.$$

По условие k е фиксирано число, тогава в следното произведение има краен брой, а именно $k - 1$ множителя, следователно

$$\lim_{n\to\infty} \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{k-1}{n}\right) = 1. \qquad \text{(IV .6)}$$

Знаем, че $np_n \to \lambda$, тогава

$$\lim_{n\to\infty} n^k p_n^k = \lim_{n\to\infty} (np_n)^k = \lambda^k. \qquad \text{(IV .7)}$$

От условието $np_n \to \lambda$ следва $p_n \sim \lambda/n$ и тогава:

$$\lim_{n\to\infty} (1 - p_n)^{n-k} = \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k}.$$

Първата граница е добре позната $\lim\limits_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$. За втората граница аналогично на (IV .6) получаваме единица. Следователно

$$\lim_{n\to\infty} (1 - p_n)^{n-k} = e^{-\lambda} \qquad \text{(IV .8)}$$

За да завършим доказателството е достатъчно да извършим граничен преход в (IV .5) и да заместим (IV .6), (IV .7) и (IV .8). ✓

За пресмятането на математическото очакване и дисперсията на поасоновото разпределение ще използваме свойствата на пораждащите функции. Нека $X \in Po(\lambda)$. За пораждащата функция на X получаваме

$$g_X(s) = \sum_{k=0}^{\infty} \mathbf{P}(X = k)s^k = \sum_{k=0}^{\infty} \frac{\lambda^k\, e^{-\lambda}}{k!} s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)}.$$

Тогава математическото очакване на X е

$$\mathbf{E}\, X = g_X'(1) = \lambda e^{\lambda(s-1)}\Big|_{s=1} = \lambda.$$

За дисперсията на X получаваме

$$\mathbf{D}\, X = g_X''(1) + g_X'(1) - \left(g_X'(1)\right)^2 = \lambda^2 e^{\lambda(s-1)}\Big|_{s=1} + \lambda - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

6. Хипергеометрично разпределение - $X \in HG(N, M, n)$

Казваме, че случайната величина X е хипергеометрично разпределена, ако тя взима целочислени стойности с вероятност, съответно

$$\mathbf{P}(X = k) = p_k = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \qquad \text{(IV .9)}$$
$$k = \max\{0, n + M - N\}, \max\{0, n + M - N\} + 1, \ldots, \min\{M, n\}.$$

Хипергеометричното разпределение се означава съкратено с $X \in HG(N, M, n)$.

Хипергеометричното разпределение съответства на урнов модел, при който от урна с M бели от общо N топки се изваждат случайно n топки. Вероятността точно k от $n-$ те да са бели е p_k. Фактът, че $\sum_k p_k = 1$ ледва от равенството

$$\binom{N}{n} = \sum_k \binom{M}{k}\binom{N-M}{n-k},$$

което следва от подреждането на всички извадки на брой $\binom{N}{n}$ в групи, в които броят на белите топки е точно k. Всяка извадка в такава група се представя като $\binom{M}{k}$ извадки на бели топки, комбинирани с $\binom{N-M}{n-k}$ на брой извадки от черни топки.

Свойства на хипергеометричното разпределение:

1.

$$\mathbf{E}\,X = \sum_k k\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} =$$

$$\sum_k k\frac{\frac{M!}{k!(M-k)!}\binom{N-M}{n-k}}{\frac{N!}{n!(N-n)!}} =$$

$$n\frac{M}{N}\sum_k \frac{\frac{(M-1)!}{(k-1)!(M-k)!}\binom{N-M}{n-k}}{\frac{(N-1)!}{(n-1)!(N-n)!}} =$$

$$n\frac{M}{N}\sum_k k\frac{\binom{M-1}{k-1}\binom{N-M}{n-k}}{\binom{N-1}{n-1}} = n\frac{M}{N}.$$

2.

$$Var(X) = n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$$

3.

$$p_k = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \frac{\binom{n}{k}\binom{N-n}{M-k}}{\binom{N}{M}}.$$

Формулата за пълната вероятност дава следното рекурентно уравнение за $p_k(n, M, N)$:

$$p_k(n+1, M, N) = p_k(n, M, N)\frac{N-M-n+k}{N-n} + p_{k-1}(n, M, N)\frac{M-n+k}{N-n}.$$

Хипергеометричното разпределение намира приложение в т. нар. статистически контрол на качеството. В партида от N изделия M са дефектни. При случайно избрани n изделия (извадка без връщане с обем n), броят на дефектните изделия в извадката е сл.в. X с хипергеометрично разпределение

$$\mathbf{P}(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

В играта Спорт-тото "6 от 49" вероятността с един фиш от 6 числа да се спечели тройка, четворка, петица или шестица се пресмята като броят на познатите числа има хипергеометрично разпределение

$$\mathbf{P}(X = k) = \frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}}$$

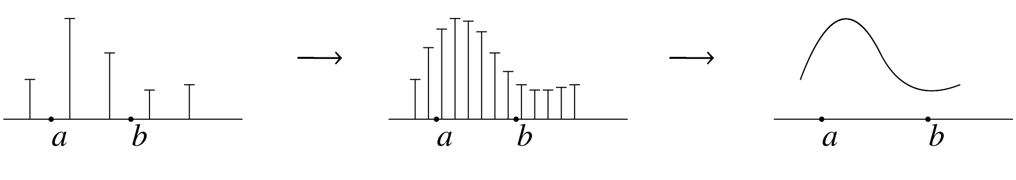
или вероятност за k познати резултата от $M = 6$ печеливши от общо $N = 49$ числа.

7. Полиномно разпределение

V . Непрекъснати случайни величини

До сега разглеждахме дискретни случайни величини, чиито стойности са краен или най-много изброим брой. Това ограничение върху броя на стойностите прави дискретните случайни величини неудобни за описване на редица явления. Например, температурата на въздуха се описва с реално число в интервала $(-45, 52)$, влажността, количеството въглероден двуокис също са реални числа. Затова се налага разглеждането на случайни величини стойностите на които са подмножество на реалните числа, т.е. могат да вземат неизброим брой различни стойности. Такива случайни величини ще наричаме непрекъснати.

За непрекъснатите случайни величини би било безсмислено да се въвежда разпределение под формата на таблица, тъй като не е възможно описването на всички стойности. Затова, като аналог на разпределенията се използва функция наречена *вероятностна плътност*, която играе ролята на вероятност. Формално, непрекъснатите случайни величини се дефинират като се извършва граничен преход по дискретните случайни величини. Можем да си представим този процес като вземем една дискретна случайна величина и увеличаваме броят на нейните стойности все повече и повече, докато нейното разпределение се превърне в непрекъснатата функция в ероятностнаплътност. Това е илюстрирано в следващата схема.



1. Вероятностна плътност

В този учебник няма да се спираме на формалната (аксиоматична) дефиниция на понятието непрекъсната случайна величина. Вместо това ще дефинираме функцията вероятностна плътност, а чрез нея и функцията на разпределеие на случайната величина.

Дефиниция - Плътност

Вероятностна плътност на непрекъснатата случайната величина X наричаме функцията $f_X(x)$, изпълняваща следните условия:

$$1) \quad f_X(x) \geq 0, \tag{V .1}$$

$$2) \quad \int_{-\infty}^{\infty} f_X(x)dx = 1, \tag{V .2}$$

$$3) \quad \mathbf{P}(a \leq X < b) = \int_a^b f_X(x)dx. \tag{V .3}$$

Тази дефиниция е до голяма степен аналогична на дефиницията на разпределение на дискретна случайна величина, т.к.:

- условието 1) отговаря на условието вероятностите да са положителни;

- условието 2) означава нормираност, т.е. съответства на условието сумата от всички вероятности в разпределението да бъде единица;

- условието 3) дава вероятността за попадане на случайната величина в някакво множество - като се сумират, в случая интегрират, вероятностите на благоприятните случаи.

2. Функцията на разпределение

Дефиниция - Функция на разпределение

Функция на разпределение на случайната величина X наричаме функцията $F_X(x)$, дефинирана за всяко реално число x с равенството:

$$F_X(x) = \mathbf{P}(X < x). \tag{V .4}$$

Съгласно (V .3) функцията на разпределение може да се изрази чрез плътността по следния начин

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt. \tag{V .5}$$

Следователно обратната връзка е от вида

$$f_X(x) = \frac{dF_X(x)}{dx}. \tag{V .6}$$

Тези свойства произтичат от определение 2. на ф.р.

Твърдение

Всяка ф.р. $0 \leq F_X(x) = \mathbf{P}(X < x)$ удовлетворява следните свойства:

1. $0 \leq F_X(x) \leq 1$;

2. Ако $x_1 \le x_2$, то $F_X(x_1) \le F_X(x_2)$;
3. $\lim_{x \uparrow x_0} F_X(x) = F_X(x_0^-)$;
4. $\lim_{x \to +\infty} F_X(x) = 1$, $\lim_{x \to -\infty} F_X(x) = 0$.

Док.Свойството 1. следва непосредствено от дефиницията на ф.р. За свойство 2., от $x_1 \le x_2$ следва, че е в сила следната връзка между събитията $\{\omega : X(\omega) \le x_1\} \subseteqq \{\omega : X(\omega) \le x_2\}$, откъдето веднага следва 2.

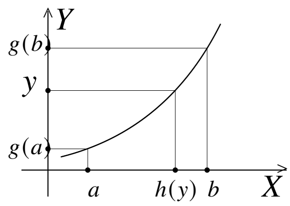
Нека $x_n = x_0 - 1/n$, т.е. $x_n < x_0$ и $x_n \uparrow x_0$, $n \to \infty$. Тогава отново редицата от събития $\{\omega : X(\omega) \le x_n\} \uparrow \{\omega : X(\omega) \le x_0\}$.

В сила е и обратното твърдение, а именно, че ако една функция $0 \le F_X(x) = \mathbf{P}(X < x)$ удовлетворява свойствата 1.-4. на твърдение 2., то тя е ф.р. на някоя сл. в. X.

3. Смяна на променливите

Нека X е произволна непрекъсната случайна величина с известно разпределение, т.е. познаваме плътността и $f_X(x)$. Нека Y е нова случайна величина зададена като функция на X, т.е. $Y = g(X)$. Ще покажем как може да се намери плътността на новата случайна величина $f_Y(y)$.

Най-напред ще разгледаме случая, в който функцията $g(x)$ е непрекъсната и монотонно растяща (виж Фигура 1). В този случай съществува обратна функция, която ще означим с $h(y) \equiv g^{-1}(y)$.



Фигура 1.

Вероятността случайната величина X да принадлежи на интервала $[a, b]$ е равна на вероятността Y да принадлежи на $[g(a), g(b)]$. Това ни дава възможност да пресметнем функцията на разпределение на Y

$$F_Y(y) = \mathbf{P}(Y < y) = \mathbf{P}(g(X) < y) = \mathbf{P}(X < h(y)) = F_X(h(y)).$$

Пресмятането на плътността $f_Y(y)$ съгласно (V .6) се свежда до намиране на производна на $F_X(h(y))$. От фомулата за диференциране на сложна функция следва

$$f_Y(y) = \frac{\partial F_X(h(y))}{\partial y} = f_X(h(y))\, h'(y). \qquad \text{(V .7)}$$

Нека сега функцията $g(x)$ е непрекъсната и монотонно намаляваща. Обратната функция $h(y)$ отново съществува.

Вероятността случайната величина X да принадлежи на интервала $[a, b]$ е равна на вероятността Y да принадлежи на $[g(b), g(a)]$. За функцията на разпределение на Y получаваме

$$F_Y(y) = \mathbf{P}(Y < y) = \mathbf{P}(g(X) < y) = \mathbf{P}(X > h(y)) =$$

$$= 1 - \mathbf{P}(X \leq h(y)) = 1 - F_X(h(y)).$$

В последното равенство използвахме факта, че X е непрекъсната и следователно вероятността да попадне във фиксирана точка е нула, т.е. $\mathbf{P}(X \leq h(y)) = \mathbf{P}(X < h(y))$ Отново ще приложим (V .6) за да определим плътността на Y

$$f_Y(y) = \frac{\partial[1 - F_X(h(y))]}{\partial y} = -f_X(h(y))\, h'(y). \qquad \text{(V .8)}$$

Тъй като $g(x)$ е намаляваща функция, то и обратната и функция $h(y)$ също е намаляваща, а производната и $h'(y)$ е отрицателна. Следователно намерената плътност е добре дефинирана $f_Y(y) \geq 0$.

Ще обобщим тези два случая на смяна на променливите в следващото твърдение.

Твърдение

Нека X е непрекъсната случайна величина, а $g(x)$ е монотонна и непрекъсната функция. Тогава плътността на случайната величина $Y = g(X)$ се пресмята по формулата

$$f_Y(y) = f_X(h(y))\, |h'(y)|, \qquad \text{(V .9)}$$

където $h(y)$ е обратната функция на $g(x)$.

Док.Доказателството следва непосредствено от (V .7) и (V .8).

4. Математическо очакване и дисперсия

Формално понятието математическо очакване също се дефинира чрез граничен преход. Ние няма да се спираме на този аксиоматичен подход. Ще използваме следният директен начин.

Дефиниция - Математическо очакване

Математическо очакване на непрекъснатата сл.в. X наричаме числото $\mathbf{E}\,X$ дефинирано с равенството:

$$\mathbf{E}\,X = \int_{-\infty}^{\infty} x\, f_X(x) dx. \qquad \text{(V .10)}$$

Ако интегралът е разходящ, казваме че математическото очакване не съществува.

Тази дефиниция на математическото очакване е аналогична на дефиницията (II .1) от дискретния случай. Поради неизброимия брой стойности на случайната величина тук сумирането е заменено с интегриране, а конкретните вероятности са заменени с плътността на случайната величина.

По-общо математическо очакване на функция от случайна величина може да се пресметне по формулата зададена в следващото твърдение.

Твърдение

Нека X е непрекъсната случайна величина, а $g(x)$ е произволна функция. Тогава, ако съществува математическото очакване $\mathbf{E}g(X)$, то се пресмята по формулата

$$\mathbf{E}\, g(X) = \int_{-\infty}^{\infty} g(x)\, f_X(x) dx.$$

Док.Ще използваме формулата за смяна на променливите. Нека $Y = g(X)$. Съгласно предходната дефиниция

Доказателството е само за непрекъснатата и монотонна $g(x)$. Твърдението е изпълнено и за произволна функция, но в този случай се изисква аксиоматично построяване на математическото очакване, което е извън рамките на учебни.

$$\mathbf{E}\, g(x) = \mathbf{E}\, Y = \int_{-\infty}^{\infty} y\, f_Y(y) dy = \int_{-\infty}^{\infty} y\, f_X(h(y))\, |h'(y)| dy.$$

В последното равенство приложихме (V .9). Сега ще направим и смяна на променливите в интеграла, като ще ползваме същата смяна $y = g(x)$, $x = h(y)$

$$\int_{-\infty}^{\infty} y\, f_X(h(y))\, |h'(y)| dy = \int_{-\infty}^{\infty} y\, f_X(h(y)) dh(y) = \int_{-\infty}^{\infty} g(x)\, f_X(x) dx.$$

✓

Аналогично на дисперсията на дискретна случайна величина се дефинира и дисперсията в непрекъснатия случай.

Дисперсия на непрекъсната сл.в.

Дисперсия на непрекъснатата сл.в. X наричаме числото $\mathbf{D}\, X$ дефинирано с равенството:

$$\mathbf{D}\, X = \int_{-\infty}^{\infty} (x - \mathbf{E}X)^2\, f_X(x) dx. \qquad \text{(V .11)}$$

Ако интегралът е разходящ, казваме че дисперсията е безкрайна.

Ще отбележим, че всички свойства на математическото очакване и диспресията от дискретния случай се запазват и няма да ги доказваме отделно.

VI . Двумерни непрекъснати случайни величини

Нека (X_1, X_2) е случаен вектор, чиито компоненти са непрекъснати сл.в. Тогава съвместна функция на разпределение или съвместно разпределение на сл.вектор (X_1, X_2) наричаме функция $F_{X_1,X_2}(x_1, x_2)$ такава, че:

$$F_{X_1,X_2}(x_1, x_2) = \mathbf{P}(X_1 < x_1; X_2 < x_2)$$

за всички $x_1, x_2 \in \mathbb{R}$.

Дефиниция - Съвместна плътност

Нека (X_1, X_2) е случаен вектор, чиито компоненти са непрекъснати сл.в. Тогава вероятностна плътност на съвместното разпределение на сл.вектор (X_1, X_2) (накратко съвместна плътност) наричаме функция $f_{X_1,X_2}(x_1, x_2)$ такава, че:

$$1. f_{X_1,X_2}(x_1, x_2) \geq 0, \quad \text{за всички} \quad x_1, x_2 \in \mathbb{R};$$

$$2. \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)dx_1dx_2 = 1;$$

$$3. \mathbf{P}(a \leq X_1 \leq b; c \leq X_2 \leq d) = \int_a^b\int_c^d f_{X_1,X_2}(x_1, x_2)dx_1dx_2.$$

Подобно на дискретния случай и тук можем да дефинираме разпределението на сл. в. X_1, X_2 поотделно, наричани също маргинални разпределения.

Дефиниция - Маргинална плътност

Маргинална плътност на сл.в. X_1 наричаме функцията:

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f_{X_1,X_2}(x_1, x_2)dx_2,$$

Аналогично се дефинира и маргиналната плътност на X_2.

Важно е да се отбележи, че в термините на съвместната ф.р. можем да дадем аналогична дефиниция на независимост на две сл.в. в непрекъснатия случай по следния начин:

Дефиниция - Независимост

Казваме, че сл.в. X_1 и X_2 са независими, ако

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2),$$

за всички $x_1, x_2 \in \mathbb{R}$;

Подобно на независимост при случайни събития, понятието независимост може да бъде въведено и за повече от две сл.в. Отново имаме два вида независимост: "две по две" и в съвкупност. Ще дадем съответните дефиниции като предварително ще отбележим, че това е в сила и в дискретния случай в съответните означения.

Дефиниция - Независимост "две по две"

Казваме, че сл.в. $X_1, \ldots, X_n$ са независими "две по две", ако

$$F_{X_i,X_j}(x_i, x_j) = F_{X_i}(x_i)F_{X_j}(x_j),$$

за всички $x_i, x_j \in \mathbb{R}, i, j = 1, \ldots, n.$

Дефиниция - Независимост в съвкупност

Казваме, че сл.в. $X_1, \ldots, X_n$ са независими в съвкупност или накратко само независими, ако за всяко $k : \quad 2 \le k \le n$

$$F_{X_i,\ldots,X_n}(x_i, \ldots, x_k) = \prod_{i=1}^{k} F_{X_i}(x_i),$$

за всички $x_i \in \mathbb{R}, i = 1, \ldots, n.$

Съгласно дефиниция (VI) съвместната функция на разпределение може да се изрази чрез съвместната плътност по следния начин

$$F_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1,X_2}(u, v)dudv.$$

Следователно обратната връзка е от вида

$$f_{X_1,X_2}(x_1, x_2) = \frac{\partial^2 F_{X_1,X_2}(x_1, x_2)}{\partial x_1 \partial x_2}.$$

Аналогична връзка между съвместната ф.р. и плътност имаме и за повече от две сл.в. Да отбележим, че в термините на съвместната плътност можем да дадем аналогична дефиниция на независимост на две сл.в. в непрекъснатия случай.

Както при събития и тук от независимост в съвкупност следва независимост "две по две", но обратното не е вярно.

VII . Непрекъснати разпределения

1. Равномерно разпределение - $X \in U(a,b)$

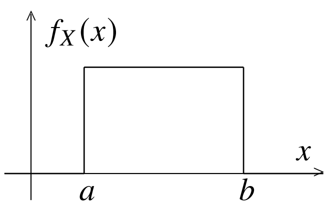
Нека $[a,b]$ е произволен интервал върху реалната права. Казваме, че случайната величина X е равномерно разпределена в $[a,b]$, ако вероятността да вземе коя да е стойност в този интервал е една и съща. Или казано по друг начин, ако X попада по случаен начин в този интервал. Равномерното разпределение се означава съкратено $X \in U(a,b)$, където $a < b$ са реални числа.

Плътността на равномерно разпределената случайна величина X е константа в $[a,b]$, т.е.

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & , x \in [a,b], \\ 0 & , x \notin [a,b] \end{cases}$$

и е представена на Фигура 2.

Константата е определена от нормиращото условие (V .2), т.е. лицето под функцията вероятностна плътност да бъде единица.



Фигура 2. Вероятностна плътност на сл.в. $X \in U(a,b)$.

Ще намерим математическото очакване на X:

$$\mathbf{E}\,X = \int_a^b x\,\frac{1}{b-a}dx = \frac{1}{b-a}\,\frac{x^2}{2}\bigg|_{x=a}^{b} = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

Както можеше да се предположи, математическото очакване е точно в средата на интервала $[a,b]$.

Ще пресметнем и дисперсията на X:

$$\mathbf{E}\,X^2 = \int_a^b x^2\,\frac{1}{b-a}dx = \frac{1}{b-a}\,\frac{x^3}{3}\bigg|_{x=a}^{b} = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}.$$

Тогава за дисперсията по съкратената формула, получаваме

$$\mathbf{D}\,X = \mathbf{E}\,X^2 - (\mathbf{E}\,X)^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

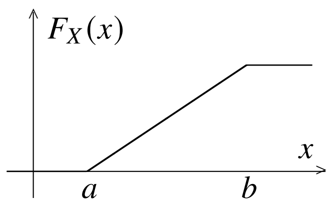
Ще използваме (V .5) за да определим функцията на разпределение (ф.р.) на X. Ясно е, че $F_X(x) = 0$ за $x < a$ и $F_X(x) = 1$ за $x \geq b$. Ще пресметнем съществения случай $a \leq x < b$

$$F_X(x) = \int_a^x \frac{1}{b-a}\,dt = \frac{t}{b-a}\bigg|_{t=a}^{x} = \frac{x-a}{b-a}.$$

Окончателно за ф.р. на $X \in U(a,b)$ получаваме

$$F_X(x) = \begin{cases} 0 & ,x < a, \\ \dfrac{x-a}{b-a} & ,a \leq x < b, \\ 1 & ,x \geq b, \end{cases}$$

чиято графика е представена на Фигура 1..



Фигура 3. Функция на разпределение на сл.в. $X \in U(a,b)$.

2. Нормално разпределение - $X \in N(\mu, \sigma^2)$

Нормално разпределените случайни величини са изключително често срещани, изучаването на техните свойства е важна задача в теория на вероятностите. Нормалната плътност е открита от Гаус при изучаване на грешките в астрономически наблюдения. Поради тази причина разпределението също се нарича Гаусово. Нормалната плътност има вида:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \qquad x \in (-\infty, \infty)$$

Където $\sigma > 0$, а μ е произволно реално число.

Известно е, че функцията e^{-x^2} няма примитивна, т.е. не може да бъде интегрирана в общия случай. Затова функцията на разпределение $F_X(x)$ няма явен вид, а също така вероятностите, свързани с нормално разпределена сл.в. не могат да бъдат пресмятани по традиционния начин чрез интегриране, а са пресметнати числено и се вадят от таблици.

Ще пресметнем математическото очакване на X.

$$\mathbf{E}\,X = \int_{-\infty}^{\infty} \frac{x}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{x-\mu+\mu}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dx$$

Ще разделим този интеграл на две части. Първата с $x-\mu$, а втората с μ. В първият интеграл ще извършим линейна смяна на променливите $y = x-\mu$. При тази смяна границите на интеграла се запазват, а $dy = dx$

$$\int_{-\infty}^{\infty} \frac{x-\mu}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{y}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{y^2}{2\sigma^2}} dy = 0.$$

Този интеграл е равен на нула, тъй като функцията е нечетна, а границите на интегриране са симетрични.

Ще решим втория интеграл, като го сведем до интеграл от нормална плътност. Ще използваме факта, че плътността отговаря на нормиращото условие 2), т.е. за всяко реално μ и $\sigma > 0$ е изпълнено

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dx = 1.$$

Това е известния от анализа Гаусов интеграл (няма да го решаваме тук). Един начин за намиране е да се вземе интегралът на квадрат, да се представи като двоен и да се извърши полярна смяна на променливите.

Като използваме Гаусовия интеграл за математическото очакване получаваме:

$$\mathbf{E}\,X = \int_{-\infty}^{\infty} \frac{\mu}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dx = \mu \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dx = \mu.$$

Ще намерим дисперсията на X. Няма да използваме традиционния начин с намирането на $\mathbf{E}\,X^2$, защото това е трудоемка задача. Вместо това ще пресметнем дисперсията директно от нейната дефиниция. Като приложим формулата за очакване на функция от сл.в. (стр.??), получаваме

$$\mathbf{D}\,X = \mathbf{E}\,(X-\mathbf{E}\,X)^2 = \mathbf{E}\,(X-\mu)^2 = \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dx$$

Интегрираме по части, като ще вкараме експонентата под знака на диференциала. Предварително ще пресметнем производната на експонентата.

$$\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)' = -\frac{x-\mu}{\sigma^2}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Следователно

$$\mathbf{D}\,X = \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = -\sigma^2 \int_{-\infty}^{\infty} \frac{(x-\mu)}{\sigma\sqrt{2\pi}} d\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) =$$

$$= -\sigma^2\, \frac{x-\mu}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\Bigg|_{-\infty}^{\infty} + \sigma^2 \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \sigma^2$$

Първото събираемо е равно на нула, тъй като $\lim\limits_{x\to\pm\infty} xe^{-x^2} = 0$. Във второто събираемо интегралът отново е Гаусов.

Така за нормално разпределената случайна величина $X \in N(\mu, \sigma^2)$ получихме

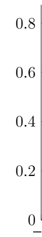
$$\mathbf{E}\,X = \mu, \qquad \mathbf{D}\,X = \sigma^2,$$

т.е. параметрите μ и σ^2, с които се задава случайната величина, са математическото очакване и дисперсията.

Графиката на нормалната плътност

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

е известната камбановидната крива, тя е симетрична, средата и е точно в математическото очакване μ.



Дисперсията, както може да се очаква дава разсейването на случайната величина. При по-малка дисперсия кривата е по стръмна и стойностите на сл.в. са скупчени около очакването (синята крива).

При по-голяма дисперсия, кривата е по плавна и има по-голяма вероятност стойностите на сл.в. да са далеч от очакването (червената крива).

Както отбелязахме, вероятностите свързани с нормално разпределени сл.в. се вадят от таблици. Таблиците се отнасят само за сл.в. от типа $N(0,1)$. Прието е, това да се нарича „стандартно нормално разпределение“, а процедурата, с която една случайна величина се привежда към този вид - „стандартизиране“.

Твърдение - Стандартизиране

Нека $X \in N(\mu, \sigma^2)$ и нека

$$Z = \frac{X - \mu}{\sigma},$$

тогава $Z \in N(0, 1)$.

Док. Ще използваме формулата за смяна на променливите(стр.??). В условието е зададена правата трансформация $z = \frac{x-\mu}{\sigma}$.

Обратната трансформация е $x = h(z) = \sigma z + \mu$ с производна $h'(z) = \sigma$. По този начин за плътността на Y, получаваме

$$f_Z(z) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(\sigma z + \mu - \mu)^2}{2\sigma^2}}\, \sigma = \frac{1}{\sqrt{2\pi}}\, e^{-\dfrac{z^2}{2}}$$

Тази функция явно е плътност от типа $N(0, 1)$. ✓

Прието е функцията на разпределение F_X на стандартното нормално разпределение да се бележи с $\mathbf{\Phi}$, т.е. ако $Z \in N(0, 1)$

$$\mathbf{\Phi}(q) = \mathbf{P}(Z < q)$$

Стойностите на функцията $\mathbf{\Phi}(q)$ са пресметнати числено и са дадени в таблицата на следващата страница. Първите две цифри на числото q са означени в началото на реда, а третата цифра в съответната колона. Така например $\mathbf{\Phi}(0.83) = 0.7967$.

Вероятностите при отрицателните стойности на q не са в таблицата, тъй като лесно могат да бъдат сметнати, като се използва симетричността на нормалната плътност.

Свойство на функцията на разпределение на $Z \in N(0, 1)$

$$\mathbf{\Phi}(-q) = 1 - \mathbf{\Phi}(q)$$

Док. $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\dfrac{z^2}{2}}$ е четна функция, т.е. $f_Z(-z) = f_Z(z)$. Тогава

$$\mathbf{\Phi}(-q) = \int_{-\infty}^{-q} f_Z(z)dz = \int_{q}^{\infty} f_Z(z)dz = \int_{-\infty}^{\infty} f_Z(z)dz - \int_{-\infty}^{q} f_Z(z)dz = 1 - \mathbf{\Phi}(q)$$

Пример

Резултатите от теста SAT проведен в 1982г. са нормално разпределени с очакване $\mu = 503$ и дисперсия $\sigma^2 = 9604$. Каква част от резултатите попадат в граници съответно $\pm$ едно стандартно отклонение от средното.

Знаем че $X \in N(503, 9604)$ и стандартното отклонение е $\sigma = \sqrt{9604} = 98$.

Търсим вероятността $p = \mathbf{P}(\mu - \sigma < X < \mu + \sigma) = \mathbf{P}(405 < X < 601)$.

За да можем да използваме таблицата ще трябва да стандартизираме X (стр.2.) , т.е. да направим трансформацията $Z = \frac{X-503}{98}$, тогава $Z \in N(0,1)$.

$$p = \mathbf{P}\left(\frac{405-503}{98} < \frac{X-503}{98} < \frac{601-503}{98}\right) = \mathbf{P}(-1 < Z < 1) =$$

$$= \mathbf{P}(Z < 1) - \mathbf{P}(Z < -1) = \Phi(1) - \Phi(-1)$$

От таблицата отчитаме: $\Phi(1) = 0.8413$,
$\Phi(-1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$.

Следователно $p = 0.8413 - 0.1587 = 0.6826$, т.е. приблизително 68% от резултатите са в интервала$(\mu - \sigma, \mu + \sigma)$.

Аналогично:
95% от резултатите са в интервала$(\mu - 2\sigma, \mu + 2\sigma)$.
99.7% от резултатите са в интервала$(\mu - 3\sigma, \mu + 3\sigma)$.

3. Експоненциално разпределение - $X \in Ex(\lambda)$

4. Гама разпределение - $X \in \Gamma(\alpha, \beta)$

5. Хи квадрат разпределение - $X \in \chi^2(n)$

6. Разпределение на Стюдънт - $X \in t(n)$

VIII . Гранични резултати

В тази глава ще разгледаме два фундаментални резултата в теория на вероятностите, а именно:

- Закон за големите числа и
- Централна гранична теорема

1. Сходимост на случайни величини

Нека $X_1, X_2, \dots X_n, \dots$ е редица от случайни величини. Интересен е въпросът, дали тази редица е сходяща и в какъв смисъл. Съществуват няколко вида сходимост на сл.в. - почти сигурно, по вероятност, по разпределение, слаба и т.н. Ние ще разгледаме два типа.

Дефиниция - Сходимост по вероятност $X_n \xrightarrow{P} X$

Казваме, че редицата е $X_1, X_2, \dots X_n, \dots$ е сходяща по вероятност към сл.в. X, ако за всяко $\varepsilon > 0$

$$\mathbf{P}\left(|X_n - X| > \varepsilon\right) \to 0 \quad \text{при} \quad n \to \infty$$

Съгласно тази дефиниция, редицата $\{X_n\}$ е сходяща, ако вероятността случайните величини да попаднат извън ε-околност на X клони към нула.

Еквивалентен начин да дефинираме сходимост по вероятност е да поискаме $\mathbf{P}\left(|X_n - X| < \varepsilon\right) \to 1$. т.е вероятността сл.в. да попаднат в ε-околността да клони към едно.

Сходимостта по вероятност е по "слаба" от традиционната сходимост позната от анализа, при която бихме поискали всички сл.в., след някое място в редицата, да са в тази околност. Ще дадем един важен контрапример, който показва разликата между двете сходимости.

Нека $\mathbf{\Omega} = [0, 1]$, събитията са интервалите, и вероятността на даден интервал е дължината му. Ще въведем следната редица от събития:

за $n = 1 : A_{1,1} = [0, 1]$ $\mathbf{P}(A_{1,1}) = 1$
за $n = 2 : A_{2,1} = [0, 1/2]$ $A_{2,2} = [1/2, 1]$ $\mathbf{P}(A_{2,k}) = 1/2$
за $n = 3 : A_{3,1} = [0, 1/3]$ $A_{3,2} = [1/3, 2/3]$ $A_{3,3} = [2/3, 1]$ $\mathbf{P}(A_{3,k}) = 1/3$
изобщо $A_{n,k} = [(k-1)/n, k/n]$ $\mathbf{P}(A_{n,k}) = 1/n$

Дефинираме случайни величини, които са индикатори на тези събития

$$X_{n,k}(\omega) = \begin{cases} 1, & \omega \in A_{n,k} \\ 0, & \omega \notin A_{n,k} \end{cases}$$

Редицата $X_{1,1}, X_{2,1}, X_{2,2}, X_{3,1}, \ldots$ ще клони по вероятност към нула, защото

$$\mathbf{P}\left(|X_{n,k} - 0| > \varepsilon\right) = \mathbf{P}\left(X_{n,k} > \varepsilon\right) = \mathbf{P}(X_{n,k} = 1) = 1/n \to 0$$

От друга страна, което и ω да фиксираме, за $\forall n$ ще има множество, което съдържа ω и такива, който не, т.е. индикатори $X_{n,i}(\omega) = 0$ и $X_{n,j}(\omega) = 1$. Следователно редицата ще има две точки на сгъстяване 0 и 1 и няма как да е сходяща в смисъла на анализа.

Дефиниция - Сходимост по разпределение $X_n \xrightarrow{d} X$

Казваме, че редицата е $X_1, X_2, \ldots X_n, \ldots$ е сходяща по разпределение към сл.в. X, ако съответната редица от функции на разпределение е сходяща, за всяка точка на непрекъснатост x, т.е.

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

Сходимостта по разпределение, носи по-скоро някаква статистическа информация. Казва как е разпределена границата на редицата, но не казва нищо за конкретната реализация. Например, при известни условия сумата на произволни случайни величини се оказва нормално разпределена и този факт ни дава възможност да направим изводи за поведението на сумата, дори без да познаваме отделните събираеми.

Следващото твърдение дава връзката между въведените сходимости. Твърдение От сходимост по вероятност следва сходимост по разпределение, т.е.

$$X_n \xrightarrow{p} X \quad \Rightarrow \quad X_n \xrightarrow{d} X$$

Док. По условие за всяко $\varepsilon > 0$ при $n \to \infty$ е изпълнено

$$\mathbf{P}\left(|X_n - X| > \varepsilon\right) \to 0$$

но

$$\mathbf{P}\left(|X_n - X| > \varepsilon\right) = \mathbf{P}\left(X_n - X > \varepsilon\right) + \mathbf{P}\left(X - X_n > \varepsilon\right)$$

Следователно

$$\mathbf{P}\left(X_n - X > \varepsilon\right) \to 0 \qquad \text{и} \qquad \mathbf{P}\left(X - X_n > \varepsilon\right) \to 0 \qquad \text{(VIII .1)}$$

Ще покажем, че от тези равенства следва

$$F_{X_n}(x) = \mathbf{P}(X_n < x) \ \to \ \mathbf{P}(X < x) = F_X(x)$$

Да допуснем, че $X_n < x$. Тогава за X има две възможности:

- или $X < x + \varepsilon$
- или $X \geq x + \varepsilon$ и в този случай $X - X_n > \varepsilon$

От допускането следва едно от двете описани събития, т.е.

$$\{X_n < x\} \ \subset \ \{X < x + \varepsilon\} \ \cup \ \{X - X_n > \varepsilon\}$$

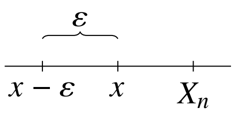
Което означава

$$\mathbf{P}(X_n < x) \ \leq \ \mathbf{P}(X < x + \varepsilon) \ + \ \mathbf{P}(X - X_n > \varepsilon)$$

По този начин получихме оценка отгоре за ф.р. $F_{X_n}(x) = \mathbf{P}(X_n < x)$.

Аналогично ще получим оценка отдолу. Да допуснем, че $X_n \geq x$. Тогава:

- или $X \geq x - \varepsilon$
- или $X < x - \varepsilon$ и в този случай $X_n - X > \varepsilon$



$$\{X_n \geq x\} \ \subset \ \{X \geq x - \varepsilon\} \ \cup \ \{X_n - X > \varepsilon\}$$

Ще изразим вероятностите на противоположните събития, които всъщност ни интересуват. Така получаваме оценка отодолу за функцията на разпределение.

$$1 - \mathbf{P}(X_n < x) \ \leq \ 1 - \mathbf{P}(X < x - \varepsilon) \ + \ \mathbf{P}(X_n - X > \varepsilon)$$

$$\mathbf{P}(X < x - \varepsilon) \ - \ \mathbf{P}(X_n - X > \varepsilon) \leq \ \mathbf{P}(X_n < x)$$

Нека да комбинираме двете оценки в една.

$$F_X(x - \varepsilon) - \ \mathbf{P}(X_n - X > \varepsilon) \ \leq \ F_{X_n}(x) \ \leq \ F_X(x + \varepsilon) + \ \mathbf{P}(X - X_n > \varepsilon)$$

По този начин от ???, при $n \to \infty$ получаваме.

$$F_X(x - \varepsilon) \ \leq \ F_{X_n}(x) \ \leq \ F_X(x + \varepsilon)$$

Ако x е точка на непрекъснатост на $F_X(x)$, от "теоремата за милиционерите" следва търсената сходимост.

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

Това твърдение показва, че ако сл.в. $X_1, X_2, \ldots X_n, \ldots$ се уеднаквяват стохастически, т.е. започват да взимат близки стойности то и разпределението им се уеднаквява. Обратното твърдение не е вярно, възможно е случайните виличини в редицата да са с подобно или даже еднакво разпределение и въпреки това да се държат съвсем различно. Както показва следващия контрапример.

Пример

Нека $X_1, X_2, \dots X_n, \dots$ са независими и еднакво разпределени $N(0,1)$. След като всички сл.в. имат едно и също разпределение то очевидно можем да приемем, че те клонят към него $X_n \xrightarrow{d} X \in N(0,1)$.

От друга страна, ако $X_n > \varepsilon/2$ и $X < \varepsilon/2$ ще следва $X_n - X > \varepsilon$, тогава

$$\left\{X_n > \frac{\varepsilon}{2}, X < -\frac{\varepsilon}{2}\right\} \subset \{X_n - X > \varepsilon\}$$

Следователно

$$\mathbf{P}(X_n - X > \varepsilon) \geq \mathbf{P}\left(X_n > \frac{\varepsilon}{2}\right) \mathbf{P}\left(X < -\frac{\varepsilon}{2}\right) = \left[1 - \Phi\left(\frac{\varepsilon}{2}\right)\right] \Phi\left(-\frac{\varepsilon}{2}\right)$$

$\Phi()$ е ф.р. на $N(0,1)$, за която знаем, че $\Phi(0) = 1/2$. Това означава, че няма как при $n \to \infty$ тази вероятност да клони към нула, защото за $\varepsilon \to 0$ получаваме

$$\mathbf{P}(X_n - X > \varepsilon) \geq \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Редицата не е сходяща по вероятност.

Все пак в някой частни случай, от сходимост по разпределение следва сходимост по вероятност.

Твърдение Ако $X_n \xrightarrow{d} C = const$, то $X_n \xrightarrow{P} C$.

Док. За начало ще намерим функцията на разпределение на константа. На константите гледаме, като на случайни величини, които с вероятност едно приемат единствена стойност. Тогава за всяко x

$$F_C(x) = \mathbf{P}(C < x) = \begin{cases} 0, & x \leq C \\ 1, & x > C \end{cases}$$

Нека редицата е сходяща по разпределение, т.е. $F_{X_n}(x) \to F_C(x)$. Ще покажем , че редицата е сходяща и по вероятност.

$$\mathbf{P}(|X_n - C| > \varepsilon) = \mathbf{P}(X_n > C + \varepsilon) + \mathbf{P}(X_n < C - \varepsilon) \leq$$

$$\leq 1 - \mathbf{P}(X_n < C + \varepsilon) + \mathbf{P}(X_n < C - \varepsilon) =$$

$$= 1 - F_{X_n}(C + \varepsilon) + F_{X_n}(C - \varepsilon) \to$$

$$\to 1 - F_C(C + \varepsilon) + F_C(C - \varepsilon) = 1 - 1 + 0 = 0$$

2. Неравенства за случайни величини

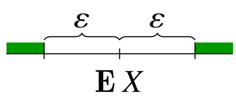
Под името неравенство на Чебишов в литературата са познати няколко неравенства. Ние ще приведем, това което в последствие ще използваме за доказване на закон за големите числа.

Дефиниция - Неравенство на Чебишов

Нека X е произволна случайна величина и $\mathbf{E}\,X$ съществува, тогава за всяко ε е изпълнено:

$$\mathbf{P}(|X - \mathbf{E}\,X| \geq \varepsilon) \leq \frac{\mathbf{D}\,X}{\varepsilon^2}$$

Това неравенство добре описва и смисъла на понятието дисперсия. Вероятността сл.в. да е извън ε-околност на математическото очакване се оценява с дисперсията. Обърнете внимание, че няма изискване ε да е пренебрежимо малко, то може да е произволно число.



Док. Неравенството е изпълнено за произволни сл.в. Ние ще го докажем поотделно за дискретни и непрекъснати сл.в.

Нека X е дискретна сл.в. За удобство ще означим $p_i = \mathbf{P}(X = x_i)$. Знаем, че $\mathbf{E}\,X$ е число и тогава $|X - \mathbf{E}\,X|$ също е дискретна сл.в. Нека S е множеството от индекси на онези стойности x_i на X, за който $|x_i - \mathbf{E}\,X| \geq \varepsilon$, т.е.

$$S = \{i : |x_i - \mathbf{E}\,X| \geq \varepsilon\}$$

В частност множеството S може и да е празно. Вероятността, която се опитваме да оценим се получава чрез сумиране върху S.

$$\mathbf{P}(|X - \mathbf{E}\,X| \geq \varepsilon) = \sum_S p_i$$

За елементите от S е изпълнено $1 \leq \frac{|x_i - \mathbf{E}\,X|}{\varepsilon}$ и следователно $1 \leq \frac{(x_i - \mathbf{E}\,X)^2}{\varepsilon^2}$. Ако в горната сума умножим събираемите с този множител, то сумата само може да нарасне. Ако разширим границите на сумиране то сумата също ще нарасне.

$$\mathbf{P}(|X - \mathbf{E}\,X| \geq \varepsilon) \leq \sum_S \frac{(x_i - \mathbf{E}\,X)^2}{\varepsilon^2}\, p_i \leq \sum_{i=1}^{\infty} \frac{(x_i - \mathbf{E}\,X)^2}{\varepsilon^2}\, p_i$$

За да завършим доказателството е достатъчно да използваме формулата за очакване на функция от сл.в.

$$\sum_{i=1}^{\infty} (x_i - \mathbf{E}\,X)^2\, p_i = \mathbf{E}\,(X - \mathbf{E}\,X)^2 = \mathbf{D}\,X$$

Нека сега X е непрекъсната сл.в. с плътност $f_X(x)$. Следваме сходен подход - представяме вероятността като интеграл върху подходящо множество.

$$\mathbf{P}(|X - \mathbf{E}\,X| \geq \varepsilon) = \int_{|X-\mathbf{E}\,X|\geq\varepsilon} 1\,.\, f_X(x)\,dx \leq$$

заменяме константата 1 в интеграла с $\frac{(x_i-\mathbf{E}\,X)^2}{\varepsilon^2} \geq 1$ и разширяваме границите на интегриране

$$\leq \int_{|X-\mathbf{E}\,X|\geq\varepsilon} \frac{(x_i - \mathbf{E}\,X)^2}{\varepsilon^2} f_X(x)dx \leq \frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}(x_i-\mathbf{E}\,X)^2 f_X(x)dx = \frac{\mathbf{D}\,X}{\varepsilon^2} \qquad \square$$

Неравенството на Чебишов дава само горна граница за вероятността. То не гарантира, че тя ще бъде достигната. Ако дисперсията е голяма, е възможно получената горна граница да бъде по-голяма от едно, тогава неравенството просто не дава никаква информация за вероятността. В частност, възможно е дисперсията да не съществува, т.е. да бъде безкрайна и неравенството ще е тривиално.

Неравенството на Чебишов е полезно, тогава когато получената вероятност е малка, респективно за големи стойности на ε.

В лекция IV разгледахме пример с играч на рулетка, който залага 10лв. При това той може да заложи на червено и тогава евентуалната му печалба е 10лв., или да заложи на едно число при печалба 350лв. С X_1 съответно Y_1 ще означим печалбите в двата случая. Пресметнахме $\mathbf{E}\,X_1 = \mathbf{E}\,Y_1 = -0.27$, а също $\mathbf{D}\,X_1 \approx 99.9$ и $\mathbf{D}\,Y_1 \approx 3408$.

Сега ще използваме неравенството на Чебишов за да оценим вероятноста след 100 изиграни игри играчът да е спечелил поне 700лв.

Нека $X = X_1+X_2+\ldots+X_{100}$ е печалбата от 100 игри. Печалбите в отделните игри са независими сл.в. Тогава $\mathbf{E}\,X = \mathbf{E}\,X_1 + \mathbf{E}\,X_2 + \ldots =$ -27лв и съответно $\mathbf{D}\,X = \mathbf{D}\,X_1+\mathbf{D}\,X_2+\ldots = 9990$. Ще изберем $\varepsilon = 727$

$$\mathbf{P}(X \geq 700) \leq \mathbf{P}(|X + 27| \geq 727) \leq \frac{9900}{727^2} \approx 0.0189$$

Аналогично за $Y = Y_1 + Y_2 + \ldots + Y_{100}$

$$\mathbf{P}(Y \geq 700) \leq \mathbf{P}(|Y + 27| \geq 727) \leq \frac{340800}{727^2} \approx 0.64$$

Както можеше да се очаква за сл.в, която е с по-малка дисперсия - в случая X, вероятността да е далеч от очакването си е по-малка, едва 2%.

Все пак това са само оценки отгоре, ако искаме да пресметнем вероятността на събитията, разполагаме и с по-точни методи.

3. Закони за големите числа

Интуитивно е ясно, че ако разполагаме с няколко наблюдениа на сл.в, средното аритметично от тях ще даде по точна информация за случайната величина, отколкото едно единствено измерване. Формалното обяснение на този факт се дава от законите за големи числа.

Закон за големите числа - ЗГЧ

Нека $X_1, X_2, \dots X_n, \dots$ е редица от случайни величини, казваме че тя изпълнява, или че за нея е в сила "закон за големите числа" (ЗГЧ), ако

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbf{E}\,X_i) \xrightarrow{P} 0 \quad \text{при} \quad n \to \infty$$

Ако сл.в. са еднакво разпределени, то математическото им очакване съвпада и ЗГЧ може да бъде записан в следния еквивалентен вид.

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mathbf{E}\,X_1 \quad \text{при} \quad n \to \infty$$

Този запис дава и смисъла на понятието "математическо очакване" - то е усреднения резултат от измерването на сл.в. при извършване на безкраен брой опити.

Разбира се, законът за големите числа не е изпълнен за всяка редица от сл.в. Ще докажем няколко теореми, които задават условия над редицата, за да е в сила ЗГЧ.

Теорема на Марков

Ако за редицата $X_1, X_2, \dots X_n, \dots$ е изпълнено

$$\frac{1}{n^2}\mathbf{D}\left(\sum_{i=1}^{n} X_i\right) \to 0 \quad \text{при} \quad n \to \infty$$

то е в сила ЗГЧ.

Док. Съгласно дефиницията за сходимост по вероятност (стр.??), за да е изпълнен ЗГЧ следната вероятност трябва да клони към нула

$$\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbf{E}\,X_i)\right| > \varepsilon\right) =$$

Ще преобразуваме този израз и ще приложим неравенството на Чебишов (стр.??) към сл.в. $\sum_{i=1}^{n} X_i$

$$= \mathbf{P}\left(\left|\sum_{i=1}^{n} X_i - \mathbf{E}\left(\sum_{i=1}^{n} X_i\right)\right| > n\,\varepsilon\right) \le \frac{\mathbf{D}\,\left(\sum_{i=1}^{n} X_i\right)}{n^2\varepsilon^2} \to 0 \qquad \square$$

4. Функция пораждаща моментите

Съществуват различни доказателства на централна гранична теорема, но при всички тях се използват помощни функции от аналитичния апарат на теория на вероятностите. За нашето доказателство ще ни е нужна функцията пораждаща моментите. В началото на лекцията ще въведем и ще изведем основните свойства на тази функция.

В лекция IV описахме свойствата на пораждащите функции. Да припомним, за сл.в. X която има за стойности цели, положителни числа въвеждаме пораждаща функция с равенството

$$g_X(s) = \sum_k P(X = k)\, s^k$$

Сега, когато познаваме свойствата на "математическо очакване", можем да запишем същата дефиниция по аналогичен начин

$$g_X(s) = \mathbf{E}\ s^X$$

Пораждащите функции са всъщност представяне на случайните величини в пространство с базис $1, s, s^2, s^3 \ldots$, т.е. представянето им с полином.

Пораждащите функции дават много възможности за изследване на свойствата на сл.в., но те са приложими само за случайни величини, който имат целочислени стойности. За по общ клас от случайни величини по аналогичен начин се въвеждат функциите пораждащи моментите.

Дефиниция - Функция пораждаща моментите

Нека X е сл.в. такава, че за всяко t в околност на нулата $\mathbf{E}\, e^{tx}$ съществува, тогава функция пораждаща моментите (ф.п.м.) наричаме

$$M_X(t) = \mathbf{E}\left(e^{t\,X}\right)$$

По този начин функция пораждаща моментите е представяне в базис експонентите. Възможно е разбира се, това математическо очакване да е безкрайно, т.е. ф.п.м. да не съществува.

Има и по-мощна аналитична функция, така наречената характеристична функция която се дефинира с равенството $\varphi_X(t) = \mathbf{E}\, e^{it\,X}$, това е познатата трансформация на Фурие, тя винаги съществува, доколкото $\left|e^{it\,X}\right| = 1$. Предвид известната формула на Ойлер $e^{it} = \sin t + i \cos t$, характеристичните функции представят случайните величини в базис синусоиди. За намиране на характеристичните функции обаче се налага пресмятането на интеграли в комплексната област, което е извън рамките на този учебник.

Начина на пресмятане на функцията пораждаща моментите зависи от типа на случайната величина.

❧ За дискретни сл.в.

$$M_X(t) = \mathbf{E}\left(e^{t\,X}\right) = \sum_k e^{tx_k}\mathbf{P}(X = x_k)$$

❧ За непрекъснати сл.в

$$M_X(t) = \mathbf{E}\left(e^{t\,X}\right) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$$

Относно името на функцията, прието е числовите характеристики на сл.в. да се наричат моменти. Като различаваме няколко вида моменти.

❧ Начален момент (или просто момент) от ред k наричаме $\mathbf{E}\,X^k$
т.е. математическото очакване е първия начален момент на сл.в.

❧ Централен момент от ред k наричаме $\mathbf{E}\,(X - \mathbf{E}\,X)^k$
т.е. дисперсията е втория централен момент на сл.в.

❧ Факториелен момент от ред k наричаме $\mathbf{E}\,\left[X(X-1)\dots(X-k+1)\right]$

Свойствата на функцията пораждаща моментите са в някакъв смисъл аналогични на свойствата на пораждащите функции. В повечето случай ни интересува поведението на $M_X(t)$ или на нейните производни при $t = 0$, затова изискваме тя да е дефинирана в някой околност на нулата, т.е за всяко $t \in (-\varepsilon, \varepsilon)$ и не обсъждаме нейната сходимост в общия случай.

Свойства:

❧ $M_X(0) = 1$

Док. Очевидно $M_X(0) = \mathbf{E}\ e^0 = 1$

❧ $M'_X(0) = EX$

Док. Производната на сума е сума от производните, т.е. знакът за сума и производна могат да разменят своите места. Същото свойство важи и за интегралите.

$$M'_X(0) = \mathbf{E}\ \left(e^{tX}\right)'\Big|_{t=0} = \mathbf{E}\ Xe^{tX}\Big|_{t=0} = \mathbf{E}\,X$$

❧ $M_X^{(k)}(0) = EX^k$

Док. Аналогично на предишното свойство, при всяко диференциране на експонентата по t, пред нея пада по едно X.

$$M_X^{(k)}(0) = \mathbf{E}\ \left(X\,e^{tX}\right)^{(k-1)}\Bigg|_{t=0} = \dots = \mathbf{E}\ X^k e^{tX}\Big|_{t=0} = \mathbf{E}\,X^k$$

❧ Ако $X \perp\!\!\!\perp Y$ то $M_{X+Y}(t) = M_X(t)M_Y(t)$

Док. От независимостта на X и Y следва, че e^X и e^Y също са независими, тогава

$$M_{X+Y}(t) = \mathbf{E}\, e^{(X+Y)t} = \mathbf{E}\, e^{Xt} \mathbf{E}\, e^{Yt} = M_X(t) M_Y(t)$$

Следващото свойство е познато, като теорема на Къртис за непрекъснатост на ф.п.м., тя твърди, че от сходимост на ф.п.м. следва сходимост по разпределение. Разглеждаме редица от сл.в. $\{X_n\}$ и нека съответните ф.п.м. $M_{X_n}(t)$ съществуват и са сходящи към $M_X(t)$ за всяко t от някоя околност на нулата, т.е.

❧ Ако $M_{X_n}(t) \to M_X(t)$, за $\forall |t| < \varepsilon$, тогава $X_n \xrightarrow{d} X$

Това свойство показва, че функцията пораждаща моментите еднозначно определя разпределението на случайната величина, т.е. ако познаваме ф.п.м то познаваме и конкретното разпределение.

Доказателството на теоремата на Къртис е сложно технически и няма да го привеждаме тук, то най-общо казано почива на теоремата за непрекъснатост на характеристични функции на Леви.

Ще изведем функцията пораждаща моментите на някой поважни разпределения, които ще използваме в статистиката.

Твърдение

Ако $X \in N(\mu, \sigma^2)$ то функцията пораждаща моментите има вида

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Док. За начало ще разгледаме стандартно нормално разпределение. Нека $Z \in N(0,1)$, т.е. $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$. Тогава

$$M_X(t) = \int_{-\infty}^{\infty} e^{zt} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2 - 2zt}{2}} dz =$$

Ще допълним показателя на експонентата до точен квадрат.

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2 - 2zt + t^2}{2} + \frac{t^2}{2}} dz = e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}} dz = e^{\frac{t^2}{2}}$$

Последният интеграл е равен на едно, като интеграл от нормална плътност $N(t, 1)$.

Нека сега $X \in N(\mu, \sigma^2)$. Знаем че $Z = \frac{X-\mu}{\sigma} \in N(0,1)$. Не е трудно да се покаже, че е в сила и обратната връзка, т.е ако $Z \in N(0,1)$, то

$$X = \sigma Z + \mu \quad \in \quad N(\mu, \sigma^2).$$

Доказателството на този факт е елементарно, прилага се формулата за смяна на променливите в едномерния случай. Ще използваме последното равенство за да изразим функцията пораждаща моментите на X чрез тази на Z.

$$M_X(t) = M_{\sigma Z + \mu}(t) = \mathbf{E}\, e^{t(\sigma Z + \mu)} = \mathbf{E}\left(e^{\sigma t Z} e^{\mu t}\right) =$$

$e^{\mu t}$ е константа от гледна точка на математическото очакване, което е по сл.в. Z.

$$= e^{\mu t}\mathbf{E}\, e^{(\sigma t)Z} = e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{\frac{\sigma^2 t^2}{2}}$$

В последното равенство използвахме ф.п.м. на $Z \in N(0,1)$ ✓

Съгласно теоремата на Къртис, изпълнено е и обратното твърдение. Ако сл.в. X има функция пораждаща моментите от вида $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$ то X е нормално разпределена, т.е. $X \in N(\mu, \sigma^2)$.

Ще изведем и функцията пораждаща моментите на гама разпределението, която ще използваме по-късно при решаване на някой статистически задачи.

Твърдение

Ако $X \in \Gamma(\alpha, \beta)$ то при $t < \beta$ функцията пораждаща моментите е

$$M_X(t) = \left(\frac{\beta}{\beta - t}\right)^{\alpha}$$

Док. Да припомним $X \in \Gamma(\alpha, \beta)$ означава, че плътността има вида

$$f_X(x) = \begin{cases} \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, & x \geq 0, \\ 0, & x < 0, \end{cases} \qquad \text{където} \qquad \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Тогава

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\beta - t)x} dx =$$

Ще пресметнем интеграла като го сведем до гама функция. За целта ще направим смяна на променливите $y = (\beta - t)x$. Трансформацията е линейна за $t < \beta$ границите на интегриране се запазват, обратната трансформация е $x = \frac{y}{\beta - t}$.

$$= \frac{\beta^\alpha}{\Gamma(\alpha)(\beta - t)^\alpha} \underbrace{\int_0^\infty y^{\alpha-1} e^{-y} dy}_{\Gamma(\alpha)} = \left(\frac{\beta}{\beta - t}\right)^{\alpha}$$

✓

Когато разглеждахме гама разпределение (???), ние не изведохме формулите за очакването и дисперсията му в общия случай. Сега, когато разполагаме с функцията пораждаща моментите, намирането им е елементарна задача, свежда се до просто диференциране.

Нека $X \in \Gamma(\alpha, \beta)$ от свойствата на ф.п.м. получаваме

$$\mathbf{E}\,X = \left(\frac{\beta^\alpha}{(\beta - t)^\alpha}\right)'\Bigg|_{t=0} = \frac{\alpha}{\beta}$$

$$\mathbf{E}\,X^2 = \left(\frac{\beta^\alpha}{(\beta - t)^\alpha}\right)''\Bigg|_{t=0} = \frac{\alpha(\alpha+1)}{\beta^2}$$

$$\mathbf{D}\,X = \mathbf{E}\,X^2 - (\mathbf{E}\,X)^2 = \frac{\alpha}{\beta^2}$$

Можем да намерим функция пораждаща моментите на хи-квадрат разпределението, позовавайки се на връзката му с гама разпределение. Знаем, че $\chi^2(n) \equiv \mathbf{\Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$ тогава за ф.п.м. на $X \in \chi^2(n)$ получаваме

$$M_X(t) = \left(\frac{1/2}{1/2 - t}\right)^{n/2} = (1 - 2t)^{-n/2}, \qquad t < \frac{1}{2}$$

По-късно ще използваме хи-квадрат разпределението при решаване на редица статистически задачи.

5. Централна гранична теорема

Нека $X_1, X_2, \dots X_n, \dots$ е редица от случайни величини. Казваме, че централна гранична теорема (ЦГТ) е в сила, ако

$$\frac{\sum_{k=1}^n X_k - \mathbf{E}\left(\sum_{k=1}^n X_k\right)}{\sqrt{\mathbf{D}\left(\sum_{k=1}^n X_k\right)}} \xrightarrow[n\to\infty]{d} N(0,1) \qquad (\star)$$

Очакването и дисперсията в тази формула са константи, в частен случай могат и да клонят към безкрайност, но те са просто нормиращ множител. Основният извод е, че сумата от какви да е сл.в. клони към нормално разпределение. Разбира се това не е изпълнено за съвсем произволни сл.в. Съществуват няколко теореми, всяка от тях носи името ЦГТ, който налагат условия върху редицата, така че горната сходимост да е в сила.

ЦГТ обяснява и защо нормалното разпределение е толкова често срещано, всяка сл.в. която може да се представи като сума, при някой най-общи условия се оказва нормалното разпределена.

Централна гранична теорема

Нека $X_1, X_2, \dots X_n, \dots$ са независими, еднакво разпределени сл.в. и нека $\mathbf{E}\,X_k = \mu$, $\mathbf{D}\,X_k = \sigma^2 < \infty$, т.е. очакването и дисперсията съществуват, тогава

$$\frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} N(0,1)$$

Док. За начало ще отбележим, че от условието следва

$$\mathbf{E}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathbf{E}\, X_k = n\mu \quad \text{и} \quad \mathbf{D}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathbf{D}\, X_k = n\sigma^2$$

т.е. твърдението на теоремата съвпада с ($\star$).

С цел опростяване на записа ще използваме сл.в. X, която е разпределена точно както зададената редица $\{X_i\}_{i=1,2,\ldots}$. Това е възможно, тъй като $X_1, X_2, \ldots X_n, \ldots$ са еднакво разпределени.

Допускаме, че $M_X(t)$ - функцията пораждаща моментите на X съществува за $|t| < \varepsilon$ (в противен случай, трябва да се направи доказателство с характеристични функции). Ще означим с $m(t)$ ф.п.м. на $X - \mu$, т.е.

$$m(t) = M_{X-\mu}(t) = \mathbf{E}\left(e^{t(X-\mu)}\right) = e^{-t\mu}\, \mathbf{E}\left(e^{tX}\right) = e^{-t\mu}\, M_X(t)$$

и явно $m(t)$ също съществува за $|t| < \varepsilon$. Освен това, съгласно свойства 1), 2) и 3) на ф.п.м. е изпълнено

$$m(0) = 1, \qquad m'(0) = \mathbf{E}\,(X-\mu) = 0, \qquad m''(0) = \mathbf{E}\,(X-\mu)^2 = \sigma^2$$

Ще развием $m(t)$ в ред на Тейлор около точката $t = 0$.

$$m(t) = m(0) + m'(0)t + \frac{m''(0)\, t^2}{2} + O(t^3) = 1 + \frac{\sigma^2 t^2}{2} + O(t^3)$$

Тук $O(t^3)$ е стандартен запис, показващ че събираемото е равно на константа по t^3.

Ще разгледаме сумата $\frac{\sum_{k=1}^{n} X_k - n\mu}{\sigma\sqrt{n}}$,и ще означим с $M(t,n)$ нейната ф.п.м.

$$M(t,n) = \mathbf{E}\left[\exp\left(t\, \frac{\sum_{k=1}^{n} X_k - n\mu}{\sigma\sqrt{n}}\right)\right] =$$

$$= \mathbf{E}\left[\exp\left(t\, \frac{X_1-\mu}{\sigma\sqrt{n}}\right) \exp\left(t\, \frac{X_2-\mu}{\sigma\sqrt{n}}\right) \ldots \exp\left(t\, \frac{X_n-\mu}{\sigma\sqrt{n}}\right)\right] =$$

По условие сл.в. $\{X_k\}$ са независими, и следователно техните експоненти също са независими.

$$= \mathbf{E}\left[\exp\left(t\, \frac{X_1-\mu}{\sigma\sqrt{n}}\right)\right] \mathbf{E}\left[\exp\left(t\, \frac{X_2-\mu}{\sigma\sqrt{n}}\right)\right] \ldots \mathbf{E}\left[\exp\left(t\, \frac{X_n-\mu}{\sigma\sqrt{n}}\right)\right] =$$

Случайните величини са еднакво разпределени, тогава очакванията им съвпадат.

$$= \left\{\mathbf{E}\left[\exp\left(t\, \frac{X-\mu}{\sigma\sqrt{n}}\right)\right]\right\}^n = \left[m\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

Очакването в лявата част на последното равенство разглеждаме като $m()$ - ф.п.м. на $X - \mu$, с аргумент $\frac{t}{\sigma\sqrt{n}}$

Сега използваме развитието на $m(t)$ в ред на Тейлор, което изведохме

$$M(t,n) = \left[m\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \left[1 + \frac{\sigma^2 t^2}{2\sigma^2 n} + O\left(\frac{t^3}{\sigma^3 n^{3/2}}\right)\right]^n =$$

$$= \left[1 + \frac{t^2}{2\,n} + O\left(\frac{t^3}{n^{3/2}}\right)\right]^n$$

Интересува ни границата на този израз при $n \to \infty$. Събираемото $O\left(\frac{t^3}{n^{3/2}}\right)$ клони към нула и то с порядък по-бързо от предишното събираемо, тъй като има в знаменател по-висока степен на n. Следователно основната тежест в асимптотиката се носи от $\dfrac{t^2}{2n}$. Прилагаме известната граница за експонента $\lim\limits_{n\to\infty}\left(1 + \dfrac{c}{n}\right)^n = e^c$. Така получаваме

$$\lim_{n\to\infty} M(t,n) = e^{\frac{t^2}{2}}$$

Това е функцията пораждаща моментите на $N(0,1)$. Според теоремата на Къртис, сходимостта на ф.п.м означава че имаме и сходимост по разпределение към същата случайна величина, т.е.

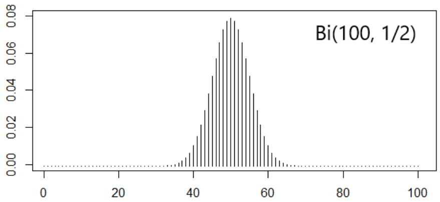
$$\frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} N(0,1) \qquad \checkmark$$

Следствия от централна гранична теорема.

Много често в статистиката се извършват независими наблюдения над една случайна величина. Например, изследват се пациенти приели едно и също лекарство, или се измерват дефектите на детайли произведени от една машина и.т.н. В този случай разполагаме с наблюдения $X_1, \ldots, X_n$, които са независими и еднакво разпределени и явно централна гранична теорема ще бъде валидна, стига наблюденията да са достатъчно на брой. Стандартното означение за средно аритметично е $\overline{X_n} = \frac{\sum X_k}{n}$, чрез него ЦГТ в статистически задачи обикновено обикновено се записва в следния вид

$$\frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{d} N(0,1)$$

Централна гранична теорема твърди, че има сходимост към нормално разпределение, но не казва нищо за скоростта на тази сходимост. Съществуват отделни изследвания по темата - теорема на Бер например. В статистически изследвания често се приема че 30 наблюдения са достатъчни за да се разглежда $\overline{X_n}$ като нормално разпределена сл.в. Този извод обаче е твърде произволен, по-скоро трябва да се изследват конкретните данни.



Първите доказателство на централна гранична теорема са свързани с нейни частни случаи в схема на Бернули. Още в 1733г. Моавър използва нормално разпределение за да опише броя на успехите при хвърляне на монета, този резултат по-късно е доразвит от Лаплас.

Нека разгледаме схема на Бернули, т.е. последователни независими опити с вероятност за успех при всеки от тях p. От една страна броят на успехите X при n опита е биномно разпределена сл.в. $X \in Bi(n, p)$.

От друга страна можем да представим X като сума от отделните опити, т.е. $X = X_1 + \ldots + X_n$, където X_k е резултата на k-тия опит. И тогава, съгласно ЦГТ, X трябва да клони към нормално разпределение. Наистина не е трудно да се види, че условието на ЦГТ е спазено - X_k са независими, еднакво разпределени с очакване $\mathbf{E}\, X_k = p$ и дисперсия $\mathbf{D}\, X_k = pq$. Следователно, би трябвало да очакваме за $n \to \infty$ биномното разпределение да клони към нормално.

На графиката са биномните вероятности $\mathbf{P}(X = k)$, с максимум при $k = 50$. Контурът наистина напомня на нормално разпределение.

За сходимостта на биномно разпределение към нормално се отнася резултатът на Моавър - Лаплас.

Интегрална теорема на Моавър - Лаплас

Нека $X \in Bi(n, p)$ и $\sqrt{npq} \to \infty$ тогава за $a < b$

$$\mathbf{P}\left(a \le \frac{X - np}{\sqrt{npq}} \le b\right) \underset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, dx$$

Това е оригиналната формулировка на теоремата. Вероятността за попадане на сл.в. в интервал $[a, b]$ клони към интеграл от нормална плътност върху същия интервал.

Класическото доказателство на теоремата на Бернули, която формулирахме в края на IX лекция, се извършва с помощта на интегралната теорема на Моавър - Лаплас. Ще предоставим на уважаемия читател възможността да се сети как точно се извършва това.

Ще завършим тази глава с един пример

Собственик на казино отчита , че средно 10 000 пъти на ден се залага жетон от 10лв. на черно или червено на масите в казиното. Той се пита каква е вероятността да спечели по-малко от 2000 лева от този залог.

В лекция IV намерихме разпределението, очакването и дисперсията на печалбата на играча при този залог. Съвсем аналогично се пресмята печалбата на казиното. Нека X е печалбата за собственика от една игра.

X	-10	10
$\mathbf{P}$	$18/37$	$19/37$

$$\mathbf{E}\,X = \frac{10}{37}, \qquad \mathbf{E}\,X^2 = 100, \qquad \mathbf{D}\,X = \frac{136800}{1369}$$

Нека Y е печалбата от 10 000 игри, т.е. Y е сумата от 10 000 сл.в. от типа на X. Тогава съвсем спокойно можем да приемем, че Y е нормално разпределена, $\mathbf{E}\,Y = 10\,000\mathbf{E}\,X \approx 2702$, $\mathbf{D}\,Y = 10\,000\mathbf{D}\,X$ и $\sigma = \sqrt{\mathbf{D}\,Y} \approx 999$

За намиране на търсената вероятност прилагаме ЦГТ.

$$\mathbf{P}(Y < 2000) = \mathbf{P}\left(\frac{Y - 2702}{999} < \frac{2000 - 2702}{999}\right) = \mathbf{P}(Z < -0.70)$$

И от таблицата за нормално разпределение за $Z \in N(0,1)$ отчитаме

$$\mathbf{P}(Z < -0.70) = \Phi(-0.7) = 1 - \Phi(0.7) = 1 - 0.7580 = 0.242$$

Приблизително в една четвърт от дните печалбата на собственика ще е под 2000лв.

Математическа статистика

IX . Точкови оценки

1. Извадка

Задачата, която се решава в статистиката обикновено е следната. Разполагаме с голяма съвкупност от еднородни обекти. Това може да са всички жители на някоя държава, или цялата продукция на някой завод или всички ракови клетки в един организъм и т.н. Всички обекти наричаме “генерална съвкупност”. Изследването на цялата генерална съвкупност най-често е невъзможно или прекалено скъпо. Затова се избират част от обектите, тази част се нарича ”извадка”, провеждат се изследвания върху нея, и по събраната информация се съди за генералната съвкупност. За да бъде проучването смислено е много важен начинът, по който е направена извадката. Един добър метод е извадката да е чисто случайна, т.е. обектите да не бъдат сортирани по никакъв признак. Това не винаги е лесно за съобразяване.

Известно със своя неуспех е проучването направено от американско списание за президентските избори през 1936г. Списанието събрало адреси от телефонните указатели и разпратило въпросници на над 4 000 000 души. Обработило върнатите отговори и обявило победителя на президентските избори, който всъщност ги загубил. Основната грешка на списанието бил начинът, по който е направена извадката. По онова време телефонът бил относително рядка вещ и само по-заможните семейства можели да си го позволят. Така допитването било извършено само сред тях, а вотът на по бедните не бил отчетен.

На същите избори социологът Дж.Галъп правилно предрича резултата използвайки само 4000 анкети. Неговия подход отразява другия начин за подбор на извадката. Той е отчел, че обществото се разпада на социални групи, който се сравнително еднородни в отношението си към кандидатите за президент. Затова в една социална група анкетите може да са малобройни, но е важно извадката да повтаря социалната структура на генералната съвкупност.

Ще дадем още един пример за да поясним същността на проблема. Ако целта ни е да определим средния ръст на жителите на една страна, разбира се измерването на всички е непосилна задача. Ако измерваме само мъжете, естествено ще допуснем грешка, тъй като мъжете са по-високи. Би трябвало да използваме извадка, в която пропорцията на мъжете и жените е такава, каквато е сред цялото население. Но полът не е единствения фактор, който влияе на ръста, друг такъв е възрастта, а също расата и т.н. Отчитането на всички фактори е сложна задача, този проблем от само себе си се решава при използването на чиста случайна извадка.

Ще запишем формално задачата на математическата статистика. Смятаме че има неизвестна сл.в. X която желаем да измерим. Предполагаме, че генералната съвкупност е достатъчно голяма и можем да направим n на брой независими наблюдения над X, който ще означим с $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$. Самите наблюдения сл.в. X_k са копия на X, т.е. имат същото разпределение, очакване и т.н.

За начало ще предполагаме, че разпределението на сл.в. X е известно, но то зависи от неизвестен за нас параметър θ. Например, знаем че X е нормално разпределена, но не знаем с какво очакване и дисперсия. Нека $F_X(x, \theta)$ е функцията на разпеделение на X. Ние търсим точната стойност на θ, като неизвестния параметър θ може да бъде и вектор, т.е. да имаме за оценяване едновременно няколко константи.

Нека $\overrightarrow{\mathrm{X}} = X_1, \ldots, X_n$ са независими наблюдения над X. Построяваме функция $\hat{\theta} = \hat{\theta}(\overrightarrow{\mathrm{X}})$ и нейната стойност приемаме за стойност на неизвестния параметър θ. Наричаме $\hat{\theta}$ точкова оценка (или статистика) за параметъра. Наблюденията $X_1, \ldots, X_n$ са случайни величини и за теоритични цели ги разглеждаме като такива, но ако извършим измерванията, т.е. в конткретен случай те са числа, и стойността на $\hat{\theta}$ е число.

Нашата задача е намирането на точкова оценка. В следващите два парагарафа ще опишем методи за целта.

, но преди това ще въведем следното означение.

Нека $f_{X_k}(x_k, \theta)$ е плътността на k-тото наблюдение. Съвместната плътност на наблюденията наричаме функция на правдоподобие

$$L(\overrightarrow{\mathrm{X}}, \theta) = \prod_{k=1}^{n} f_{X_k}(x_k, \theta)$$

Тя отразява вероятността на наблюденията.

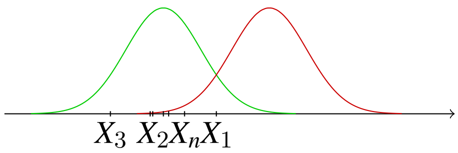
2. Метод на максимално правдоподобие

В началото ще въведем следното означение. Нека $f_{X_k}(x_k, \theta)$ е плътността на k-тото наблюдение. Съвместната плътност на наблюденията наричаме функция на правдоподобие

$$L(\overrightarrow{\mathrm{X}}, \theta) = \prod_{k=1}^{n} f_{X_k}(x_k, \theta)$$

Тя отразява вероятността на наблюденията.

Нека например се опитваме да оценим очакването μ на нормално разпределение, т.е. търсим къде е центрирана кривата. Естествено е да предпочетем μ, при което наблюденията са с голяма вероятност (зелената крива), пред такава стойност, при която наблюденията са малко вероятни (червената крива).



В това се състои идеята на метода на максимално правдоподобие. Неизвестният параметър се избира по такъв начин, че направените наблюдения да се окажат с възможно най-голяма вероятност. В някакъв смисъл напасваме теорията към действителността. Този метод е бил познат още на Гаус, а в началото на 20 век е доразвит от Фишер. Тъй като познаваме функцията на правдоподобие, методът се свежда до намиране на максимума на тази функция по неизвестния параметър.

Дефиниция - Максимално правдоподобна оценка (м.п.о.)

Нека $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$ са независими наблюдения над сл.в. $X - F(x, \theta)$.

Максимално правдоподобна оценка (м.п.о.) за неизвестен параметър $\boldsymbol{\theta}$ е тази стойност $\hat{\boldsymbol{\theta}}$, за която функцията на правдоподобие достига максимум.

Намирането на максимума обикновено се извършва по стандартния начин с намиране и нулиране на производната, т.е. $\hat{\theta}$ е решение на уравнението

$$\frac{\partial L(\overrightarrow{\mathrm{X}}, \theta)}{\partial \theta} = 0$$

Много често при решаването на тази задача се работи с логаритъм. Идеята е, че логаритъмът е монотонна функция и затова L и $\ln L$ достигат максимум в една и съща точка, а от друга страна тъй като L е дефинирана като произведение от плътности, то $\ln L$ е сума и има по-прост вид, съответно намирането на максимума на $\ln L$ е по-лесна изчислителна задача. Затова спрямо θ се решава следното уравнение

$$\frac{\partial \ln L(\overrightarrow{\mathrm{X}}, \theta)}{\partial \theta} = 0$$

В дефиницията на метода на максимално правдоподобие параметъра θ може да бъде и многомерен, т.е. да се търси максимум едновременно по няколко променливи. Тогава се решава система от съответните частни производни.

Пример

Нека сл.в. X попада по случаен начин в интервала $[0, \theta]$, т.е. $X \in U(0, \theta)$. Ще намерим м.п.о. оценка за θ, ако разполагаме с вектора на наблюденията.

Да припомним, вероятността X да е навсякъде в интервала е една и съща, което значи, че плътността е константа, в случая $f_X(x) = \frac{1}{\theta}$ за $x \in [0, \theta]$. Така за функцията на правдоподобие получаваме

$$L(\overrightarrow{\mathrm{X}}, \theta) = \prod_{k=1}^{n} f_{X_k}(x_k) = \begin{cases} \frac{1}{\theta^n} & , \forall x_k \in [0, \theta] \\ 0 & , \exists x_k \notin [0, \theta] \end{cases}$$

По метода на максимално правдоподобие, трябва да намерим стойността на θ, за която функцията $L(\overrightarrow{\mathrm{X}}, \theta)$ достига максимум. Ясно е, че тази функция е намаляваща по θ, тогава колкото по-малки стойности взима θ толкова по голяма е $L(\overrightarrow{\mathrm{X}}, \theta)$. От друга страна, ако θ стане по-малка от някое наблюдение X_k, то ще има наблюдение извън интервала $[0, \theta]$ и тогава $L(\overrightarrow{\mathrm{X}}, \theta) = 0$.

Следователно максимум ще се достига при най-малкото θ, по-голямо или равно на всички X_k, т.е оценката е

$$\hat{\theta} = \max(X_1, X_2, \ldots, X_n)$$

3. Точкови оценки за нормално разпределени данни

В тази част ще изведем максимално правдоподобни оценки за параметрите на данни, който са нормално разпределени. В огромна част от случайте данните са именно нормално разпределени.

Нека $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$ са независими наблюдения над $X \in N(\mu, \sigma^2)$. Ще намерим оценка за всяка от двете константи при условие, че другата е известна или неизвестна, т.е. ще разгледаме четири случая. В началото ще изведем функцията на правдоподобие, която е обща за четирите случая. Знаем плътността на нормалното разпределение. Точно същата е и плътноста на наблюденията, така за функцията на правдоподобие получаваме:

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$$L(\overrightarrow{\mathrm{X}}, \mu, \sigma) = \prod_{k=1}^{n} f_{X_k}(x_k) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} =$$

$$= \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \exp\{-\sum_{k=1}^{n} \frac{(x_k - \mu)^2}{2\sigma^2}\}$$

Ще пресметнем логаритъм на функцията на правдоподобие,

$$\ln L(\overrightarrow{\mathrm{X}}, \mu, \sigma) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{k=1}^{n} (x_k - \mu)^2$$

а също и производните по всеки от двата параметъра

$$\frac{\partial \ln L(\overrightarrow{\mathrm{X}},\mu,\sigma)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{k=1}^{n}(x_k-\mu)$$

$$\frac{\partial \ln L(\overrightarrow{\mathrm{X}},\mu,\sigma)}{\partial\sigma} = -\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{k=1}^{n}(x_k-\mu)^2$$

Сега ще разгледаме четирите възможни случая.

❧ Оценка за очакването μ, ако дисперсията σ^2 е известна.

В този случай трябва да намерим максимум на функцията на правдоподобие L по μ, т.е. трябва да решим уравнението $\frac{\partial \ln L}{\partial\mu}=0$ спрямо μ

Трябва да признаем, че този случай не е особено достоверен.

$$\frac{1}{\sigma^2}\sum_{k=1}^{n}(x_k-\mu)=0$$

$$\sum_{k=1}^{n}x_k-n\mu=0$$

От тук получаваме оценката

$$\hat{\mu}=\frac{1}{n}\sum_{k=1}^{n}x_k=\overline{X_n}.$$

.

$\overline{X_n}$ е стандартно означение за средноаритметично)

❧ Оценка за дисперсията σ^2, ако очакването μ е известно.

Сега трябва да намерим максимум на L по σ, т.е. $\frac{\partial \ln L}{\partial\sigma}=0$

$$-\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{k=1}^{n}(x_k-\mu)^2=0$$

Следователно търсената оценка е

$$\hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{n}(x_k-\mu)^2$$

❧ Оценка за очакването μ, ако дисперсията σ^2 е неизвестна.

И двата параметъра са неизвестни, което означава че трябва да намерим максимум на функцията на правдоподобие по две променливи, т.е. трябва да изразим μ от системата

$$\left|\begin{array}{l}\dfrac{\partial \ln L}{\partial\mu}=0\\[2ex] \dfrac{\partial \ln L}{\partial\sigma}=0\end{array}\right. \Leftrightarrow \left|\begin{array}{l}\dfrac{1}{\sigma^2}\displaystyle\sum_{k=1}^{n}(x_k-\mu)=0\\[2ex] -\dfrac{n}{\sigma}+\dfrac{1}{\sigma^3}\displaystyle\sum_{k=1}^{n}(x_k-\mu)^2=0\end{array}\right. \qquad \text{(IX .1)}$$

Първото уравнение има решение идентично с това от първия случай, а второто не влияе на стойността на μ. Така получената оценка съвпада с първи случай $\hat{\mu} = \overline{X_n}$, т.е. средното аритметично е оценка за очакването на нормално разпределена сл.в. независимо дали знаем или не дисперсията.

❧ Оценка за дисперсията σ^2, ако очакването μ е неизвестно.

Отново трябва да е в сила горната система (IX .1), но в този случай изразяваме σ^2. От второто уравнение получаваме търсената оценка, като μ заместваме с решението на първо уравнение

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \mu)^2 = \frac{1}{n}\sum_{k=1}^{n}\left(x_k - \overline{X_n}\right)^2$$

Тези статистики се използват не само за точково оценяване на параметрите на нормално разпределение, но са отправна точка при построяването на доверителни интервали или за проверката на хипотези свързани с нормалното разпределение. Ние подробно ще изследваме техните свойства, но преди това разгледаме още един начин за получаване на точкови оценки.

4. Метод на моментите

Ще опишем още един начин за намиране на точкови оценки. Нека сл.в. X има функция на разпределение $F_X(x,\theta)$ която ни е известна като вид, но не знаем стойността на параметъра θ. Ние можем да пресметнем математическото очакване на X, което също ще зависи от θ, т.е. $\mathbf{E}\,X = \mu(\theta)$ е функция на θ.

Нека $\overrightarrow{\mathrm{X}} = (X_1,\dots,X_n)$ е векторът на наблюденията, тези сл.в. са независими, еднакво разпределени и следователно за тях е изпълнен законът за големите числа.

$$\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k \quad \xrightarrow[n\to\infty]{P} \quad \mathbf{E}\,X = \mu(\theta)$$

Тогава, съвсем естествено е да изберем тази стойност на θ, за която ще имаме равенство, т.е. теоретичният момент $\mu(\theta)$ ще съвпадне със средното аритметично на наблюденията $\overline{X_n}$

Прието е $\overline{X_n}$ да се нарича емперичен момент, т.е. момент пресметнат от данните.

Дефиниция - оценка по метода на моментите

Нека $\overrightarrow{\mathrm{X}} = (X_1,\dots,X_n)$ са независими наблюдения над сл.в. $X - F(x,\theta)$, и нека съществува математическото очакване $\mathbf{E}\,X = \mu(\theta)$.

Оценка по метода на моментите (м.м.) за неизвестния параметър θ наричаме решението $\hat{\boldsymbol{\theta}}$, на уравнението $\mu(\theta) = \overline{X_n}$

По метода на моментите също могат да бъдат намирани оценки на повече от един параметър. Ако $\theta = (\theta_1,\theta_2,\dots,\theta_s)$, т.е. ако искаме да оценим едновременно s параметъра, тогава ще работим с първите s момента на сл.в. X и $\hat{\theta}$ е решение на следната система

$$\left| \begin{array}{l} \mu^1(\theta) = \overline{X_n^1} \\ \mu^2(\theta) = \overline{X_n^2} \\ \dots \\ \mu^s(\theta) = \overline{X_n^s} \end{array} \right.$$

където $\mu^i(\theta) = \mathbf{E}\, X^i$ са теоритичните моменти, а $\overline{X_n^i} = \frac{1}{n}\sum_{k=1}^{n}(X_k)^i$ емперичните. За да работи методът предполагаме, че теоритичните моменти до ред s съществуват, освен това системата трябва да има решение. Обикновено намирането на оценка по метода на моментите е по-лесно технически от откриването на м.п.о.

Пример

Отново ще намерим оценка за горната граница θ на равномерното разпределение, $X \in U(0, \theta)$.

Знаем че очакването е точно в средата на интервала, т.е. $\mathbf{E}\, X = \frac{\theta}{2}$ уравнението има вида

$$\frac{\theta}{2} = \overline{X_n}$$

Търсената оценка е $\hat{\theta} = 2\,\overline{X_n}$. Тази оценка явно е различна от м.п.о. която получихме преди (стр.??).

Ще изведем оценки за $N(\mu, \sigma^2)$ по метода на моментите. Ще приемем че и двата параметъра са неизвестни. Ще трябва да разгледаме системата от две уравнения. Нека както обикновено $X_1, \dots, X_n$ са независими наблюдения.

$$\left| \begin{array}{l} \mathbf{E}\, X = \overline{X_n} \\ \mathbf{E}\, X^2 = \overline{X_n^2} \end{array} \right.$$

Знаем, че при нормално разпределение $\mathbf{E}\, X = \mu$, така че първото уравнение директно води до решение $\hat{\mu} = \overline{X_n}$.

Също така знаем, че $\mathbf{D}\, X = \sigma^2$, но освен това $\mathbf{D}\, X = \mathbf{E}\, X^2 - (\mathbf{E}\, X)^2$, следователно $\mathbf{E}\, X^2 = \mathbf{D}\, X + (\mathbf{E}\, X)^2$ и можем да запишем второто уравнение на система във вида $\mathbf{E}\, X^2 = \sigma^2 + \mu^2 = \overline{X_n^2}$. Решението спрямо σ^2 е

$$\hat{\sigma}^2 = \overline{X_n^2} - \mu^2 = \overline{X_n^2} - \left(\overline{X_n}\right)^2$$

Оценките, който получихме по метода на моментите напълно съвпадат с максимално правдоподобните оценки. За $\hat{\mu}$ това е очевидно, а за $\hat{\sigma}^2$ следва от следното представяне на м.п.о.

$$\hat{\sigma}^2(\text{м.п.о.}) = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \overline{X_n}\right)^2 = \frac{1}{n}\sum_{k=1}^{n}\left(X_k^2 - 2X_k\overline{X_n} + \left(\overline{X_n}\right)^2\right) =$$

$$= \frac{1}{n}\sum_{k=1}^{n} X_k^2 - 2\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)\overline{X_n} + \frac{1}{n} n \left(\overline{X_n}\right)^2 = \overline{X_n^2} - 2\overline{X_n}.\overline{X_n} + \left(\overline{X_n}\right)^2 = \hat{\sigma}^2 (\text{м.м.})$$

5. Неизместеност

До тук разгледахме два начина за намиране на точкови оценки. Получените оценки могат да са еднакви, но могат и да са различни. Както видяхме от примерите, ако имаме наблюдения попадащи по случаен начин в интервал $[0, \theta]$, разполагаме с два начина да оценим параметъра. Единия е да вземем θ равно на най-голямото наблюдение, другият да приемем θ равно на два пъти средно аритметично на наблюденията. Въпросът, който възниква е кой от двата начина е по-добър. Имаме нужда от критерии, с които да оценяваме точковите оценки. Ще въведем някой важни характеристики на точковите оценки.

Дефиниция - Неизместеност

Нека $\overrightarrow{\mathrm{X}}$ са независими наблюдения над $X - F(x, \theta)$. Казваме, че $\hat{\theta}$ е неизместена оценка за θ, ако

$$\mathbf{E}\ \hat{\theta}(\overrightarrow{\mathrm{X}}) = \theta$$

Наблюденията $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$ са случайни величини, оценката е функция от сл.в. и като такава има своето математическо очакване. Логично е да поискаме това очакване да съвпадне с оценявания параметър. Ако последното не е изпълнено, то при използването на оценката ще се натрупва грешка. Прието е разликата $\mathbf{E}\ \hat{\theta} - \theta$ да се нарича "систематична грешка".

Ще проверим за неизместеност получените оценки на параметрите на нормалното разпределение (стр.??).

❧ Оценката за очакването $\hat{\mu}$ е неизместена, защото

$$\mathbf{E}\ \hat{\mu} = \mathbf{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}\, X_k = \frac{1}{n}\sum_{k=1}^{n}\mu = \mu$$

❧ Оценката за дисперсията σ^2 при известно очакване μ, е неизместена

$$\mathbf{E}\ \hat{\sigma}^2 = \mathbf{E}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)^2\right) = \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}\,(X_k - \mu)^2 = \frac{1}{n}\sum_{k=1}^{n}\mathbf{D}\, X_k = \sigma^2$$

❧ Оценката за дисперсията $\hat{\sigma}^2$ при неизвестно очакване е изместена. Ще използваме вида на $\hat{\sigma}^2$ изведен на стр.??.

$$\mathbf{E}\ \hat{\sigma}^2 = \mathbf{E}\left(\overline{X_n^2}\right) - \mathbf{E}\left(\overline{X_n}\right)^2 = \mathbf{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^2\right) - \mathbf{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)^2 =$$

Втората сума разделяме на две части - сума от квадратите и сума от смесените произведения

$$= \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}\,X_k^2 - \frac{1}{n^2}\mathbf{E}\left(\sum_{k=1}^{n} X_k^2 + \sum_{i=1}^{n}\sum_{j\neq i} X_i X_j\right) =$$

От независимостта на наблюденията следва $\mathbf{E}\,(X_i X_j) = \mathbf{E}\,X_i\,\mathbf{E}\,X_j$

$$= \frac{n-1}{n^2}\sum_{k=1}^{n}\mathbf{E}\,X_k^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\mathbf{E}\,(X_i X_j) =$$

$$= \frac{n-1}{n^2}\sum_{k=1}^{n}\mathbf{E}\,X^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\mathbf{E}\,X_i\,\mathbf{E}\,X_j =$$

Наблюденията са копия на сл.в. X, следователно $\mathbf{E}\,X_i = \mathbf{E}\,X_j = \mathbf{E}\,X$. Така всички събираеми в двойната сума са еднакви, а броят на събираемите е $n(n-1)$.

$$= \frac{n-1}{n}\mathbf{E}\,X^2 - \frac{n-1}{n}(\mathbf{E}\,X)^2 = \frac{n-1}{n}\,\mathbf{D}\,X = \frac{n-1}{n}\,\sigma^2 \neq \sigma^2$$

Оказа се, че оценката за дисперсията при неизвестно очакване

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \overline{X_n}\right)^2$$

е изместена с множител $\frac{n-1}{n}$.

Лесно можем да получим неизместена оценка за дисперсията, достатъчно е да умножим $\hat{\sigma}^2$ с реципрочния множител $\frac{n}{n-1}$. По този начин се получава известната оценка S^2.

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X_n}\right)^2$$

Дотук определяхме неизместеността като използвахме свойствата на математическото очакване, понякога при по-нетипични точкови оценки този подход е невъзможен.

Пример

Ще проверим за неизместеност статистиката получена в първия пример $\hat{\theta} = \max(X_1, X_2, \ldots, X_n)$, т.е. интересува ни дали $\mathbf{E}\,\hat{\theta} = \theta$. За целта първо ще намерим функцията на разпределение на $\hat{\theta}$. Нека $y \in [0, \theta]$

$$F_{\hat{\theta}}(y) = \mathbf{P}(\hat{\theta} < y) = \mathbf{P}(\max(X_1, X_2, \ldots, X_n) < y) =$$

$$= \mathbf{P}(X_1 < y, X_2 < y, \ldots, X_n < y) = \mathbf{P}(X_1 < y)\mathbf{P}(X_2 < y), \ldots, \mathbf{P}(X_n < y)$$

Наблюдението X_k е копие на X, тогава $\forall k$: $\mathbf{P}(X_k < y) = \int_0^y \frac{1}{\theta} dx = \frac{y}{\theta}$. Така получаваме функцията на разпределение, а от нея и плътността

$$F_{\hat{\theta}}(y) = \left(\frac{y}{\theta}\right)^n, \qquad f_{\hat{\theta}}(y) = \frac{\partial F_{\hat{\theta}}(y)}{\partial y} = \frac{n\, y^{n-1}}{\theta^n}, \qquad 0 \le y \le \theta$$

Тогава очакването на $\hat{\theta}$ е следното

$$\mathbf{E}\,\hat{\theta} = \int_0^\theta y\, \frac{n\, y^{n-1}}{\theta^n} dy = \frac{n}{\theta^n}\, \frac{y^{n+1}}{n+1}\bigg|_{y=0}^{\theta} = \frac{n}{n+1}\,\theta \neq \theta$$

Оценката явно е изместена, но не е проблем да се превърне в неизместена, $\frac{n+1}{n}\max(X_1, X_2, \ldots, X_n)$ е такава.

6. Състоятелност

Друго съвсем естествено изискване към точковите оценки е - когато броя на наблюденията расте оценките да стават по-точни, т.е. вероятността за грешка да намалява.

Състоятелност

Нека $X_1, X_2, \ldots X_n$ са независими наблюдения над X с функция на разпределение $F(x, \theta)$. Казваме, че $\hat{\theta}$ е състоятелна оценка за θ, ако

$$\hat{\theta} \xrightarrow[n\to\infty]{\mathbf{P}} \theta$$

Ще припомним, сходимостта по вероятност от дефиницията означава, че за всяко ε е изпълнено

$$\mathbf{P}(|\hat{\theta} - \theta| \ge \varepsilon) \xrightarrow[n\to\infty]{} 0$$

Съгласно закона за големи числа

$$\overline{X_n} \xrightarrow[n\to\infty]{P} \mathbf{E}\,X = \mu(\theta)$$

По метода на моментите, ние избираме $\hat{\theta}$ така, че горната сходимост се превръща в равенство. Затова оценките получени по метода на моментите винаги са състоятелни. Следователно и оценките за нормално разпределение, който ние изведохме са състоятелни.

7. Ефективност

Обобщение

Математическо очакване

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} X_k = \overline{X_n}$$

Дисперсия

при μ известно	при μ неизвестно
$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - \mu)^2$	$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} \left(X_k - \overline{X_n}\right)^2$

X . Доверителни интервали

В предишната глава разгледахме начините за оценяване на неизвестен параметър θ, от който зависи наблюдаваната случайна величина. При това търсехема точкова оценка на параметъра, т.е. функция $\hat{\theta}(\overrightarrow{\mathrm{X}})$, зависеща от наблюденията $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$, като стойността на тази функция приемахме за истинска стойност на θ.

Ясно е, че в действителност оценката, която получаваме дава приблизителна стойност за истинският параметър. Наистина добавянето на дори само едно допълнително наблюдение най-често води до промяна на получената оценка. Тази оценка ще бъде безинтересна от практическа гледна точка, ако не познаваме нейната точност, т.е. вероятността да сме допуснали грешка. Намирането на доверителен интервал решава точно този проблем - вместо една единствена стойност за неизвестен параметър θ ние ще намерим интервал (т.е. границите) в който θ попада с достатъчно висока вероятност.

Ще формулираме задачата формално. Нека $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$ са независими наблюдения над сл.в. X. Ще предполагаме, че типа на разпределението на X е известно, но то зависи от неизвестен параметър θ, възможно е $\theta = (\theta_1, \theta_2, \ldots, \theta_s)$ да бъде многомерен. Търсим интервал $I = I(\overrightarrow{\mathrm{X}})$, такъв че

$$\mathbf{P}(\theta \in I) \geq \gamma$$

Константата γ се нарича **ниво на доверие**, то е предварително зададено, определянето му не е математическа задача. Нивото на доверие зависи от най-голямата вероятност с която можем да си позволим да допуснем грешка, от гледна точка на щетите, който ще понесем при грешка в сатистиката, финансовите загуби и т.н.

Обикновено се избира γ в интервал $[0.9, 0.99]$. Колкото по-високо е нивото на доверие, толково по-широк става интервала и затова носи по-малко информация за оценявания параметър. При $\gamma = 1$ интервала, който ще получим за θ е $[-\infty, \infty]$, което не казва нищо за конкретната стойност на θ. Например, с вероятност единица можем да твърдим, че ръстът на хората е между нула и три метра. Ползата от подобна статистика обаче е съмнителна. При ниски нива на доверие интервалът е по тесен, по информативен, но тогава има по-голяма вероятност за грешка, т.е. възможност θ изобщо да не лежи в този интервал.

Намирането на доверителен интервал може да се разглежда като намиране на две точкови оценки $I_1(\overrightarrow{\mathrm{X}})$ и $I_2(\overrightarrow{\mathrm{X}})$, които са съответно долна и горна граница на интервала

$$\mathbf{P}(I_1 < \theta < I_2) \geq \gamma$$

1. Централна статистика

Начин за конструиране на доверителни интервали е предложен от Нейман и Уилкс в 40-те години на миналия век. Идеята е да се конструира функция, която зависи от от наблюденията и от неизвестния параметър, но разпределението и не зависи от параметъра.

Дефиниция - Централна статистика

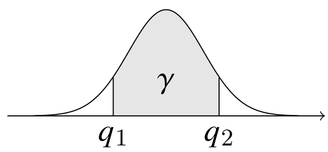
Нека $\mathrm{T}(\overrightarrow{\mathrm{X}}, \theta)$ е функция, която отговаря на следните условия:

1) $\mathrm{T}(\overrightarrow{\mathrm{X}}, \theta)$ е непрекъсната и монотонна по θ.

2) разпределението на $\mathrm{T}(\overrightarrow{\mathrm{X}}, \theta)$ не зависи от θ, т.е. $\mathbf{P}(\mathrm{T}(\overrightarrow{\mathrm{X}}, \theta) < \mathrm{x})$ зависи само от x, но не и от θ.

Намирането на централна статистикаи не е тривиална задача. Обикновено за основа се използват точковите оценки за параметъра, но се изискват задълбочени познания за разпределенията за да се конструира.

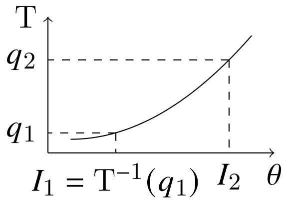
Все пак, ако разполагаме с централна статистика, лесно можем да построим доверителен интервал за параметъра θ. Процедурата се състои в следното. Тъй като разпределението на T е известно можем да намерим квантили q_1 и q_2, такива, че $\mathbf{P}(q_1 < \mathrm{T} < q_2) = \gamma$.



На схемата разпределението е нормално, но методът работи с произволно познато разпределение.

Разбира се, квантилите q_1 и q_2, не се определят еднозначно от горното равенство. Съществуват много двойки квантили с това свойство лицето между тях да бъде точно γ. Естествено е да предпочетем този интервал, който има най-малка дължина. При симетричните разпределения на това изискване отговарят квантили q_1 и q_2 такива, че лявата и дясната опашка съвпадат, т.е. $\mathbf{P}(\mathrm{T} < q_1) = \mathbf{P}(\mathrm{T} > q_2) = \dfrac{1-\gamma}{2}$. На практика и при произволни разпределения (независимо дали симетрични или не) се работи точно с тези квантили от съображения за простота.

След като определим квантилите не е проблем да обърнем функцията T, тя е обратима по θ поради свойство 1), и така да получим доверителния интервал за θ.



На практика решаваме неравенствата спрямо θ.

$$\gamma = \mathbf{P}(q_1 < \mathrm{T} < q_2) = \mathbf{P}(I_1 < \theta < I_2)$$

2. Свойства на нормалните извадки

В тази глава ще изведем доверителни интервал за очакването μ и дисперсия σ^2 на нормално разпределена случайна величина. Като централна статистика ще използваме точковите оценки (стр.105), които получихме в предходната глава. За начало ще докажем техни важни свойства.

Ще се наложи да изследваме разпределението на централната статистика

$$\mathrm{T} = \frac{\overline{X} - \mu}{S/\sqrt{n}} \qquad (\star)$$

Твърдение

Нека $X_k \in N(\mu, \sigma^2)$, $k = 1, \dots, n$, Тогава $\overline{X}$ и S^2 са независими сл.в.

Забележка. Твърдението на пръв поглед е парадоксално, доколкото и $\overline{X_n}$ и S^2 зависят от едни и същи сл.в. То кореспондира с други известни резултати за нормално рапределение. Например, ако X и Y са независими то $X + Y$ и $X - Y$ също са независими.

Док. В началото ще представим S^2 по следния начин.

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2 = \frac{1}{n-1}\left[\left(X_1 - \overline{X}\right)^2 + \sum_{k=2}^{n}\left(X_k - \overline{X}\right)^2\right]$$

Ще преработим първото събираемо. Лесно се вижда, че $\sum\left(X_k - \overline{X}\right) = 0$.

$$\sum_{k=1}^{n}\left(X_k - \overline{X}\right) = \sum_{k=1}^{n} X_k - n\overline{X} = \sum_{k=1}^{n} X_k - \sum_{k=1}^{n} X_k = 0$$

Тогава

$$\left(X_1 - \overline{X}\right) + \sum_{k=2}^{n}\left(X_k - \overline{X}\right) = 0$$

Следователно

$$\left(X_1 - \overline{X}\right)^2 = \left(-\sum_{k=2}^{n}\left(X_k - \overline{X}\right)\right)^2 = \left(\sum_{k=2}^{n}\left(X_k - \overline{X}\right)\right)^2$$

Ще заместим в израза за S^2

$$S^2 = \frac{1}{n-1}\left[\left(\sum_{k=2}^{n}\left(X_k - \overline{X}\right)\right)^2 + \sum_{k=2}^{n}\left(X_k - \overline{X}\right)^2\right]$$

Това показва, че S^2 е функция само на $\left(X_2 - \overline{X}, X_3 - \overline{X}, \ldots, X_n - \overline{X}\right)$. Ще докажем, че тези случайни величини и $\overline{X}$ са независими.

Без ограничение на общността можем да считаме, че $X_k \in N(0,1)$. Наистина, ако това не е така можем да стандартизираме X_k, при което $\overline{X}$ и S^2 също ще се променят с константа, а промяната с константа не влияе на независимостта. Съвместната плътност на $X_1, \ldots, X_n$ е

$$f(x_1, \ldots, x_n) = \prod_{k=1}^{n} f_{X_k}(x_k) = \frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left(-\frac{1}{2}\sum_{k=1}^{n} x_k^2\right)$$

Сега ще направим следната смяна на променливите

$$\left|\begin{array}{l} y_1 = \overline{x} \\ y_2 = x_2 - \overline{x} \\ \ldots \\ y_n = x_n - \overline{x} \end{array}\right.$$

Тогава обратната трансформация и якобианът са съответно

$$\left|\begin{array}{l} x_1 = y_1 - \sum_{k=2}^{n} y_k \\ x_2 = y_2 + y_1 \\ \ldots \\ x_n = y_n + y_1 \end{array}\right. \qquad J = \left|\begin{array}{rrrcr} 1 & -1 & -1 & \ldots & -1 \\ 1 & 1 & 0 & \ldots & 0 \\ \ldots & & & & \\ 1 & 0 & 0 & \ldots & 1 \end{array}\right| = n$$

Якубианът се пресмята елементарно, ако всеки ред се прибави към първия.

Съвместната плътност на новите променливи има вида

$$f(y_1, \ldots, y_n) = \frac{n}{(2\pi)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2}\left[\left(y_1 - \sum_{k=2}^{n} y_k\right)^2 + \sum_{k=2}^{n}(y_k + y_1)^2\right]\right\}$$

Можем да разделим израза в експонентата на две части, едната от които зависи само от y_1. За целта използваме формула за съкратено умножение и групираме.

$$\left(y_1 - \sum_{k=2}^{n} y_k\right)^2 + \sum_{k=2}^{n}(y_k + y_1)^2 = n\, y_1^2 + \left(\sum_{k=2}^{n} y_k\right)^2 + \sum_{k=2}^{n} y_k^2$$

$$= y_1^2 - 2y_1 \sum_{k=2}^{n} y_k + \left(\sum_{k=2}^{n} y_k\right)^2 + \sum_{k=2}^{n} y_k^2 + 2y_1 \sum_{k=2}^{n} y_k + (n-1)y_1^2 =$$

По този начин представяме съвместната плътност като произведение от две плътности.

$$f(y_1, \dots, y_n) = \left[\left(\frac{n}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{n y_1^2}{2}\right\}\right]\left[\frac{n^{\frac{1}{2}}}{(2\pi)^{\frac{n-1}{2}}} \exp\left\{-\frac{1}{2}\left(\sum_{k=2}^{n} y_k\right)^2 - \frac{1}{2}\sum_{k=2}^{n} y_k^2\right\}\right]$$

Не е трудно да разпознаем първата $N(0, \frac{1}{n})$, точно това се очаква и за плътността на $Y_1 = \overline{X}$. Фактът, че съвместната плътност се разпада на произведение от две маргинални плътности означава, че съответните сл.в. са независими. Следователно $\overline{X}$ е независимо с $\left(X_2 - \overline{X}, X_3 - \overline{X}, \dots, X_n - \overline{X}\right)$, а тогава и $X \perp\!\!\!\perp S^2$, тъй като S^2 е функция само от тях, както доказахме. ✓

Вече знаем, че числителят и знаменателят на централната статистика (⋆) T са независими, освен това познаваме разпределението в числителя, остава да определим как е разпределен знаменателя.

Твърдение

Нека $X_k \in N(\mu, \sigma^2)$, $k = 1, \dots, n$, Тогава

$$\frac{(n-1)}{\sigma^2} S^2 \in \chi^2(n-1) \qquad \text{(X .1)}$$

Док. Ще използваме равенството $\sum\left(X_k - \overline{X}\right) = 0$, което доказахме по-горе. Тогава

$$\sum_{k=1}^{n}(X_k - \mu)^2 = \sum_{k=1}^{n}\left[(X_k - \overline{X}) + (\overline{X} - \mu)\right]^2 =$$

$$= \sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2 + 2(\overline{X} - \mu)\sum_{k=1}^{n}\left(X_k - \overline{X}\right) + \sum_{k=1}^{n}\left(\overline{X} - \mu\right)^2$$

$$= \sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2 + n\left(\overline{X} - \mu\right)^2$$

Ще разделим двете страни на полученото равенство на σ^2.

$$\frac{1}{\sigma^2}\sum_{k=1}^{n}(X_k-\mu)^2 = \frac{1}{\sigma^2}\sum_{k=1}^{n}\left(X_k-\overline{X}\right)^2 + n\left(\frac{\overline{X}-\mu}{\sigma}\right)^2$$

Използваме дефиницията на S^2.

$$\underbrace{\sum_{k=1}^{n}\left(\frac{X_k-\mu}{\sigma}\right)^2}_{Z_1} = \underbrace{\frac{n-1}{\sigma^2}S^2}_{Z_2} + \underbrace{\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2}_{Z_3}$$

Лесно се вижда, че Z_1 и Z_3 имат хи-квадрат разпределение. Наистина, по условие $X_k \in N(\mu,\sigma^2)$, следователно $\frac{X_k-\mu}{\sigma} \in N(0,1)$. В лекция VIII доказахме че сумата от квадратите на нормални сл.в. има хи-квадрат разпределение

$$Z_1 = \sum_{k=1}^{n}\left(\frac{X_k-\mu}{\sigma}\right)^2 \in \chi^2(n)$$

При решаването на първата задача показахме $(*)$, че

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \in N(0,1) \qquad \Longrightarrow \qquad Z_3 = \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \in \chi^2(1)$$

Освен това знаем, че $\overline{X_n}$ и S^2 са независими, тогава $Z_2 \perp\!\!\!\perp Z_3$ защото се различават само с константи.

Да обобщим

$$Z_1 = Z_2 + Z_3$$

Освен това $Z_1 \in \chi^2(n)$, $Z_3 \in \chi^2(1)$ и $Z_2 \perp\!\!\!\perp Z_3$. Тогава единствената възможност е $Z_2 \in \chi^2(n-1)$, с което всичко е доказано. Все пак това не е прецизно доказателство, ще направим такова за по-любознателните.

В глава VI доказахме, че за функцията пораждаща моментите е изпълнено

$$Z = X + Y,\ X \perp\!\!\!\perp Y \quad \Rightarrow \quad M_Z(t) = M_X(t)\,M_Y(t)$$

В същата глава изведохме функцията пораждаща моментите на хи-квадрат рапределение, а именно

$$X \in \chi^2(m) \qquad \Longleftrightarrow \qquad M_X(t) = (1-2t)^{-m/2}$$

Лесно можем да докажем следните две твърдения:

$$1) \quad X \in \chi^2(m),\ Y \in \chi^2(k) \quad \Rightarrow \quad Z \in \chi^2(m+k)$$

Защото

$$M_Z(t) = (1-2t)^{-m/2}(1-2t)^{-k/2} = (1-2t)^{-(m+k)/2}$$

2) $X \in \chi^2(m),\ Z \in \chi^2(m+k) \quad \Rightarrow \quad Y \in \chi^2(k)$

Понеже

$$M_Y(t) = \frac{M_Z(t)}{M_X(t)} = \frac{(1-2t)^{-(m+k)/2}}{(1-2t)^{-m/2}} = (1-2t)^{-k/2}$$

✓

3. Доверителни интервали за нормални данни

Ще намерим доверителни интервали за параметрите на нормално разпределена сл.в. $X \in N\left(\mu, \sigma^2\right)$.

❧ Доверителен интервал за очакването μ при известна дисперсия σ^2

Като цяло този случай не е особено достоверен. Да познаваме дисперсията но да не знаем очакването на една случайна величина рядко се реализира на практика. Възможно е например, при повторно изследване на пациенти да се предполага, че дисперсията зависи от индивидуалните особености на всеки човек, а очакването от действието на изследваното лекарство. Така дисперсията ще бъде известна от предишното изследване. По-разумно е разбира се да се провери доколко дисперсиите съвпадат.

Ние ще разгледаме този случай защото е по-прост от теоретична страна и е отправна точка за следващия по-сложен случай. В предишната лекция изведохме точкава оценка за параметър μ, а именно $\hat{\mu} = \overline{X}$. Ще превърнем тази точкова оценка в централна статистика. Знаем, че сумата на нормално разпределени сл.в. е нормално разпределена, от $X_k \in N\left(\mu, \sigma^2\right)$ следва

$$\sum_{k=1}^{n} X_k \in N\left(n\mu, n\sigma^2\right) \quad \Longleftrightarrow \quad \overline{X} = \frac{\sum X_k}{n} \in N\left(\mu, \frac{\sigma^2}{n}\right)$$

Това равенство показва защо средноаритметичното е по-добра статистика от едно отделно наблюдение - има по-малка дисперсия.

Стандартизираме

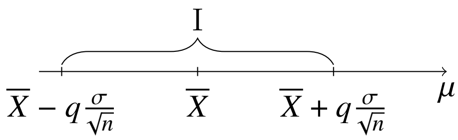
$$\mathrm{T} = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n\sigma^2}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \in N(0,1) \qquad (*)$$

Видно е, че T е монотонно намаляваща и непрекъсната по μ, а разпределението и е стандартно нормално, т.е. не зависи от μ. Следователно, T е търсената централна статистика.

Ако имаме зададено ниво на доверие γ, избираме съответните симетрични квантили $-q$ и q и решаваме получените неравенства спрямо μ

$$\gamma = \mathbf{P}\left(-q < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < q\right) = \mathbf{P}\left(\overline{X} - q\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + q\frac{\sigma}{\sqrt{n}}\right)$$

Доверителния интервал, който получаваме $\mathrm{I} = \overline{X} \pm q\frac{\sigma}{\sqrt{n}}$ е околност на средното аритметично $\overline{X}$.

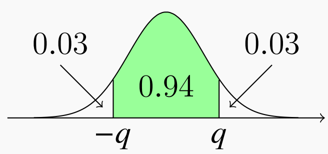


Дължината на интервала зависи от нивото на доверие, дисперсията на сл.в. X, както и от броя на наблюденията. Когато броят на наблюденията расте интервала става по-тесен, т.е. по-точен.

Пример

Няколко опитни полета са третирани с нов тип тор, след което е измерен добива: 4.5, 6.3, 5, 4.8, 5.2. Да се построи доверителен интервал, с ниво на доверие $\gamma = 0.94$ за средния добив, ако от предишни изследвания знаем, че данните са нормално разпределени с дисперсия 0.5.

Разполагаме с пет наблюдения над сл.в. $X \in N(\mu, 0.5)$ и търсим доверителен интервал за μ. Пресмятаме $\overline{X} = 5.16$ и $\sigma = \sqrt{0.5} = 0.7071$



От таблица за нормално разпределение вадим квантил $q = q_{0.97} = 1.88$. Доверителния интервал, който получаваме е

$$\mathrm{I} = 5.16 \pm 1.88\frac{0.7071}{\sqrt{5}} = (4.56; 5.75)$$

Това означава че има 94% вероятност средния добив да е в този интервал, съответно 6% вероятност за грешка. Обърнете внимание! Ние не твърдим, че 94% от полетата имат такъв добив. Интервалът се отнася за средния добив от полетата, а не за конкретно поле.

❧ Доверителен интервал за очакването μ при неизвестна дисперсия σ^2

Единствената разлика с предходния случай е, че сега дисперсията е неизвестна. Въпросът който стои е какво би станало, ако използваме същата централна статистика за построяване на доверителен интервал, но заместим неизвестната дисперсия σ^2 с оценката за нея. Ще се върнем към централната статистика Т (★), ще разделим на σ и ще пренаредим

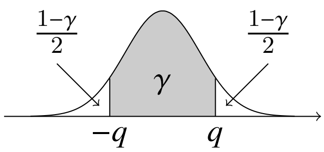
$$\mathrm{T} = \frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \Big/ (n-1)}}$$

Както многократно отбелязвахме числителят е с разпределение $N(0,1)$, знаменателят е от типа $\sqrt{\chi^2(n-1)/(n-1)}$ и те са независими. От последното твърдение в лекция VIII следва, че Т има разпределение на Стюдънт с $n-1$ степени на свобода

$$\mathrm{T} = \frac{\overline{X} - \mu}{S/\sqrt{n}} \in t(n-1)$$

Разпределението на Стюдънт е подобно на нормалното разпределение, също е симетрично, затова и квантилите се определят по същия начин, т.е. при зададено ниво на доверие γ намираме квантили $-q$ и q от таблица с $t(n-1)$ разпределение, такива че

$$\mathbf{P}(-q < \mathrm{T} < q) = \gamma$$



Ясно е че

$$\mathbf{P}(T < q) = \frac{1-\gamma}{2} + \gamma = \frac{1+\gamma}{2}$$

Тогава q е $\frac{1+\gamma}{2}$ квантил на разпределение на Стюдънт $n-1$ степени на свобода. А търсеният доверителен интервал за μ е $\mathrm{I} = \overline{X} \pm q\frac{S}{\sqrt{n}}$

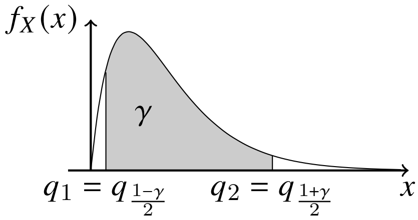
И двата доверителни интервала за μ, които построихме, при известна и при неизвестна дисперсия, използват статистика Т, в която основна роля играе $\overline{X} = \frac{\sum X_k}{n}$. Ние предполагахме, че $X_k \in N(\mu, \sigma^2)$ и затова $\overline{X}$ е нормално разпределено. В действителност $\overline{X}$ ще бъде нормално разпределено, дори първоначалните наблюдения да не са, стига те да са достатъчно на брой, това твърди централна гранична теорема от глава VI. Поради тази причина, статистиката Т (★) може да се използва за намиране на доверителен интервал за математическото очакване на произволни случайни величини, достатъчно е наблюденията да са много на брой. Под "много" най-често се разбира повече от 30.

❧ Доверителен интервал за дисперсия σ^2 при известно очакване μ

Като основа за конструиране на централна статистика отново ще ползваме съответната точкова оценка. $\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \mu)^2$. Нека

$$T = \frac{1}{\sigma^2}\sum_{k=1}^{n}(X_k - \mu)^2$$

По горе (стр.??) показахме, че $T \in \chi^2(n)$, осван това функцията явно е монотонна и непрекъснато по σ, следователно тя е централна статистика. Избираме съответните квантили на хи-квадрат разпределението и решаваме спрямо σ^2



$$\gamma = \mathbf{P}(q_1 < T < q_2) = \mathbf{P}\left(\frac{\sum_{k=1}^{n}(X_k - \mu)^2}{q_2} < \sigma^2 < \frac{\sum_{k=1}^{n}(X_k - \mu)^2}{q_1}\right)$$

❧ Доверителен интервал за дисперсия σ^2 при неизвестно очакване μ

В централната статистика от предната задача ще заместим очакването μ с точковата оценка за него $\overline{X}$.

$$T = \frac{1}{\sigma^2}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2$$

T се оказва точно статистиката (X .1) от Твърдение 2.

$$T = \frac{n-1}{\sigma^2}S^2 \in \chi^2(n-1)$$

Разликата с предишния случай е в степените на свобода тук те са $n-1$. Доверителния интервал е аналогичен.

$$I = \left(\frac{\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2}{q_2}; \frac{\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2}{q_1}\right)$$

Обобщение

Доверителни интервали за $N(\mu, \sigma^2)$

- Доверителен интервал за μ при известна σ^2

$$\mathrm{I} = \left(\overline{X} - q\frac{\sigma}{\sqrt{n}}\ ;\ \overline{X} + q\frac{\sigma}{\sqrt{n}}\right) \qquad q \in N(0,1)$$

- Доверителен интервал за μ при неизвестна σ^2

$$\mathrm{I} = \left(\overline{X} - q\frac{S}{\sqrt{n}}\ ;\ \overline{X} + q\frac{S}{\sqrt{n}}\right) \qquad q \in t(n-1)$$

- Доверителен интервал за σ^2 при известно μ

$$\mathrm{I} = \left(\frac{\sum_{k=1}^{n}(X_k - \mu)^2}{q_2}\ ;\ \frac{\sum_{k=1}^{n}(X_k - \mu)^2}{q_1}\right) \qquad q_1, q_2 \in \chi^2(n)$$

- Доверителен интервал за σ^2 при неизвестно μ

$$\mathrm{I} = \left(\frac{\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2}{q_2}\ ;\ \frac{\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2}{q_1}\right) \qquad q_1, q_2 \in \chi^2(n-1)$$

XI . Проверка на хипотези

В статистиката много често ни се налага да проверяваме истинността на някое твърдение. То може да е съвсем свободно формулирано, например - “лекарството подобрява състоянието на болните”, или да носи в себе си конкретна информация - “5% от заболелите нямат определен симптом”, или да бъде формално - “броят на заразяванията е експоненциално разпределена сл.в. с параметър $\lambda = 2$”. Във всички случай ни е необходим критерии, с който да отсъдим дали твърдението е истина. Също така искаме да знаем вероятността за допускане на грешка.

Проблемът с проверката на хипотези е изследван от Нейман и Пирсън през 30 години на XX век. Предложения подход в някакъв смисъл е аналогичен на разсъждение с допускане на противното. Допускаме, че хипотезата е вярна. Конструираме събитие A, което би се изпълнило с голяма вероятност при вярна хипотеза. Извършваме опит, правим наблюдения и ако събитието A не настъпи имаме основание да отхвърлим хипотезата. Това е така, защото вероятността събитието A да не настъпи при вярна хипотеза е много малка, т.е. наблюдаваме сбъдването на невероятното събитие $\overline{A}$. Хипотезата, която сме направили влиза в противоречие с наблюденията, затова я отхвърляме.

Ако събитието A настъпи ние нямаме основание да отхвърляме хипотезата.

1. Критична област. Грешка от I и II тип.

Ще зададем подходящ математически модел за проверка на хипотези, като ще започнем от най-простия възможен случай. Ще предполагаме, че хипотезата се отнася за параметър θ, от който зависи сл.в. X. Както обикновено $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$ са независими наблюдения над нея. Проверяваме

$$\begin{array}{lrcl} & \text{хипотеза } H_0 & : & \theta = \theta_0 \\ \text{срещу} & & & \\ & \text{алтернатива } H_1 & : & \theta = \theta_1 \end{array}$$

Съвсем естествено е хипотезата и алтернативата да са взаимноизключващи, т.е. не могат да се едновременно верни.

Константите θ_0 и θ_1 са известни, предварително зададени, не се определят от данните. В случай като този хипотезата и алтернативата са прости, доколкото става дума за константи. За сложна алтернатива говорим тогава, когато вместо една единствена стойност θ_1 е зададено цяло множество T_1. Най-често използваните сложни алтернативи са от вида:

$$\underbrace{H_1 : \theta < \theta_0, \qquad H_1 : \theta > \theta_0,}_{\text{едностранни}} \qquad\qquad \underbrace{H_1 : \theta \neq \theta_0}_{\text{двустранна}}$$

Основната ни цел е да конструираме множество W в n-мерното пространство, което наричаме **критична област**, и което е критерий за проверка на хипотезата:

$$\begin{aligned} &\text{Ако } \overrightarrow{\mathrm{X}} \in W \quad \Rightarrow \quad \text{Отхвърляме } H_0 \\ &\text{Ако } \overrightarrow{\mathrm{X}} \notin W \quad \Rightarrow \quad \text{Приемаме } H_0 \end{aligned}$$

Възможно е разбира се да отхвърлим хипотезата H_0 дори когато е вярна. Тогава допускаме **грешка от I род**. Вероятността за грешка от първи род се нарича **ниво на съгласие**

$$\alpha = \mathbf{P}\left(\overrightarrow{\mathrm{X}} \in W | H_0\right)$$

Желателно е това число да бъде малко, обикновено то се избира предварително. Нивото на съгласие α не определя критичната област W еднозначно, съществуват много области с равно ниво на съгласие.

Възможно е, също така, хипотезата да е невярна, но ние да я приемем, това се нарича **грешка от II род**, вероятността за нея бележим с

$$\beta = \mathbf{P}\left(\overrightarrow{\mathrm{X}} \notin W | H_1\right)$$

Противоположната вероятност $\pi = 1 - \beta$ да отхвърлим хипотезата H_0, ако тя е погрешна, наричаме **мощност на критерия**. Стремим се да направим мощността възможно най-голяма.

		хипотеза която приемаме	
		H_0	H_1
вярна хипотеза	H_0	✓	грешка от I род
	H_1	грешка от II род	✓

Ако изберем по-голяма област W вероятността да попаднем в нея ще нарастне, съответно грешката от I род α ще е по-голяма, а β ще намалее. Обратно, ако изберем по-малка W вероятността да попаднем извън нея ще нарастне, т.е. β ще се увличи, а α ще намалее. В крайния случай $W = \emptyset$ явно $\alpha = 0$. Когато едната грешка расте другата намалява. Коя от двете грешки е по-опасна зависи от конкретните хипотези.

Например

$$\begin{aligned} H_0 &: \text{виното е отровно} \\ H_1 &: \text{виното не е отровно} \end{aligned}$$

Грешката от I род е “виното е отровно, а ние приемаме че не е”, очевидно тове е изключително неприятна грешка.

Грешката от II род “виното не е отровно, но ние приемаме че е”, няма да доведе до трагични последствия.

Изборът на основна хипотеза се определя и от това коя от двете грешки искаме да избегнем. Така фиксираме максималната грешка от I род, която можем да си позволим да допуснем, обикновено това е число от порядък $\alpha = 0.05,\ 0.01,\ 0.001$ и т.н. и при това условие избираме такава област W, че грешката от II род да бъде възможно най-малка, т.е. фиксираме α и търсим минимум по β.

Ако съществува област върху която се изпълняват тези условия, то казваме че това е **оптимална критична област** (о.к.о.) и я бележим с W^*.

2. Лема на Нейман Пирсън

Намирането на оптималната критична област се извършва по лемата на Нейман - Пирсън. Ще предполагаме, че разпределението на сл.в. X е известно, но то зависи от параметър θ, за който формулираме проста хипотеза срещу проста алтернатива.

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \end{aligned}$$

Нека $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$ са наблюденията над X, а $L(x, \theta)$ е съответната функция на правдоподобие

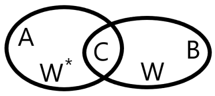
$$L(x,\theta) = \prod_{k=1}^{n} f_{X_k}(x_k, \theta), \qquad x = (x_1, x_2, \ldots, x_n)$$

Ще въведем означенията $L_0(x) = L(x, \theta_0)$ и $L_1(x) = L(x, \theta_1)$.

Лема на Нейман - Пирсън

При проверка на проста хипотеза срещу проста алтернатива с ниво на съгласие α, ако W^* е такава, че $\mathbf{P}\left(\overrightarrow{\mathrm{X}} \in W^* \mid H_0\right) = \alpha$ и съществува константа $K = K(\alpha)$ за която

$$\begin{aligned} L_1(x) &\geq K\, L_0(x), \qquad \forall x \in W^* \\ L_1(x) &\leq K\, L_0(x), \qquad \forall x \notin W^* \end{aligned}$$



Тогава W^* е оптимална критична област.

Док. Нека W е произволна друга критична област за която грешката от I род е точно α

$$\mathbf{P}\left(\overrightarrow{\mathrm{X}} \in W|H_0\right) = \alpha$$

Ще докажем че върху W^* грешката от II род е по-малка отколкото върху W, или което е същото, че мощността е по-голяма. Ще въведем означенията

$$C = W \cap W^*, \qquad A = W \setminus W^*, \qquad B = W^* \setminus W$$

Ако е изпълнена хипотеза H_0, то съвместната плътност на наблюденията е L_0 и α може да се пресметне чрез съответния n-мерен интеграл.

$$\alpha = \mathbf{P}\left(\overrightarrow{\mathrm{X}} \in W^*|H_0\right) = \int_{W^*} L_0(x)\,dx = \int_A L_0(x)\,dx + \int_C L_0(x)\,dx$$
$$\alpha = \mathbf{P}\left(\overrightarrow{\mathrm{X}} \in W\,|H_0\right) = \int_W L_0(x)\,dx = \int_B L_0(x)\,dx + \int_C L_0(x)\,dx$$

Следователно

$$\int_A L_0(x)\,dx = \int_B L_0(x)\,dx$$

Нека π^* и π са мощностите съответно върху W^* и W.

$$\pi^* = \mathbf{P}\left(\overrightarrow{\mathrm{X}} \in W^* \mid H_1\right), \qquad \pi = \mathbf{P}\left(\overrightarrow{\mathrm{X}} \in W \mid H_1\right)$$

Ще разгледаме разликата

$$\pi^* - \pi = \int_{W^*} L_1(x)\,dx - \int_W L_1(x)\,dx =$$

$$= \left(\int_A + \int_C\right) L_1(x)\,dx - \left(\int_B + \int_C\right) L_1(x)\,dx = \int_A L_1(x)\,dx - \int_B L_1(x)\,dx$$

$A \subset W^*$ следователно за $\forall x \in A$ е изпълнено $L_1(x) \geq K\,L_0(x)$, аналогично
$B \cap W^* = \emptyset$, тогава от $x \in B$ следва $x \notin W^*$ и съгласно условието на лемата $L_1(x) \leq K\,L_0(x)$, т.е. $-L_1(x) \geq -K\,L_0(x)$. Тези неравенства ни дават възможност да оценим $\pi^* - \pi$

$$\pi^* - \pi \geq \int_A K L_0(x) dx - \int_B K L_0(x) dx = K\left(\int_A L_0(x)\,dx - \int_B L_0(x)\,dx\right) = 0$$

Областта W беше произволно избрана, доказахме, че върху нея мощността е по-малка, отколкото мощността върху W^*. Това означава, че върху W^* мощността достига максимум. ✓

3. Хипотези за очакването на нормално разпределени сл.в.

Лемата на Нейман-Пирсън дава начин за конструиране на оптимална критична област при проверка на хипотези. С нейна помощ ще разгледаме хипотези за математическото очакване на нормално разпределена сл.в.

Нека $X \in N(\mu, \sigma^2)$ като предполагаме, че дисперсията σ^2 е известна. При зададено ниво на съгласие α ще проверим

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu = \mu_1 \end{aligned}$$

Тук μ_0 и μ_1 са известни константи, за определеност ще приемем, че $\mu_0 < \mu_1$. В глава VII (стр.98) изведохме функцията на правдоподобие при нормално разпределени наблюдения.

$$L(\overrightarrow{\mathrm{X}}, \mu) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} e^{-\sum_{k=1}^{n} \frac{(x_k-\mu)^2}{2\sigma^2}}$$

Съгласно лемата трябва да намерим област W^* такава, че

$$\alpha = \mathbf{P}\left(\overrightarrow{\mathrm{X}} \in W^* \mid H_0\right) = \mathbf{P}\left(L_1(x) \geq K\, L_0(x) \mid H_0\right)$$

Където L_0 и L_1 са функциите на правдоподобие, ако са изпълнени съответно H_0 и H_1. Ще заместим в горния израз и ще преобразуваме.

Целта ни е да достигнем до достатъчно прост вид на израза такъв, че да можем да пресметнем вероятността.

$$\alpha = \mathbf{P}\left(\frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} e^{-\sum_{k=1}^{n} \frac{(x_k-\mu_1)^2}{2\sigma^2}} \geq K\, \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} e^{-\sum_{k=1}^{n} \frac{(x_k-\mu_0)^2}{2\sigma^2}} \mid H_0 \right) =$$

$$= \mathbf{P}\left(e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}\left[(x_k-\mu_1)^2-(x_k-\mu_0)^2\right]} \geq K \mid H_0 \right)$$

$$= \mathbf{P}\left(e^{\frac{2(\mu_1-\mu_0)}{2\sigma^2}\sum_{k=1}^{n} x_k} e^{-\frac{n(\mu_1^2-\mu_0^2)}{2\sigma^2}} \geq K \mid H_0 \right) =$$

Втората експонента $e^{-\frac{n(\mu_1^2-\mu_0^2)}{2\sigma^2}}$ е константа от гледна точка на наблюденията x_k. Ще направим тази константа част от K и ще означим новата константа с K_1

$$= \mathbf{P}\left(e^{\frac{2(\mu_1-\mu_0)}{2\sigma^2}\sum_{k=1}^{n} x_k} \geq K_1 \mid H_0 \right) =$$

Ще логаритмуваме двете страни на израза и нека $K_2 = \ln K_1$

$$= \mathbf{P}\left(\frac{2(\mu_1 - \mu_0)}{2\sigma^2}\sum_{k=1}^{n} x_k \geq K_2 \mid H_0\right) =$$

По допускане $\mu_0 < \mu_1$, следователно $\frac{2(\mu_1-\mu_0)}{2\sigma^2}$ е положителна константа, ако разделим двете страни на на нея, то знакът на неравенството няма да се промени. По този начин окончатено получаваме

$$= \mathbf{P}\left(\sum_{k=1}^{n} x_k \geq K_3 \mid H_0\right) = \alpha$$

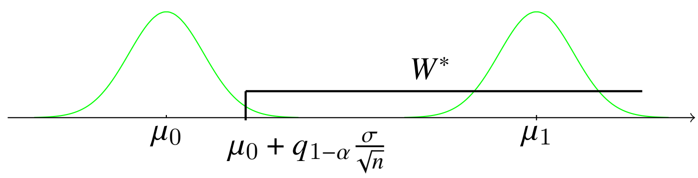
При зададена конкретна стойност на α не е проблем да се определи K_3, достатъчно е да се знае разпределението на $\sum x_k$. При изпълнена хипотеза $H_0 : X_k \in N(\mu_0, \sigma^2)$. В лекция XII показахме, че $\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \in N(0,1)$. Това ни позволява да запишем вероятността по следния начин.

$$\mathbf{P}\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq K_4 \mid H_0\right) = \alpha$$

Константата K_4 се намира от таблица за $N(0,1)$ като $\alpha - 1$ квантил, т.е. $K_4 = q_{1-\alpha}$. Последното събитие, което получихме е еквивалентно на $\overrightarrow{\mathrm{X}} \in W^*$, или казано по друг начин, то задава вида на оптималната критична област

$$W^* = \left\{\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq K_4\right\} = \left\{\overline{X} \geq \mu_0 + q_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right\}$$

Проверявахме хипотези за очакването μ на нормално разпределена сл.в. като имахме две възможности, μ_0 и μ_1, т.е трябва да изберем една от двете плътности показани по-долу в зелено.



При допускането $\mu_0 < \mu_1$ съвсем естествено получихме оптимална критичната област във вид на интервал за средноотоаритметичното на наблюденията. Ако $\overline{X}$ е по-голямо от някаква гранична стойност отхвърляме H_0 и приемаме H_1.

Аналогично, ако $\mu_0 > \mu_1$ тогава критичната област ще е в обратна посока

$$W^* = \left\{\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \leq K_4\right\}$$

В този случай константата $K_4 = q_\alpha = -q_{1-\alpha}$ ще е отрицателна.

$$W^* = \left\{ \overline{X} \le \mu_0 - q_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$$

Пример

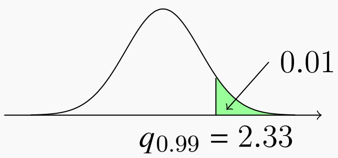
Нека $X \in N(\mu, 1)$. С ниво на съгласие $\alpha = 0.01$ проверяваме

$$\begin{aligned} H_0 &: \mu = 1 \\ H_1 &: \mu = 3 \end{aligned}$$

Ако направените наблюдения са: 1.5, 1.9, 2.9 можем ли да примем за вярна хипотезата H_0?

Пресмятаме $\overline{X} = 2.1$, тази стойност е по близо до 3, отколкото до 1. Дали това е достатъчно да отхвърлим H_0 и да приемем H_1?

В нашия случай критичната област е от типа $W^* = \left\{ \overline{X} \ge \mu_0 + q_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$



От $\mu_0 = 1$, $\sigma = 1$ и $q = 2.33$ за о.к.о. получаваме $W^* = \left\{ \overline{X} \ge 2.35 \right\}$

Но $2.1 \not\ge 2.35$ следователно $\overrightarrow{\mathrm{X}} \notin W^*$, тогава приемаме H_0. Наблюденията са по-големи от 1, но не достатъчно за да отхвърлим H_0 с ниво на съгласие $\alpha = 0.01$. При друго ниво на съгласие изводът би могъл да е различен.

След като разполагаме с критичната област не е проблем да пресметнем мощността на критерия π, т.е. вероятността да отхвърлим хипотезата H_0, ако тя наистина е погрешна. При $\mu_0 < \mu_1$

$$\pi = \mathbf{P}\left(\overrightarrow{\mathrm{X}} \in W^* \mid H_1 \right) = \mathbf{P}\left(\overline{X} \ge \mu_0 + q_{1-\alpha} \frac{\sigma}{\sqrt{n}} \mid H_1 \right)$$

Вероятността трябва да бъде сметната при условие, че е изпълнена алтернативата H_1. Тогава $\overline{X} \in N(\mu_1, \frac{\sigma}{\sqrt{n}})$, следователно $Z = \frac{\overline{X} - \mu_1}{\sigma/\sqrt{n}} \in N(0, 1)$. След съответното стандартизиране получаваме

$$\pi^* = \mathbf{P}\left(Z \ge \frac{\sqrt{n}}{\sigma}(\mu_0 - \mu_1) + q_{1-\alpha} \right)$$

Пример

Ще пресметнем мощността в предишния пример.

$$\pi = \mathbf{P}\left(Z \geq \frac{\sqrt{3}}{1}(1-3) + 2.33\right) = \mathbf{P}\left(Z \geq -1.13\right) = 0.8708$$

Мощността не е голяма, приблизително 87%, т.е. съществува 13% вероятност за грешка от II род.

Увеличаването на броя на наблюденията води до нарастване на мощността на критерия, т.е. повечето наблюдения водят до по-точни критерии. Възможно е да фиксираме мощността π, която искаме да достигнем и при това условие да намерим броя на необходимите наблюдения. За целта трябва да решим спрямо n неравенството

$$\mathbf{P}\left(Z \geq \frac{\sqrt{n}}{\sigma}(\mu_0 - \mu_1) + q_{1-\alpha}\right) \geq \pi$$

Нека $q_{1-\pi}$ е граничната стойност $\mathbf{P}\left(Z \geq q_{1-\pi}\right) = \pi$, тогава решението на неравенството е

$$\frac{\sqrt{n}}{\sigma}(\mu_0 - \mu_1) + q_{1-\alpha} \leq q_{1-\pi}$$

Откъдето не е трудно да се изрази n.

Пример

В примера, който изследвахме ще определим n такова, че мощността на критерия да е по-голяма от 0,95.

$$\pi = \mathbf{P}\left(Z \geq \frac{\sqrt{n}}{1}(1-3) + 2.33\right) \geq 0.95$$

$$\frac{\sqrt{n}}{1}(1-3) + 2.33 \leq q_{0.05} = -q_{0.95} = -1.645$$

Оттук пресмятаме $n = 4$.

Разглеждаме проста хипотеза срещу сложна алтернатива.

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = T_1 \end{aligned}$$

където T_1 е множество, което не съдържа θ_0. Наричаме H_1 сложна, защото не задава единствена стойност, а цяло множество. Тази задача лесно се свежда до предишния случай на проверка на проста хипотеза срещу проста алтернатива. Избираме фиксирано $\theta_1 \in T_1$ и проверяваме

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \end{aligned}$$

По лемата на Нейман Пирсън получаваме съответната оптимална критична област W^*. Ако се окаже, че получената област не зависи от конкретния избор на θ_1, т.е. при всяка възможна стойност получаваме една и съща област, тогава логично тя е областа за проверка на проста хипотеза срещу сложна алтернатива. Ако областта зависи от избора на θ_1, този метод не работи.

Ако се вгледате внимателно в критичната област за проверка на проста срещу проста хипотеза, която изведохме (стр.3.), ще забележите, че ние работихме в общ случай с две константи $\mu_0 < \mu_1$, а критичната област изобщо не зависи от μ_1, нейното значение се загуби. Оказва се, че когато алтернативите са едностранни описаният метод работи и за проверка на "проста срещу сложна" могат да се използват вече изведените критични области.

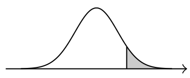
Нека $X \in N(\mu, \sigma^2)$ и основната хипотеза е

$$H_0 \ : \ \mu = \mu_0$$

- При алтернатива

$$H_1 \ : \ \mu > \mu_0$$

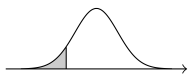
критичната област е $W^* = \left\{\overline{X} \geq \mu_0 + q_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right\}$



- При алтернатива

$$H_1 \ : \ \mu < \mu_0$$

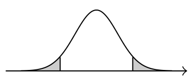
критичната област е $W^* = \left\{\overline{X} \leq \mu_0 - q_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right\}$



- При алтернатива

$$H_1 \ : \ \mu \neq \mu_0$$

критичната област е $W^* = \left\{|\overline{X}| \geq \mu_0 + q_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\}$

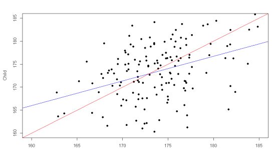


В този случай се използва съвсем резлична теория, на която ние няма да се спираме.

XII . Линейна регресия

Първото ститистическо изследване за намиране на линейна връзка между две променливи, както и едно от първите статистически изследвания изобщо е проведено от Галтон в 1885г. Той е сравнявал височината на бащите и височината на синовете. Първоначалното предположение е било, че по-високите бащи имат и по-високи синове при това пропорционално така, че правата задаваща връзката е ъглополовящата с ъглов коефицент 1.

В действителност се оказва, че по високите бащи имат по високи синове, но не чак толкова, например бащи по високи с 10 см. от средното имат синове само с 6 см. по-високи и аналогично по-ниските бащи имат по-ниски синове, но не чак толкова. Като цяло ръстът на синовете е по-близо до средния, правата е с ъглов коефициент 0.6. Галтон нарича това “регрес към посредствеността”. Така по исторически причини линейната връзка между случайни величини се нарича линейна регресия.



1. Описание на модела

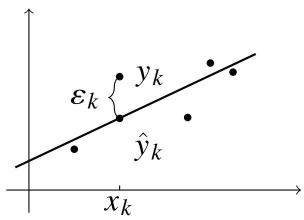
Предполагаме че съществува линейна връзка между променливите X и Y, като евентуално тя е нарушена от някаква грешка ε.

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Прието е X да се нарича независима променлива или предиктор, а Y зависима променлива или отклик. Възможно е X изобщо да не е случайна величина, а нейните стойности да бъдат предварително планувани, например дозата от някакво лекарство давана на пациента.

Разполагаме с наблюдения над X и съответните стойности на Y, т.е. наблюденията са сдвоени (x_k, y_k) за $k = 1, \dots, n$. Ние не знаем истинските стойности на коефициентите β_0 и β_1, целта ни е да намерим оценки за тях по данните с който разполагаме. Нека b_0 и b_1 са сответните оценки, а $\hat{y}_k$ е точката от правата която съответства на x_k. Тогава

$$y_k = b_0 + b_1 x_k + \varepsilon_k, \qquad \hat{y}_k = b_0 + b_1 x_k$$



Нека ε_k е грешката на k-тото наблюдение. Оценките се намират по метода на най-малките квадрати, т.е. избираме такива стойности за b_0 и b_1, при които сумата от квадратите на грешките е минимална.

Прието е със SSR (Sum of Squared Residuals) да се означава сумата на грешките.

$$SSR = \sum_{k=1}^{n} \varepsilon_k^2 = \sum_{k=1}^{n} (y_k - b_0 - b_1 x_k)^2$$

Ще трябва да намерим минимума на тази функция по b_0 и b_1. Ясно е, че функцията е непрекъсната и диференцируема с една единствена особена точка, която е минимум и той ще се достига за

$$\left| \begin{array}{lll} 0 = \frac{\partial SSR}{\partial b_0} = & -2\sum_{k=1}^{n} (y_k - b_0 - b_1 x_k) & = 2\sum_{k=1}^{n} (\hat{y}_k - y_k) \\ 0 = \frac{\partial SSR}{\partial b_1} = & -2\sum_{k=1}^{n} (y_k - b_0 - b_1 x_k)\, x_k & = 2\sum_{k=1}^{n} (\hat{y}_k - y_k)\, x_k \end{array} \right.$$

От първото уравнение на системата получаваме

$$0 = \sum_{k=1}^{n} (y_k - b_0 - b_1 x_k) = \sum_{k=1}^{n} y_k - n b_0 - b_1 \sum_{k=1}^{n} x_k = n\left(\overline{y} - b_0 - b_1 \overline{x}\right)$$

Тук с $\overline{y}$ и $\overline{x}$ сме означили средното аритметично. Следователно

$$\overline{y} = b_0 + b_1 \overline{x} \qquad (\star)$$

Сега ще изразим b_1, за целта ще разгледаме разликата

$$\hat{y_k} - \overline{y} = b_0 + b_1 x_k - b_0 - b_1 \overline{x} = b_1 (x_k - \overline{x})$$

Ще умножим двете страни с $(x_k - \overline{x})$ и ще сумираме по k.

$$b_1 \sum_{k=1}^{n} (x_k - \overline{x})^2 = \sum_{k=1}^{n} (\hat{y_k} - \overline{y})(x_k - \overline{x}) = \sum_{k=1}^{n} \left[(\hat{y_k} - y_k) + (y_k - \overline{y})\right](x_k - \overline{x}) =$$

Ще разделим сумата на части

$$= \sum_{k=1}^{n} (\hat{y_k} - y_k) x_k - \overline{x} \sum_{k=1}^{n} (\hat{y_k} - y_k) + \sum_{k=1}^{n} (y_k - \overline{y})(x_k - \overline{x})$$

Първата сума е нула заради второто уравнение на ситемата, втората сума е нула от първото уравнение на системата. Тогава

$$b_1 \sum_{k=1}^{n} (x_k - \overline{x})^2 = \sum_{k=1}^{n} (y_k - \overline{y})(x_k - \overline{x})$$

От тук и от $(\star)$ елементарно се получават търсените оценки.

Твърдение

Оценките по метод на най-малките квадрати за коефициентите β_0 и β_1 в линейната зависимост

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

са съответно

$$\begin{aligned} b_1 &= \frac{\sum_{k=1}^{n} (y_k - \overline{y})(x_k - \overline{x})}{\sum_{k=1}^{n} (x_k - \overline{x})^2} \\ b_0 &= \overline{y} - b_1 \overline{x} \end{aligned}$$

Както виждаме оценките винаги съществуват. Това все още не означава, че моделът е добър, т.е. че описва данните по подходящ начин, или че между X и Y наистина съществува линейна зависимост. Ще ни трябват критерии, по които да тестваме приложимостта на модела.

Ако разглеждаме наблюденията като случайни величини, каквито те в действителност са, то оценките b_0 и b_1 също са случайни величини. Намирането на техните характеристики ще ни позволи да строим доверителни интервали и да проверяваме хипотези свързани с тях.

Ще въведем някой ограничения на модела.

2. Хипотези за коефициентите

Ако разглеждаме наблюденията като случайни величини, каквито те в действителност са, то оценките b_0 и b_1 също са случайни величини. Намирането на техните характеристики ще ни позволи да строим доверителни интервали и да проверяваме хипотези свързани с тях.

Ще въведем някой ограничения на модела.

❧ Грешките $\varepsilon_1, \dots, \varepsilon_n$ са независими в съвкупност - това означава, че и наблюденията $Y_1, \dots, Y_n$ са независими, което е нормално изискване в статистиката.

❧ $\mathbf{E}\,\varepsilon_k = 0$, $k = 1, \dots, n$ - ако съществува някакво изместване, т.е. има не нулево очакване, то би трябвало да се отрази на коефициента b_0, а не на грешката.

❧ $\mathbf{D}\,\varepsilon_k = \sigma^2$, $k = 1, \dots, n$ - големината на грешките не зависи от $X_1, \dots, X_n$, т.е. очакваме зависимостта между променливите да се описва от модела.

❧ Грешките са нормално разпределени, т.е. $\varepsilon_k \in N(0, \sigma^2)$. Още Гаус при изследването на грешки в астрономически наблюдения е установил, че те са нормално разпределени. Така, че това изискване не е толкова ограничаващо и често се реализира на практика.

Сега ще определим математическото очакване на коефициента b_1. Разглеждаме отклиците $Y_1, \dots, Y_n$ като случайни величини, а предикторите x_k като известни зададени стойности.

$$b_1 = \frac{\sum_{k=1}^n (Y_k - \overline{Y})(x_k - \overline{x})}{\sum_{k=1}^n (x_k - \overline{x})^2}$$

Ще разделим сумата в числителя на две суми

$$b_1 = \frac{\sum_{k=1}^n Y_k(x_k - \overline{x}) - \overline{Y} \sum_{k=1}^n (x_k - \overline{x})}{\sum_{k=1}^n (x_k - \overline{x})^2}$$

Втората сума е нула, тъй като

$$\sum_{k=1}^n (x_k - \overline{x}) = \sum_{k=1}^n x_k - n\overline{x} = \sum_{k=1}^n x_k - \sum_{k=1}^n x_k = 0$$

Тогава можем да разглеждаме b_1 като линейна комбинация на $Y_1, \dots, Y_n$, с коефициенти

$$b_1 = \sum_{k=1}^n \overbrace{\frac{(x_k - \overline{x})}{\sum_{m=1}^n (x_m - \overline{x})^2}}^{v_k} Y_k \;=\; \sum_{k=1}^n v_k\, Y_k$$

Ако сумираме коефициентите v_k, то сумата в числителя както показахме е нула. Тогава

$$\sum_{k=1}^{n} v_k = \frac{\sum_{k=1}^{n}(x_k - \overline{x})}{\sum_{m=1}^{n}(x_m - \overline{x})^2} = 0$$

и освен това

$$\sum_{k=1}^{n} v_k (x_k - \overline{x}) = \frac{\sum_{k=1}^{n}(x_k - \overline{x})^2}{\sum_{m=1}^{n}(x_m - \overline{x})^2} = 1$$

Следователно

$$\sum_{k=1}^{n} v_k x_k = \sum_{k=1}^{n} v_k x_k - \overline{x} \overbrace{\sum_{k=1}^{n} v_k}^{=0} = \sum_{k=1}^{n} v_k (x_k - \overline{x}) = 1$$

Знаем, че $\mathbf{E}\, Y_k = \mathbf{E}\, (\beta_0 + \beta_1 x_k + \varepsilon_k) = \beta_0 + \beta_1 x_k$, така от линейното представяне на b_1 получаваме

$$\mathbf{E}\, b_1 = \sum_{k=1}^{n} v_k\, \mathbf{E}\, Y_k = \beta_0 \overbrace{\sum_{k=1}^{n} v_k}^{0} + \beta_1 \overbrace{\sum_{k=1}^{n} v_k\, x_k}^{1} = \beta_1$$

Това означава, че b_1 е неизместена оценка за β_1.

Отново ще използваме $\mathbf{E}\, Y_k$ за да пресметнем

$$\mathbf{E}\, \overline{Y} = \frac{1}{n} \sum_{k=1}^{n} \mathbf{E}\, Y_k = \frac{1}{n} \sum_{k=1}^{n} (\beta_0 + \beta_1 x_k) = \beta_0 + \beta_1 \overline{x}$$

За математическото очакване на b_0 получаваме

$$\mathbf{E}\, b_0 = \mathbf{E}\, (\overline{Y} - b_1 \overline{x}) = \mathbf{E}\, \overline{Y} - \overline{x}\, \mathbf{E}\, b_1 = \beta_0 + \beta_1 \overline{x} - \overline{x} \beta_1 = \beta_0$$

Това означава, че и оценката b_0 е неизместена.

За да пресметнем дисперсията на b_1 отново ще се върнем към представянето му като линейна комбинация

$$b_1 = \sum_{k=1}^{n} v_k\, Y_k = \sum_{k=1}^{n} v_k\, (\beta_0 + \beta_1 x_k + \varepsilon_k) = \overbrace{\sum_{k=1}^{n} v_k\, (\beta_0 + \beta_1 x_k)}^{Const} + \sum_{k=1}^{n} v_k\, \varepsilon_k$$

Първата сума е равна на константа следователно дисперсията и е нула. Тогава от $\mathbf{D}\, \varepsilon_k = \sigma^2$ следва

$$\mathbf{D}\, b_1 = \sum_{k=1}^{n} \mathbf{D}\, (v_k\, \varepsilon_k) = \sigma^2 \sum_{k=1}^{n} v_k^2 = \sigma^2 \sum_{k=1}^{n} \left[\frac{(x_k - \overline{x})}{\sum_{m=1}^{n}(x_m - \overline{x})^2} \right]^2 =$$

$$= \sigma^2 \frac{\sum_{k=1}^n (x_k - \overline{x})^2}{\left[\sum_{m=1}^n (x_m - \overline{x})^2\right]^2} = \frac{\sigma^2}{\sum_{k=1}^n (x_k - \overline{x})^2}$$

Както показахме по-горе b_1 е линейна комбинация на Y_k, които са нормално разпределени. Следователно b_1 също е нормално разпределена, при това ние изведохме очакването и дисперсията и.

$$b_1 \in N\left(\beta_1, \frac{\sigma^2}{\sum_{k=1}^n (x_k - \overline{x})^2}\right) \qquad (*)$$

Ако използваме стандартното означение S^2 на неизместената оценка за дисперсията, можем да запишем

$$\mathbf{D}\, b_1 = \frac{\sigma^2}{(n-1)S^2}$$

$$b_0 = \overline{Y} - b_1 \overline{x} = \frac{1}{n} \sum_{k=1}^n Y_k - \overline{x} \sum_{k=1}^n v_k\, Y_k = \sum_{k=1}^n \left(\frac{1}{n} - \overline{x}\, v_k\right) Y_k =$$

$$= \sum_{k=1}^n \left(\frac{1}{n} - \overline{x}\, v_k\right) (\beta_0 + \beta_1 x_k + \varepsilon_k) =$$

$$= \underbrace{\sum_{k=1}^n \left(\frac{1}{n} - \overline{x}\, v_k\right) (\beta_0 + \beta_1 x_k)}_{const} + \sum_{k=1}^n \left(\frac{1}{n} - \overline{x}\, v_k\right) \varepsilon_k$$

Първата сума е константа, тогава за дисперсията получаваме

$$\mathbf{D}\, b_0 = \sum_{k=1}^n \left(\frac{1}{n} - \overline{x}\, v_k\right)^2 \mathbf{D}\, \varepsilon_k = \sigma^2 \sum_{k=1}^n \left(\frac{1}{n} - \overline{x}\, v_k\right)^2 =$$

$$= \sigma^2 \left(\sum_{k=1}^n \frac{1}{n^2} - \frac{2\,\overline{x}}{n} \sum_{k=1}^n v_k + \overline{x}^2 \sum_{k=1}^n v_k^2\right) = \sigma^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{k=1}^n (x_k - \overline{x})^2}\right)$$

Статистиката b_0 също е линейна комбинация на Y_k, следователно и тя е нормално разпределена.

$$b_0 \in N\left(\beta_0,\ \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum_{k=1}^n (x_k - \overline{x})^2}\right]\right)$$

$$= \sigma^2 \left(\frac{\sum_{k=1}^n \left(x_k^2 - 2x_k\, \overline{x} + \overline{x}^2\right) + n\, \overline{x}^2}{n \sum_{k=1}^n (x_k - \overline{x})^2}\right) = \sigma^2 \left(\frac{\sum_{k=1}^n x_k^2}{n \sum_{k=1}^n (x_k - \overline{x})^2}\right)$$

Дотук намерихме дисперсията на оценките b_0 и b_1, но тя зависи от дисперсията на грешката σ^2. Ако тя е неизвестна, а в практически задачи най-често е така, ще трябва да оценим σ^2 също от данните.

Съгласно предположенията, които направихме за модела $\varepsilon_k \in N(0, \sigma^2)$. Тогава за Y_k също е случайна величина с нормално разпределение, доколкото β_0, β_1 и x_k са константи, т.е. $Y_k = \beta_0 + \beta_1 x_k + \varepsilon_k \in N(\beta_0 + \beta_1 x_k, \sigma^2)$. Следователно

$$\frac{Y_k - \beta_0 + \beta_1 x_k}{\sigma} \in N(0,1)$$

Както показахме в Лекция VIII сумата от квадратите на тези случайни величини ще има хи-квадрат разпределение с n степени на свобода.

$$\sum_{k=1}^{n} \frac{(Y_k - \beta_0 + \beta_1 x_k)^2}{\sigma^2} \in \chi^2(n)$$

Така можем да намерим разпределението на SSR. Оказва се, че

$$\frac{SSR}{\sigma^2} = \frac{\sum_{k=1}^{n}(Y_k - b_0 + b_1 x_k)^2}{\sigma^2} \in \chi^2(n-2)$$

В SSR два от параметрите , а именно β_0 и β_1 са оценени от данните затова и степените на свобода падат с две. Няма да доказваме този факт формално. Подобно доказателство приведохме за S^2 (Твърдение 2 от Лекция XII).

Знаем, че хи-квадрат разпределението е частен случай на гама

$$X \in \chi^2(n) \qquad \Longleftrightarrow \qquad X \in \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

Така можем да изведем формула за очакването на хи-квадрат, като знаем очакването на гама разпределението, а именно $\mathbf{E}\, X = \frac{n/2}{1/2} = n$. Следователно

$$\mathbf{E}\left(\frac{SSR}{\sigma^2}\right) = n - 2$$

Това означава, че SSR корегирано със съответната константа, може да се разглежда като неизместена оценка за параметъра σ^2

$$\hat{\sigma}^2 = \frac{SSR}{n-2}$$

Ще обърнем внимание, че в SSR участват оценените параметри b_0 и b_1, тогава оценката $\hat{\sigma}^2$ всъщност е зависима от модела и при друг модел естествено ще бъде друга.

Също така е възможно да се използва метода на максимално правдоподобие за да се построи статистика за оценка на параметъра σ^2. Получената в този случай оценка е различна от дадената по-горе, при това тя е и изместена. Затова ще предпочетем $\hat{\sigma}^2$ за по-нататъшните изчисления.

Вече разполагаме с всичко необходимо за да пристъпим към тестване на хипотези свързани с модела.

Най често първата хипотеза, който проверяваме е

$$H_0 \ : \ \beta_1 = 0$$

Тази хипотеза е важна, защото ако $\beta_1 = 0$ то модела се изражда до $Y = \beta_0 + \varepsilon$, което означава, че изобщо не съществува линейна връзка между X и Y, т.е. линейната регресия е безмислена.

Ние ще разгледаме тази хипотеза като частен случай на по общата

$$H_0 \ : \ \beta_1 = b$$
$$H_1 \ : \ \beta_1 \neq b$$

където b е известна константа. Нека α е нивото на значимост.

Както показахме по-горе ($*$) b_1 е нормално разпределена, което ни позволява да конструираме критична област за проверка на хипотезата.

- Ако σ^2 е известна ще използваме статистика

$$Z = \frac{b_1 - b}{\sqrt{\dfrac{\sigma^2}{\sum_{k=1}^{n}(x_k - \overline{x})^2}}}$$

При изпълнена хипотеза H_0, следва $Z \in N(0, 1)$.

Критичната област има вида

$$W = \left\{|Z| \geq q_{\alpha/2}\right\}$$

където $q_{\alpha/2}$ е квантил на $N(0, 1)$.

Идеята тук е, че ако е изпълнена H_0, то стойността на b_1 ще бъде близо до b и съответно статистиката Z ще е около нулата. Тогава, ако Z е прекалено малка или пракалено голяма трябва да отхвърлим H_0 и да приемем H_1. Това обуславя критична зона от типа $|Z| \geq Const$.

Съответно, ако проверяваме същата хипотеза срещу едностранна алтернатива $H_1 \ : \ \beta_1 > b$, ще използваме критична област

$$W = \{Z \geq q_\alpha\}$$

А при едностранна алтернатива $H_1 : \beta_1 < b$, критичната област е

$$W = \{Z \leq q_\alpha\}$$

Статистиката Z може да бъде използвана и като централна статистика (Лекция XII) за построяване на доверителен интервал за β_1. Достатъчно е в нея b да бъде заменено с β_1 и по познатия начин да се изрази β_1. Полученият интервал е

$$I = \left\{b_1 \pm q_{\alpha/2}\sqrt{\frac{\sigma^2}{\sum_{k=1}^{n}(x_k - \overline{x})^2}}\right\}$$

❧ Ако σ^2 е неизвестна използваме подобна на Z статистика, в която заместваме дисперсията с оценката $\hat{\sigma}^2$за нея

$$T = \frac{b_1 - b}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{k=1}^{n}(x_k - \overline{x})^2}}} = \frac{b_1 - b}{\sqrt{\frac{SSR}{(n-2)\sum_{k=1}^{n}(x_k - \overline{x})^2}}}$$

Не е трудно да се съобрази, че

$$T = \frac{Z}{\sqrt{\frac{SSR}{(n-2)\sigma^2}}}$$

С подобни статистики работихме при конструирането на доверителен интервал за очакването на нормално разпределение (Лекция XII). Знаем, че $Z \in N(0,1)$ и $\frac{SSR}{\sigma^2} \in \chi^2(n-2)$ следователно T има разпределение на Стюдънт с $n-2$ степени на свобода, т.е. $T \in t(n-2)$. Критичните области се конструират както в случая с известна диспресия, единствената разлика е, че квантилите се взимат от таблици на Стюдънт.

$$H_1 \ : \ \beta_1 \neq b \qquad H_1 \ : \ \beta_1 > b \qquad H_1 \ : \ \beta_1 < b$$

$$W = \left\{|Z| \geq q_{\alpha/2}\right\} \qquad W = \{Z \geq q_\alpha\} \qquad W = \{Z \leq q_\alpha\}$$

Аналогично на предходния случай T може да се използва за построяване на доверителен интервал за b_1.

Проверка на хипотези за b_0. Както доказахме по-горе

$$b_0 \ \in \ N\left(\beta_0,\ \sigma^2\left[\frac{1}{n} + \frac{\overline{x}^2}{\sum_{k=1}^{n}(x_k - \overline{x})^2}\right]\right)$$

Ако е изпълнена хипотеза $H_0 \ : \ \beta_0 = b$, ще знаем разпределенията на Z и T съответно за случайте на известна и неизвестна дисперсия.

$$Z = \frac{b_0 - b}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{k=1}^{n}(x_k - \overline{x})^2}\right)}} \qquad \in \ N(0,1)$$

$$T = \frac{b_0 - b}{\sqrt{\frac{SSR}{n-2}\left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{k=1}^{n}(x_k - \overline{x})^2}\right)}} \qquad \in \ t(n-2)$$

Критичните области за проверка на хипотези и доверителните интервали се построяват аналогично на тези за b_1. Ще оставим любознателния читател да довърши подробностите.

XIII . Примерна Глава

1. Секция

Просто текст някакъв ....

И още....

Кутия без граници

Използвам ги за примери и т.н.

Текст в полето се прави с \margintext{ }

И още....

Кутия СЪС граници

Използвам ги за теореми, дефиниции и т.н.

$$\sum x = \mathbf{P}(A), \mathbf{E}\,X, \mathbf{D}\,X$$

Математическите символи:
$\mathbf{P}(A)$ = \P
$\mathbf{E}\,X$ = \E
$\mathbf{D}\,X$ = \D
$A \perp\!\!\!\perp B$ = \indep
Край на доказателството ✓ = \yes

Текстът в полето може да се измества вертикално с \margintext[2cm]{ }

За всяка глава създавам отделна директория, секциите са отделни файлове.

Основният файл е BooK.tex

Всички команди са в BooKstyle.tex