**Лекция 1**
**Случайна Величина**

Химията е наука, която се занимава с преобразуването и строежа на веществото. Основни понятия в химията са химичните и физикохимичните величини, чието измерване е една от главните задачи на химичния експеримент. Понятието величина е удобна научна фикция, служеща успешно в практиката на човека, въпреки че в абсолютен смисъл величината няма стойност. Едно по-точно измерване на концентрацията на даден хомогенен разтвор би показало, че тя се изменя по място и по време, вследствие на случайните флуктуации на разтвореното вещество. Аналогично, масата на определено количество вещество се променя непрекъснато от влагата и от изпарението на веществото. Но като понятие величината предполага една стойност, наречена ***истинска стойност*** на величината, която експериментаторът се стреми да определи. В неговите възможности е само ***оценка*** на тази стойност посредством набор от опитни стойности и то ***приближена оценка***, която той дава с определена вероятност за нейната истинност.

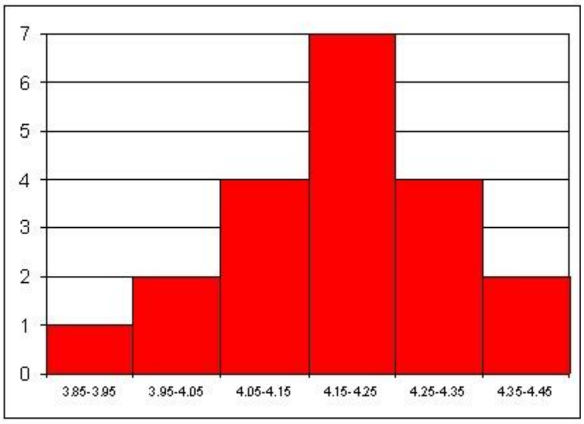
**1.1. Представа за случайна величина.** Опитът показва, че при повтаряне на дадено измерване, се получават стойности на величината, които могат да се различават една от друга, независимо от постоянството на експерименталните условия. Тези стойности не могат да се предскажат преди опита. Това са така наречените случайни величини. За разлика от тях, неслучайните величини имат точно определена стойност, която може да се предскаже преди опита или получаваните в резултат на измерванията стойности са равни една на друга. Различават се ***дискретни*** и ***непрекъснати*** случайни величини. Когато възможните значения на една случайна величина са краен или *изброимо безкраен брой*, тя е дискретна. В повечето случаи при химичните измервания стойностите на случайната величина могат да се разполагат в даден интервал, като заемат коя да е реална числена стойност в него (например концентрацията на едно вещество е в интервала 0 - 100%). Това са така наречените ***непрекъснати случайни величини***.

При измерването на дадена величина се забелязва, че получените стойности (случайни величини) са групирани по определен начин около дадена стойност. В таблица 1.1 са представени резултатите от двадесет измервания на съдържанието на цинк в питейна вода.

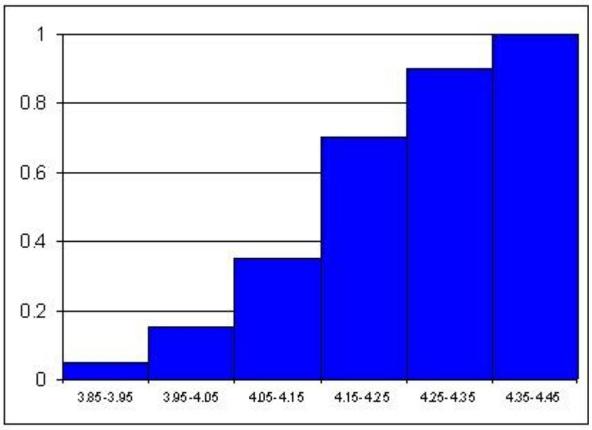
**Таблица 1.1.** Съдържание на цинк в питейна вода.

| No на измерването | Съдържание на цинк (μg/ml) | No на измерването | Съдържание на цинк (μg/ml) |
|---|---|---|---|
| 1 | 4.23 | 11 | 4.11 |
| 2 | 3.97 | 12 | 4.05 |
| 3 | 4.18 | 13 | 4.27 |
| 4 | 4.29 | 14 | 4.43 |
| 5 | 4.00 | 15 | 4.31 |
| 6 | 4.17 | 16 | 4.15 |
| 7 | 4.12 | 17 | 4.24 |
| 8 | 4.08 | 18 | 4.18 |
| 9 | 4.20 | 19 | 4.25 |
| 10 | 3.88 | 20 | 4.35 |

За да се добие представа за характера на получаваните случайни величини, може да се извърши следното: най-малката стойност е 3.88, а най-голямата е 4.43, т.е. всички стойности са в интервала (3.85, 4.45). Този интервал се разделя на шест интервала с дължина на всеки равна на 0.10 и се построява хистограмата, показана на фигура 1.1. Височината на всеки от правоъгълниците е равна на броя на резултатите, намиращи се в съответния интервал, включително и тези равни на дясната граница на интервала.

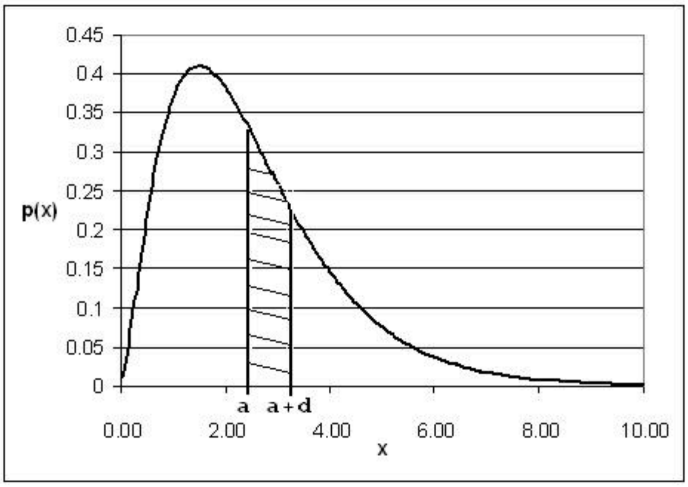



**Фигура 1.1.** Хистограма на резултатите от таблица 1.1




**Фигура 1.2.** Емпирична функция на разпределение на резултатите от таблица 1.1

Прави впечатление, че в интервала (4.15, 4.25), който е разположен почти в средата, има най-много стойности (седем) и че с отдалечаване от него на другите интервали в тях намалява броят на експерименталните стойности. Това е обща закономерност на получаваните резултати при почти всички използувани методи за измерване. Същата хистограма е по-правилно да се построи, като височината на стълбовете се нормира, т.е. се раздели на общия брой резултати. Така получените числа 1/20, 2/20, 4/20, 7/20, 4/20 и 2/20 се наричат ***относителни честоти на поява на случайните величини*** в съответните интервали. Може да се построи и друга хистограма, в която височината на даден стълб, се определя като сума от височините на всички стълбове, разположени в ляво от дясната граница на съответния интервал. На фигура 1.2 е представена тази хистограма, като височините на стълбовете са нормирани, т.е. сумирани са нормираните височини от хистограмата на фигура 1.1. Втората хистограма показва честота на появата на случайна величина по-малка от дадено число (дясната граница на интервала) и се нарича ***емпирична функция на разпределение на случайната величина***. При увеличаване на броя на измерванията относителните честоти, както на първата хистограма, така и на втората, се стабилизират, което означава, че при провеждане на все по-голям брой опити те клонят във вероятностен смисъл към определени числа, наречени вероятности. Практически това означава, че при нарастване на броя на измерванията и намаляване на големината на интервалите, височините на правоъгълниците на хистограмите от фигури 1.1 и 1.2 клонят към определени стойности, разположени на непрекъснати криви. Хистограмата от фигура 1.1 (ако се нормира, тя тогава представлява относителните честоти на поява на случайната величина) клони към крива, наречена ***плътност на (функцията на) разпределение на случайната величина***, а тази от фигура 1.2 към ***функция на разпределение на случайната величина***. Функцията на разпределение напълно характеризира случайната величина: тя показва каква е вероятността случайната величина да бъде по-малка от дадена стойност. За непрекъснатите случайни величини тя е нарастваща, непрекъсната и диференцируема функция и нейната първа производна е плътността на разпределение.




**Фигура 1.3.** Плътност на разпределение **p(x)** на произволна случайна величина **x**.

На фигура 1.3 е показана примерна крива на плътност на (функцията на) разпределение **y = p(x)**. Тя е неотрицателна функция. Защрихованата площ между правите **x = a** и **x = a + d** е вероятността случайната величина да заема стойности между **a** и **a + d** и при малък интервал **d** тя може да се апроксимира с площта на правоъгълник с ширина **d** и височина **p(a)**. Вероятността случайната величина **X** да е в произволен интервал **(a,b)** се дава с интеграла

$$p(a \le X < b) = \int_a^b p(x)dx$$

Вероятността случайната величина да е в интервала ( -∞,+∞) е единица и затова плътността на разпределение трябва да е нормирана (т.е. умножена с подходяща константа), така че

$$p(-\infty \le X < \infty) = \int_{-\infty}^{\infty} Cp(x)dx = 1$$

Функцията на разпределение на непрекъсната случайна величина се дава с интеграла:

$$F(x) = p(X < x) = \int_{-\infty}^{x} p(x)dx$$

Дискретната случайна величина може напълно да се зададе със своя вероятностен ред, който е подобен на плътността на разпределение и който дава вероятностите за получаване на съответните стойности. Таблица 1.2 показва вероятностния ред на случайната величина "сума от точките на две хвърлени зарчета".

**Таблица 1.2.** Вероятностен ред на случайната величина "сума от точките на две хвърлени зарчета". $X_k$ - стойност на случайната величина, $p_k$ - вероятност за нейното заемане.

| $X_k$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_k$ | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 6/36 | 5/36 | 4/36 | 3/36 | 2/36 | 1/36 |

Функцията на разпределение на същата случайна величина може да се даде таблично. По аналогия с тази на непрекъснатата, тя се получава по формулата:

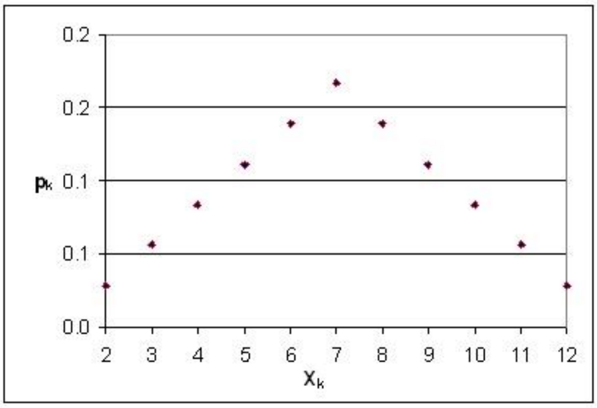
$$F(X_i) = \sum_{k=1}^{i} p_k$$

Тя е представена в таблица 1.3 и числата от долния ред показват вероятностите случайните величини да заемат стойности по-малки или равни от стойностите в горния ред.

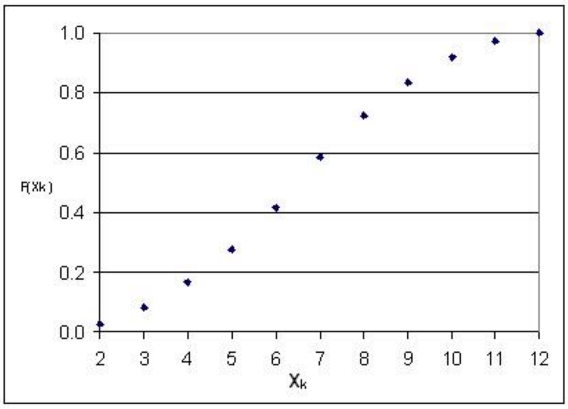
**Таблица 1.3.** Функция на разпределение на случайната величина "сума от точките на две хвърлени зарчета". $X_k$ - стойности на дискретната случайна величина; $F(X_k)$ - нейната функция на разпределение.

| $X_k$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $F(X_k)$ | 1/36 | 3/36 | 6/36 | 10/36 | 15/36 | 21/36 | 26/36 | 30/36 | 33/36 | 35/36 | 3636 |

Данните от таблици 1.2 и 1.3 са изобразени на фигури 1.4. и 1.5.




**Фигура 1.4.** Плътност на разпределение на дискретната случайна величина таблица 1.2.




**Фигура 1.5.** Функция на разпределение на случайната величина от таблица 1.2.

**1.2. Числови характеристики на случайните величини.** Функцията на разпределение напълно характеризира случайната величина, но в повечето случаи не е нужен нейният вид. Достатъчно е да се даде оценка за стойността на измерваната величина и да се посочи с определена вероятност интервалът, в който тя се намира. Поради тези причини важна роля играят две числови характеристики на разпределението на случайната величина: ***математическото очакване*** и ***дисперсията***.

Математическо очакване на *дискретна* случайна величина се дефинира като сумата:

$$M(X) = \sum_{k=1}^{\infty} p(X_k) X_k \qquad \textbf{(1.1)}$$

а за *непрекъсната* случайна величина чрез интеграла:

$$M(X) = \int_{-\infty}^{\infty} x p(x) dx \qquad \textbf{(1.2)}$$

Дисперсията на случайна величина се дефинира като математическо очакване от квадрата на разликата между случайната величина и нейното математическо очакване.

$$\mathbf{D(X) = M[X - M(X)]^2}$$

Или използвайки формули (1.1) и (1.2) получаваме:

- за дискретна случайна величина

$$D(X) = \sum_{k=1}^{\infty} p(X_k)[X_k - M(X)]^2$$

за непрекъсната случайна величина

$$D(X) = \int_{-\infty}^{\infty} [x - M(X)]^2 p(x) dx$$

Обърнете внимание, че **M(X)** и **D(X)** са числа, а не някакви функции. Те просто се означават със скоби, за да се покаже на коя случайна величина са числови характеристики, така както е направено по-долу при представяне на седемте свойства.

Плътността на вероятността **p(X)**, съответно вероятностите **p($X_к$)**, са неотрицателни величини и поради това дисперсията е неотрицателно число. Математическото очакване може да бъде, както положително, така и отрицателно число. Неговата стойност, както се вижда от (1.1) и (1.2), се определя в най-голяма степен от най-вероятните стойности. Физическият смисъл на дисперсията е, че тя е мярка за разсейването на случайните величини около математическото очакване.

**1.3. Свойства на числовите характеристики на случайните величини.** Математическото очакване и дисперсията са числа, които характеризират дадена случайна величина. Сумата, разликата, произведението и частното от случайни величи е отново случайна величина. За някои от тези действия могат да се изведат зависимости между числените характеристики на резултантната случайна величина и тези на оригиналните. Тези зависимости, дадени по-долу, не са аксиоми, а могат да се изведат от дефиницията на математическото очакване и дисперсията (вижте задачи 1.1 – 1.5).

Математическото очакване и дисперсия имат следните свойства:

1. Математическото очакване на *неслучайна* величина (*константа*) е равно на неслучайната величина (нейната единствена стойност).

$$M(C) = C$$

2. Неслучайната величина (константа) може да се изнася извън знака на математическото очакване.

$$\mathbf{M(CX) = CM(X)}$$

3. Математическото очакване на сума от случайни величини е равно на сума от математическите им очаквания.

$$\mathbf{M(X + У + \ldots + Z) = M(X) + M(У) + \ldots + M(Z)}$$

3а. Прилагайки това свойство и предишното получаваме

$$\mathbf{M(X - У) = M[X + (-1)У] = M(X) + M[(-1)У] = M(X) + (-1)M(У) = M(X) - M(У),}$$

т.е.

$$\mathbf{M(X - У) = M(X) - M(У),}$$

4. Математическото очакване на произведение на ***независими*** случайни величини е равно на произведение на математическите им очаквания.

$$\mathbf{M(XУ) = M(X)M(У)}$$

5. Дисперсията на неслучайна величина (константа) е нула.

$$\mathbf{D(C) = 0}$$

6. Неслучайната величина може да се изнася извън знака на дисперсията. Изнесената неслучайна величина се повдига на квадрат.

$$\mathbf{D(CX) = C^2D(X)}$$

7. Дисперсията на сума от ***независими*** случайни величини е равна на сума от техните дисперсии.

$$\mathbf{D(X + У + \ldots + Z) = D(X) + D(У) + \ldots + D(Z)}$$

7а. Прилагайки това свойство и предишното получаваме

$$\mathbf{D(X - У) = D[X + (-1)У] = D(X) + D[(-1)У] = D(X) + (-1)^2D(У) = D(X) + D(У)}$$

т.е.

$$\mathbf{D(X - У) = D(X) + D(У)}$$

Обърнете внимание, че няма формула за дисперсия на произведение от случайни величи!

Използувайки свойства 1 - 7, всяка една случайна величина може да се нормира по формулата.

$$Y = \frac{X - M(X)}{\sqrt{D(X)}} \qquad \textbf{(1.3)}$$

при което

$$M(У) = 0 \text{ и } D(У) = 1 \qquad (1.4)$$

**Задача 1.1.** Като използвате дефиницията за математическо очакване на *дискретна* случайна величина докажете свойство 1. ***Упътване:*** фактически константата ***C*** е дискретна случайна величина, която заема само една стойност, $X_1 = C$ с вероятност $p_1 = 1$.

**Задача 1.2.** Като използвате дистрибутивното свойство на умножението (т.е. линейността на една сума) и дефиницията за математическо очакване на дискретна случайна величина докажете свойства 2 и 3 *за дискретна случайна величина.*

**Задача 1.3.** Като използвате линейността на определения интеграл и дефиницията за математическо очакване на *непрекъсната* случайна величина докажете свойства 2 и 3 *за непрекъсната случайна величина.* От статистиката е известно, че **p(Cx) = p(x)** и **p(x+y)** е **p(x)** за случайната величина **X** и **p(y)** за случайна величина **У**.

**Задача 1.4.** Като използвате дефинициите за математическо очакване и дисперсия на случайна величина докажете свойства 5 и 6.

**Задача 1.5.** Докажете формулите в (1.4) с помощта на свойства 1 - 7!

**1.4. Равномерно разпределение.** Това е такова разпределение, за което плътността на вероятността $p(X)$ е постоянна и различна от нула в даден интервал и равна на нула извън него.

$$p(X) = \begin{cases} C \text{ за } a <= X <= b \\ 0 \text{ за } X < a \text{ и } b > X \end{cases}$$

Вероятността случайната величина да заема стойности в интервала $(-\infty, +\infty)$ е единица,

$$p(-\infty < X < \infty) = \int_{-\infty}^{\infty} p(x)dx = \int_{-\infty}^{a} 0dx + \int_{a}^{b} Cdx + \int_{b}^{\infty} 0dx = C(b-a) = 1$$

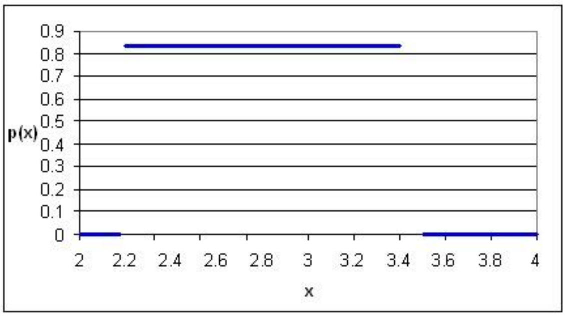
като първият и третият интеграл са нули поради равенство на подинтегралната функция на нула. При интегрирането се получава

$$C(b - a) = 1, \text{ т.е. } C = 1/(b - a).$$

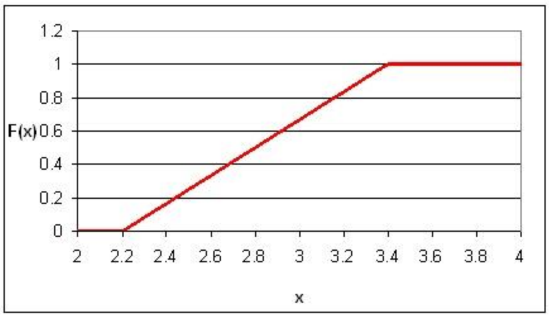
Функцията на разпределение е интеграл от плътността на вероятността и е равна на:

$$F(X) = \begin{cases} 0, \text{ за } X < a \\ (X - a)/(b - a), \text{ за } a \le X \le b \\ 1, \text{ за } X > b \end{cases}$$

На фигури 1.6. и 1.7. са изобразени двете функции $p(X)$ и $F(X)$.



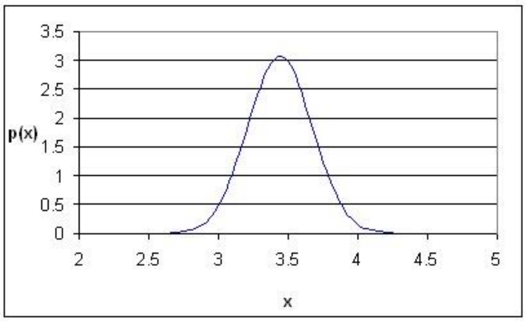

**Фигура 1.6.** Плътност на вероятността на равномерно разпределение с $a = 2.2$ и $b = 3.4$.



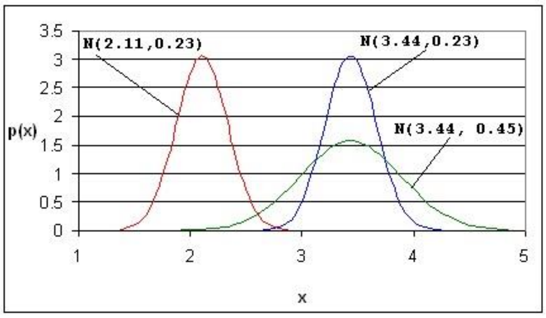

**Фигура 1.7.** Графично представяне на функцията на равномерно разпределение с $a = 2.2$ и $b = 3.4$.

**1.5. Нормално разпределение.** Непрекъсната случайна величина е нормално разпределена и тя се нарича нормална, ако плътността и на разпределение има вида:

$$p(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(X-\mu)^2}{2\sigma^2}} \quad (1.5)$$




**Фигура 1.8.** Плътност на нормалното разпределение с $\mu = 3.44$ и $\sigma = 0.23$.




**Фигура 1.9.** Плътност на нормалото разпределение за различни стойности на $\mu$ и $\sigma$; $\mu = 2.11$ и $\sigma = 0.23$; $\mu = 3.44$ и $\sigma = 0.23$; $\mu = 3.44$ и $\sigma = 0.45$.

Другото име на това разпределение е ***Гаусово разпределение***. Дадено нормално разпределение се означава със записа $N(\mu,\sigma)$, където $\mu$ и $\sigma$ са така наречените параметри на разпределението. Графичният вид на тази функция (вижте фигура 1.8) е камбановидна крива, симетрична относно правата $X = \mu$, с максимум за $X = \mu$ и с инфлексни точки за $X = \mu - \sigma$ и $X = \mu + \sigma$. При $X = \mu + 3\sigma$ нейната стойност е почти нула в сравнение с максимума: $p(X) = 0.011p(\mu)$. Коефициентът пред експонента осигурява нормираността на плътността на разпределението, т.е. вероятността за намиране на случайната величина в интервала $(-\infty, +\infty)$ е единица (вижте задача 6):

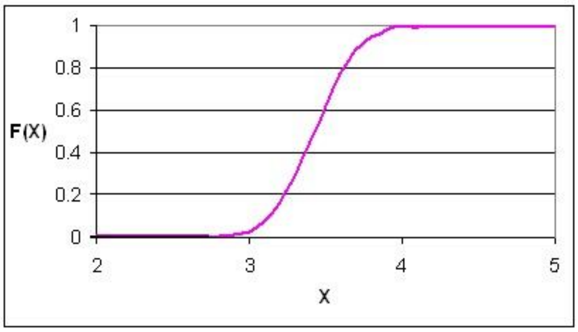
$$\int_{-\infty}^{\infty} p(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(X-\mu)^2}{2\sigma^2}} dx = 1$$

Промяната на стойността на $\mu$ измества кривата по координата $X$ без да променя наклона и, а изменението на $\sigma$ променя само нейния наклон. При по-големи стойности на $\sigma$ кривата е по-полегата и с по-нисък максимум, вижте фигура 1.9.

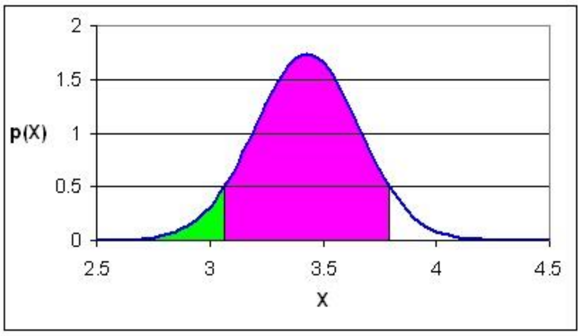
Функцията на разпределение се дава с интеграла,

$$F(X) = \int_{-\infty}^{X} p(x)dx = \int_{-\infty}^{X} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(X-\mu)^2}{2\sigma^2}} dx$$

който се нарича ***интеграл на Лаплас*** и е нерешим в явен вид. Графичният вид на тази функция е представен на фигура 1.10. Математическото очакване на нормалното разпределение е равно на $\mu$, а дисперсията на $\sigma^2$ - вижте задача 2 и нейното решение.




**Фигура 1.10.** Функция на разпределение на нормална случайна величина с $\mu = 3.44$ и $\sigma = 0.23$.




**Фигура 1.11.** Интеграли на Лаплас; F(X) в граници от $-\infty$ до $+X$ (двете оцветени площи) и F'(X) в граници от $-X$ до $+X$ (виолетово оцветената площ).

Ако в (1.5) се смени променливата $X$ с $Y=(X-\mu)/\sigma$, се получава нормирано нормално разпределение или така нареченото ***стандартно разпределение***. То има следната плътност на разпределение:

$$p(X) = \frac{1}{\sqrt{2\pi}} e^{\frac{-X^2}{2}}$$

Според формула (1.4) от математическото очакване и дисперсията на стандартното разпределение са съответно 0 и 1 - вижте също и задача 1.7.

Както видяхме стандартното разпределение е нормално разпределение с параметри $\mu$=0 и $\sigma^2$=1. В приложение 1 са дадени числените решения **F(X)** на интеграла на Лаплас за стандартното разпределение в граници от $-\infty$ до **X**. Често е необходимо намирането на негово решение **F'(X)** в граници от **–X** до **+X**. Между двете стойности съществува връзката **F'(X) = 2F(X) - 1**, която е следствие от симетричността на плътността на разпределение, вижте фигура 1.11. Двата интеграла от приложението **F'(X)** и **F(X)** могат да се използват за произволно нормално разпределение, като е необходимо границите на интегриране за последното са изразени в единици $\sigma$. Например за $\mu$ = **3.44** и $\sigma$ = **0.23** границата на интегриране **4.00** в единици $\sigma$ е равна на **(4.00–3.44)/0.23 = 2.44**.

**Задача 1.6.** Вероятността случайната величина да заема стойности в интервала $(-\infty, +\infty)$ е единица. Докажете това за нормалното разпределение - съответният интеграл няма решение, но двойният интеграл по-долу може да се реши със замяна на променливите **x** и **y** с радиални координати $\rho$ и $\theta$.

$$\int_0^\infty\int_0^\infty e^{-(x^2+y^2)}dxy = \int_0^\infty\int_0^\infty e^{-x^2}e^{-y^2}dxy = \int_0^\infty e^{-x^2}dx\int_0^\infty e^{-y^2}dy = (\int_0^\infty e^{-y^2}dy)^2$$

**Задача 1.7.** Намерете математическото очакване и дисперсията на равномерното разпределение.

**Задача 1.8.** Докажете, че математическото очакване и дисперсията на нормалното разпределение са съответно $\mu$ и $\sigma^2$.

**Задача 1.9.** За стандартното разпределение докажете, че **F'(X) = 2F(X) - 1**, където **F(X)** е интеграл в граници от $-\infty$ до **X**, а **F'(X)** - интеграл в граници от **–X** до **+X**. Използвате симетричността на Гаусовата крива.

**Задача 1.10.** За стандартното разпределение намерете от приложение 1 **F'(2.87)** и **F(2.87)**, където **F(2.87)** е интеграл в граници от $-\infty$ до **2.87**, а **F'(2.87)** - интеграл в граници от **–2.87** до **+2.87**. Използвайте, че **F'(X) = 2F(X) - 1**.

**Задача 1.11. Задача 6.** За нормално разпределение с $\mu = 3.44$ и $\sigma = 0.39$ намерете от приложение 1 **F(4.56)**, където **F(4.56)** е интеграл в граници от $-\infty$ до 4.56.

**Задача 1.12.** За нормално разпределение с $\mu = 3.44$ и $\sigma = 0.39$ намерете от приложение 1 интеграл в граници от 2.32 до 4.56. Използвайте, че **F'(X) = 2F(X) - 1**.

**Задача 1.13*.** В теория на вероятностите две събиния са независими, ако вероятността да се случат и двете е равна на произведение на вероятностите да се случат събитията поотделно: P(X и Y) = P(X)P(Y). В статистиката плътността на разпределение на произведението на две независими събития се дава с p(xy) = p(x)p(y).

Като използвате линейността на определения интеграл и дефиницията за математическо очакване на непрекъсната случайна величина докажете свойство 4 за непрекъсната случайна величина.

## Литература

1. Л. Футеков, П. Пенчев, ***Теория на експеримента***. Издателство на Пловдивския университет, Пловдив, 1992, 1999.