

ВЪЗМОЖНОСТИТЕ НА МАТЛАВ, В ПОМОЩ НА ПОДГОТОВКАТА НА ЛЕКЦИИТЕ И УПРАЖНЕНИЯТА ПО СТАТИСТИКА

Павлина Йорданова
pavlina_kj@abv.bg

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА,
ШУ "ЕПИСКОП КОНСТАНТИН ПРЕСЛАВСКИ"

В нашия век на почти неограничени компютърни възможности и все по-нарастваща нужда от усвояване на все повече и повече информация за по-кратко време, използването на съвременен софтуер и подходящи нагледни материали (графики и таблици) в обучението по Вероятности и Статистика става все по-необходимо. Поднасянето на теорията, след кратък пример за съответния практически проблем, който тя решава, мотивира студентите и ги кара сами да търсят връзки и зависимости между изучаваните понятия. Визуализирането на почти реални (симулирани) данни им дава много по-щирок поглед върху дискутирания проблем от този, който бихме им дали ако използваме само словесно описание. Това позволява скъсяване на времето за превръщане на информацията от лекцията в практически полезно за тях знание.

В тази разработка споделям очарованието си от възможностите на MATLAB, които можем да използваме при подготовката на лекциите и упражненията по Статистика.

Основни цели на разработката

1. Да илюстрира колко лесно могат да се получат полезни за работата ни статистически графики или да се реализират изучаваните статистически методи;
2. Да покаже как могат да се онагледят част от изучаваните теоретични понятия, гранични теореми и статистически методи;
3. Да демонстрира по какъв начин можем да провокираме изследователския дух на студентите и да ги накараме сами, с евристични разсъждения да достигнат до изучаваните резултати.

Предполага се, че лекциите и упражненията от основния курс по Статистика са оформени като слайдове и тук са дадени допълнителни нагледни материали, изработени на MATLAB, които могат да бъдат добавени към тях.

Защо точно MATLAB? Виж (2).

Най-често използвани разпределения

Вградени функции за

- Симулиране на реализации на случайни величини;
- Изчертаване на графиката на емпиричната функция на разпределение;
- Изчертаване на хистограма на абсолютните или относителните емпирични честоти на разпределение;
- Графичен прозорец за изчертаване на графиките на теоретичните функции на разпределение, плътности на разпределение или визуализиране на вероятностите от различните редове на разпределение;
- Пресмятане на числовите характеристики.

При различни стойности на параметрите.

- Равномерни (дискретно и абсолютно непрекъснато);
- Биномно;
- Отрицателно биномно (Геометрично);
- Хипергеометрично;
- Поасоново (Simeon Poisson, 1837);
- Гама (експоненциално и χ^2);
- Логаритмично нормално;
- Fisher;
- t, (W. S. Gossett, 1908);
- Бета;
- Weibull, (Waloddi Weibull, 1939).

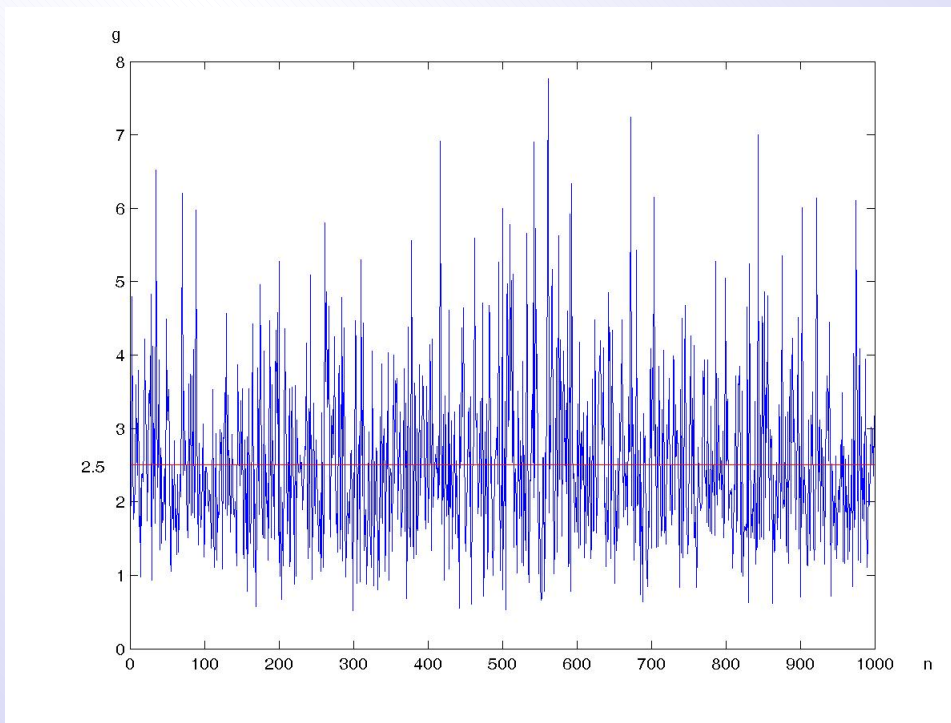
Виж (1).

rstab.m е адаптирана от доц.д-р Димитър Върндев

- α -устойчиво.

(1)

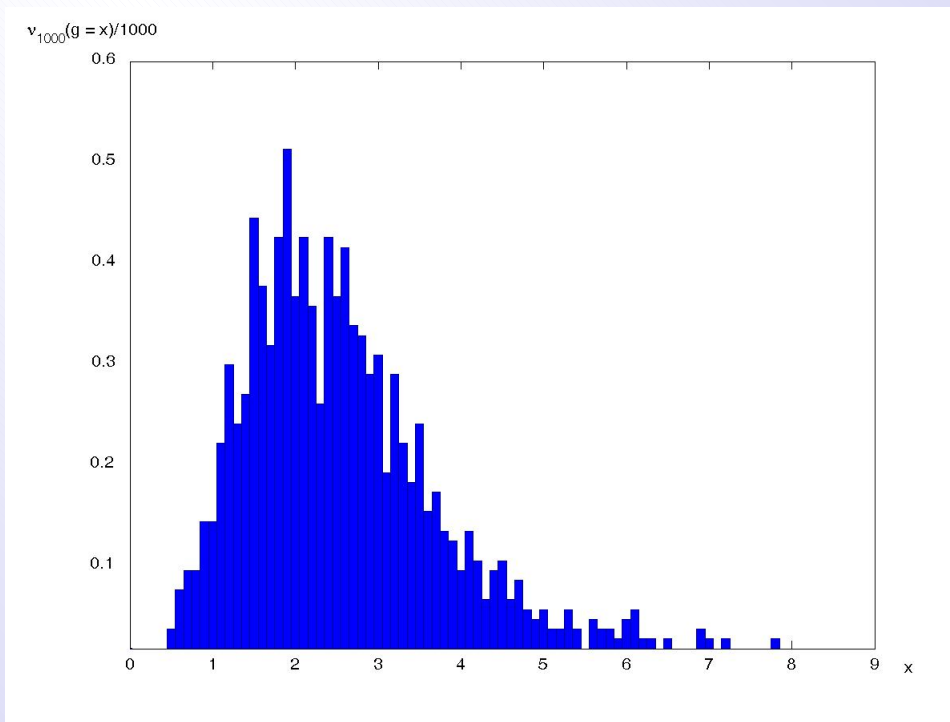
Пример 1: Проведени са 1000 независими наблюдения над $\eta \sim \Gamma(5, 0.5)$. Резултатите от наблюдението са изобразени на Фиг 1.



Фиг. 1

За получаването на тази графика виж (6).

Извършена е групировка, с $w = 0.1$ и резултатите от групировката са изобразени на следната хистограма на относителните честоти.



Фиг. 2

Виж техническото изпълнение в (7).

- $\eta \sim \Gamma(\alpha, \beta)$;

$$P_{\eta}(x) = \begin{cases} \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)} & , x \geq 0 \\ 0 & , x < 0 \end{cases} .$$

- Как влияе изменението на α на графиката на плътността?
Виж Фиг. 3.
- Как влияе изменението на β на графиката на плътността?
Виж Фиг. 4.

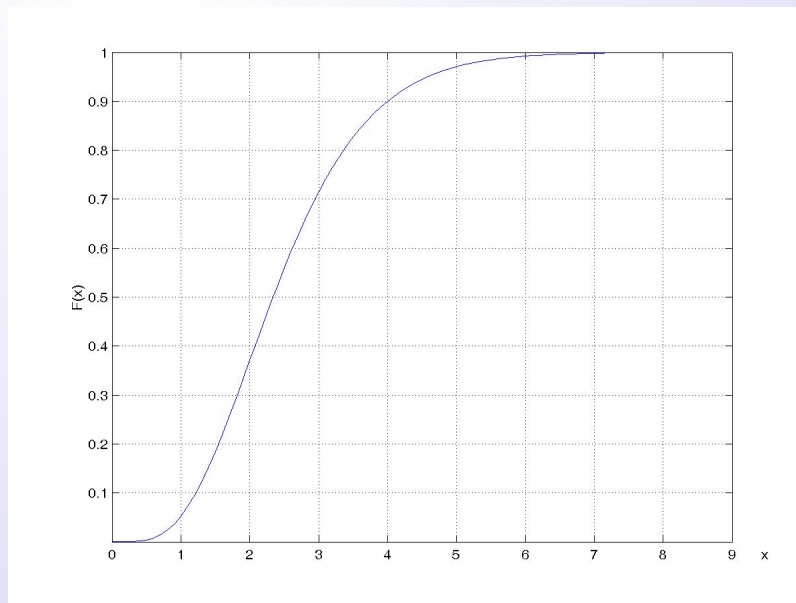
Тези графики могат да бъдат получени посредством вградената функция *disttool*, която създава интерактивни графики на вероятностни разпределения.

Графичен прозорец за работа с вероятностни разпределения

Виж Фиг. 5.

- $\eta \sim \Gamma(\alpha, \beta)$

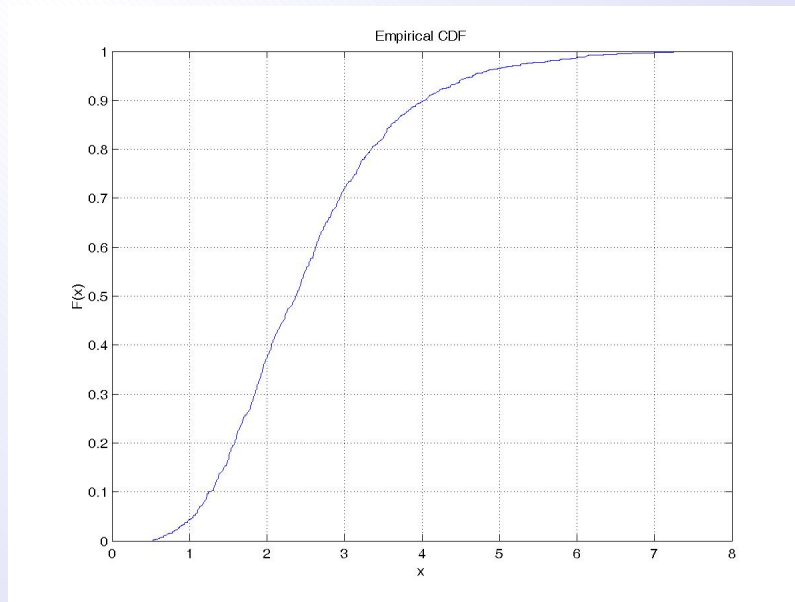
$$F_{\eta}(x) = \begin{cases} 0 & , x \leq 0 \\ \int_0^x \frac{y^{\alpha-1} e^{-y/\beta}}{\beta^{\alpha} \Gamma(\alpha)} dy & , x > 0 \end{cases} .$$



Фиг. 7 Функцията на разпределение на $\Gamma(5, 0.5)$.

Графична интерпретация на връзката между ф.р., плътност на разпределение и квантили. Виж Фиг. 8.

Определени са кумулативните относителни честоти и резултатите са дадени на следната графика на е.ф.р.



Фиг. 6

След генерирането на вектора g за получаването на тази графика използваме само

`cdfplot(g)`

Най-често използвани гранични теореми

• Теорема на Поасон (1837); Виж 28.

• Близост между $t(n)$ и $N(0, 1)$ за $n \geq 30$.

Виж 29.

За онагледяването на резултатите, получени в следващите теореми няма вградени `.m` функции в MATLAB, но с минимални познания по скрипта на `.m` файловете не е сложно да се направят.

• Теорема на Бернули; Виж 30.

• Теорема на Моавър-Лаплас; Виж 31.

• Теорема на Колмогоров; Виж 32.

• Централна гранична теорема за н.е.р. с крайна дисперсия.

Виж 33.

Виж 34.

Визуализиране на съществени моменти от някои статистически методи

- χ^2 -критерий на Пирсън; Виж 35.
- Критерий на Колмогоров - Смирнов; Виж 40.
- Проверки на хипотези за равенство между средни; Виж 41.
- Често срещана грешка при анализ на оценката на корелационния коефициент; Виж 22.
- Регресионен анализ. Виж 23.
Виж 24.

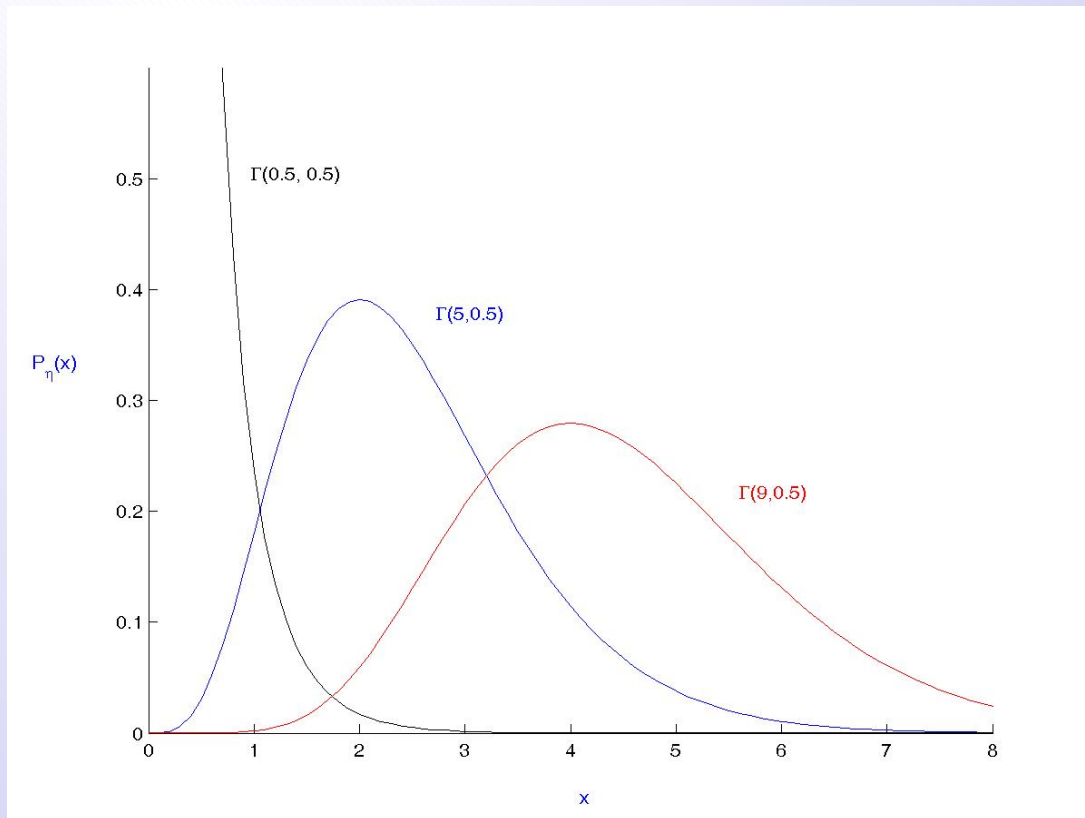
Благодаря Ви за вниманието!

Защо точно MATLAB?

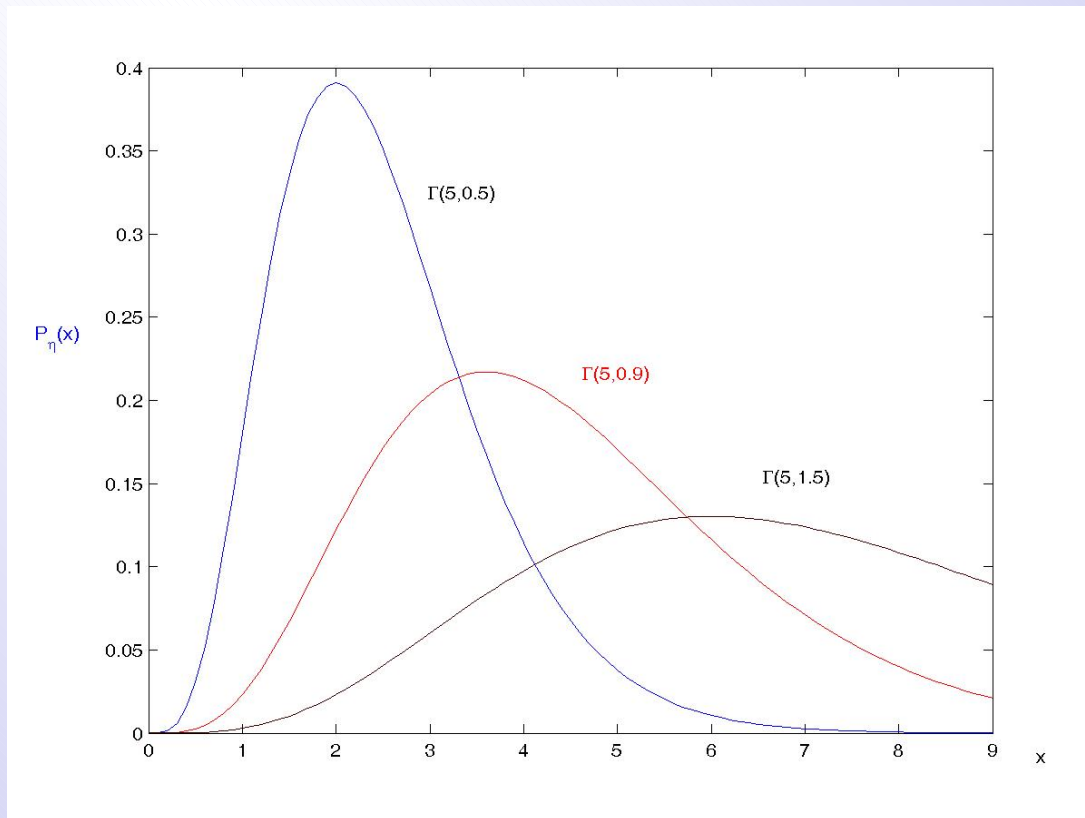
- Сравнително лесен за работа;
- Може да се използва като калкулатор с много добра точност при сложни изчисления;
- Разполага с много добри функции за симулации;
- Графичните възможности на MATLAB са практически неограничени;
- Има много добри възможности за експортиране на файлове в други продукти;
- Много добри демонстрационни примери;
- Възможностите му силно провокират изследователския дух в различни области на човешкото познание.

Недостатък: Висока за страната ни цена на лицензите.

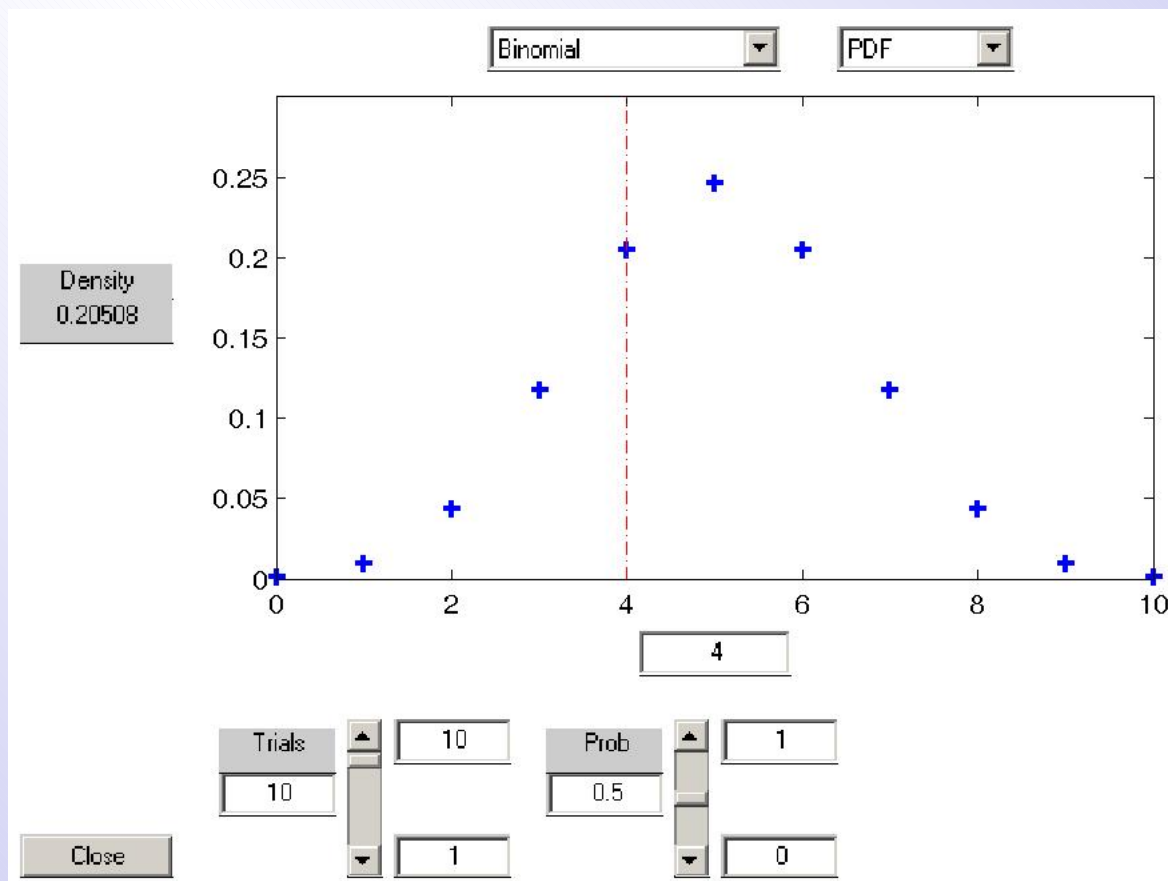
(3)



Фиг. 3



Фиг. 4



Фиг. 5

За симулиране на реализациите

$$g = \text{gamrnd}(5, 0.5, 1, 1000) \quad (6)$$

За да дефинираме координатите на абсцисите на точките

$$x = 1 : 1 : 1000$$

Изчертаването на синята линия става с

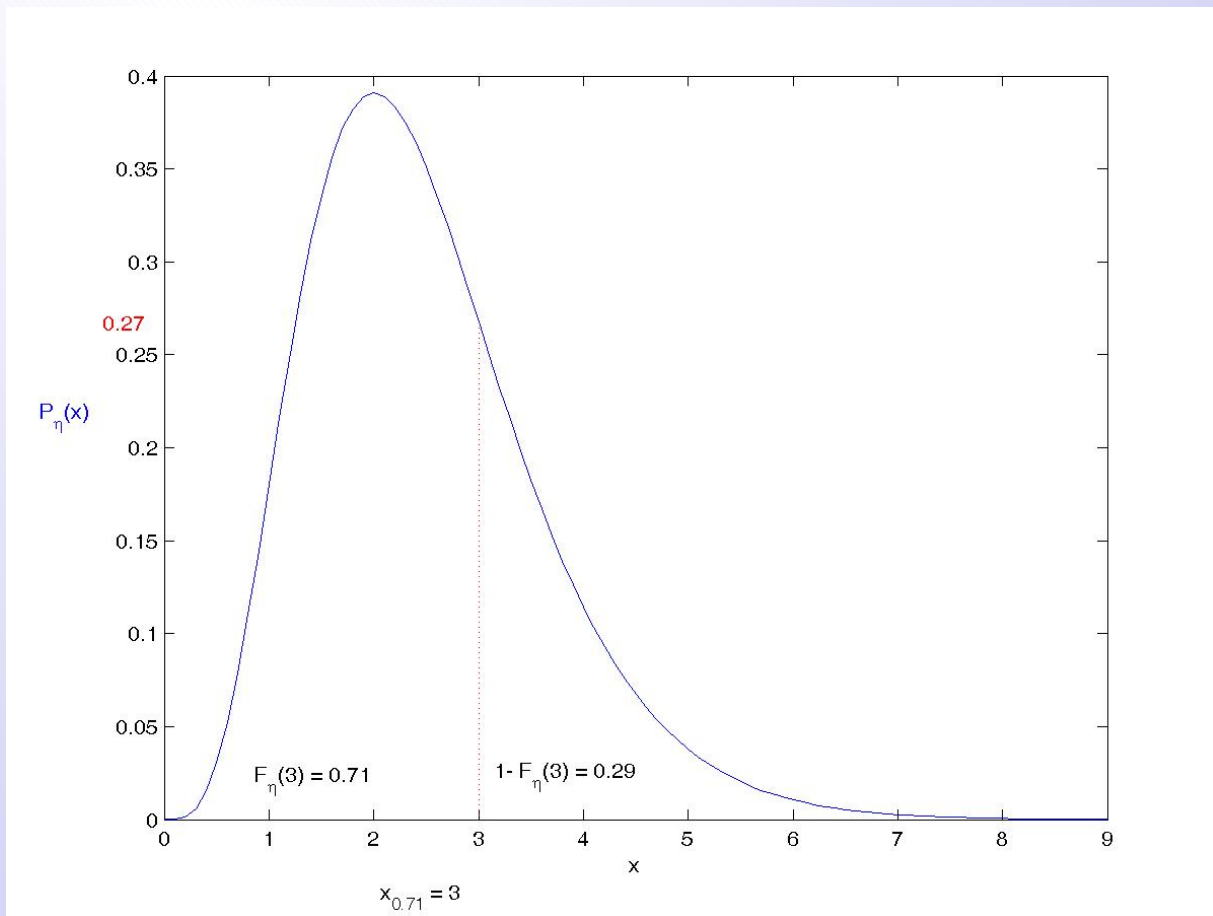
$$\text{plot}(x, g)$$

След като разполагаме с вектора g тази графика се получава непосредствено след изпълнението на командата

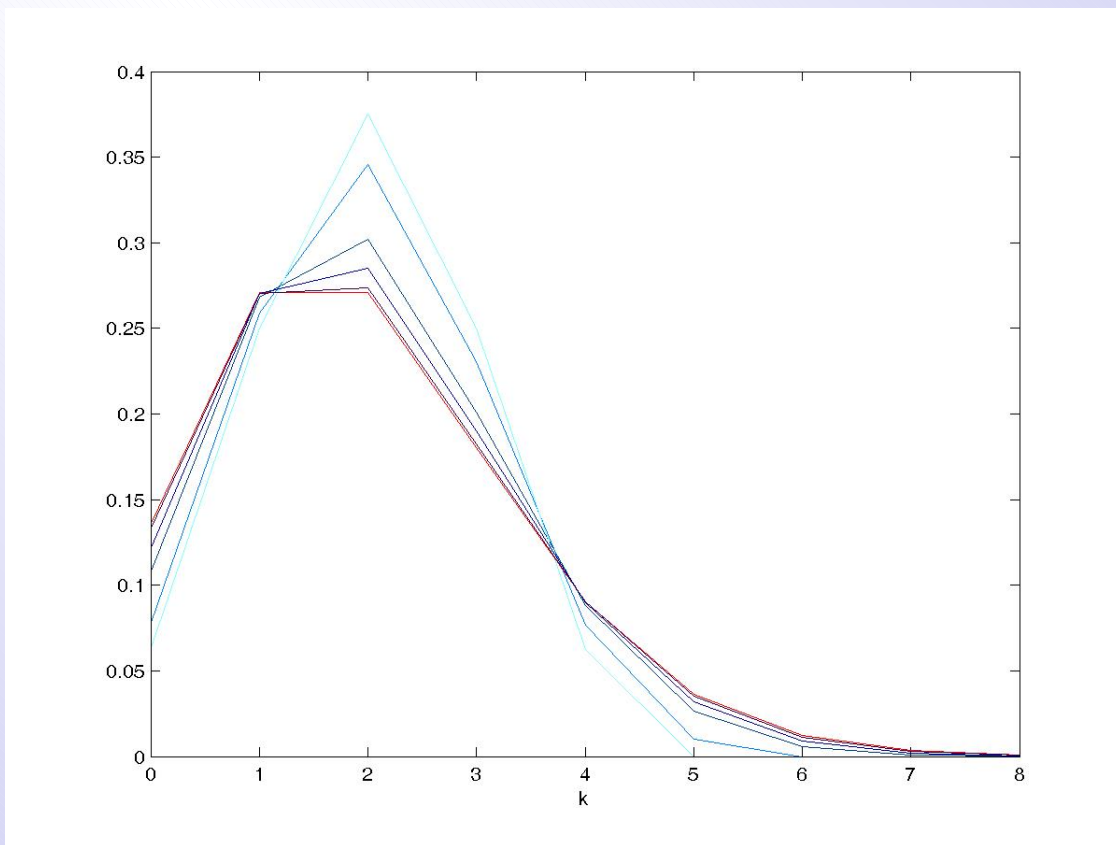
$$x1 = 0 : 0.1 : 9 \quad (7)$$

и използването на функцията

$$hist(g, x1)$$

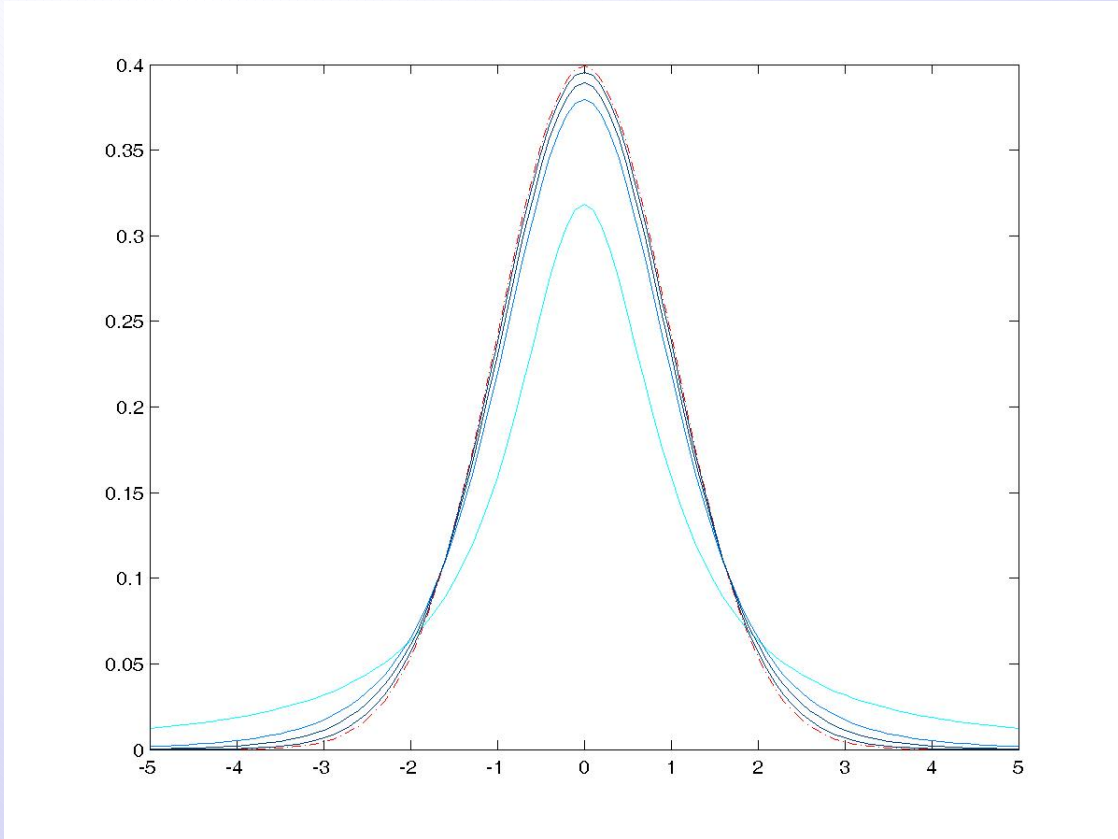


Фиг. 8



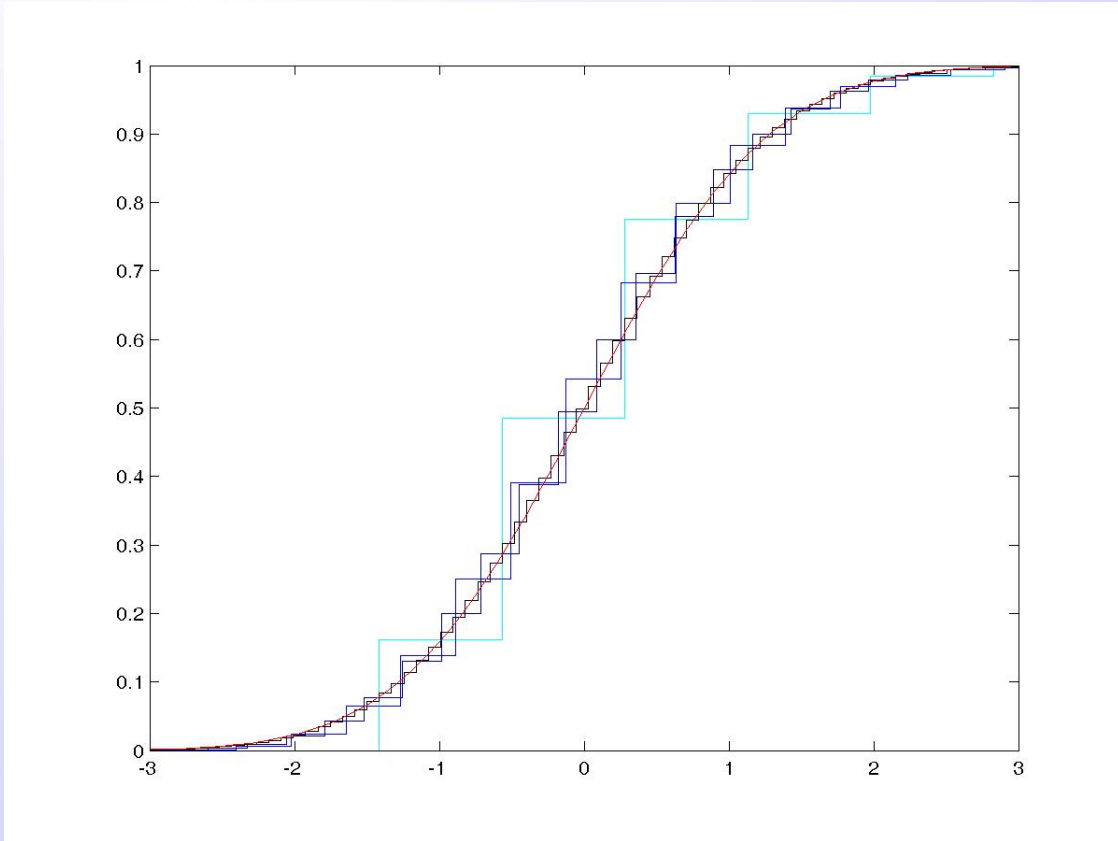
ФИГ. 9.

(10)



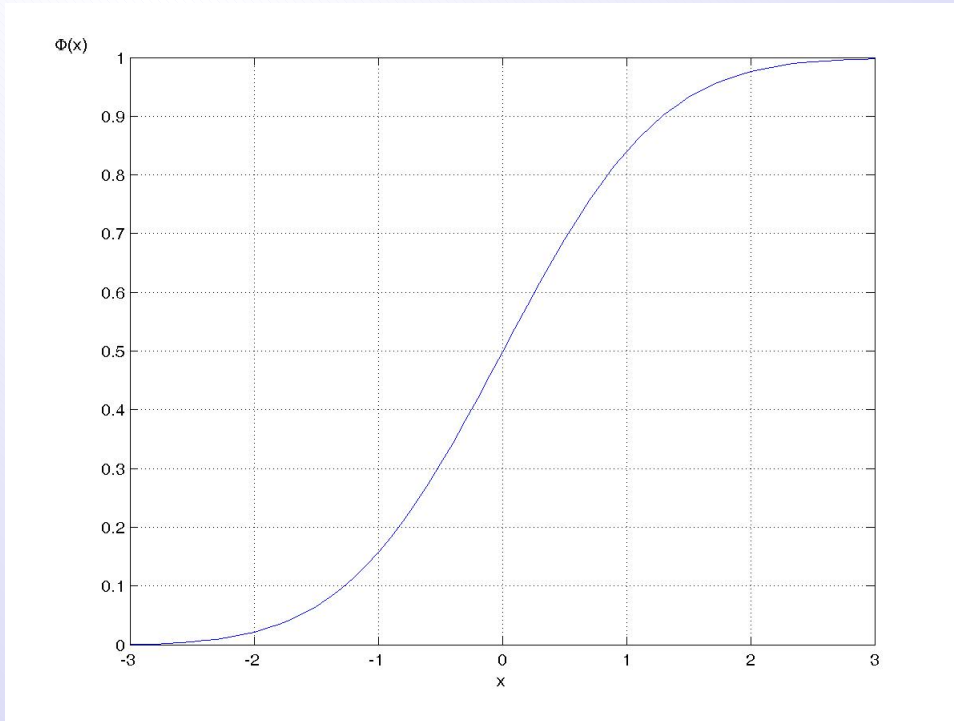
Фиг. 10.

(11)

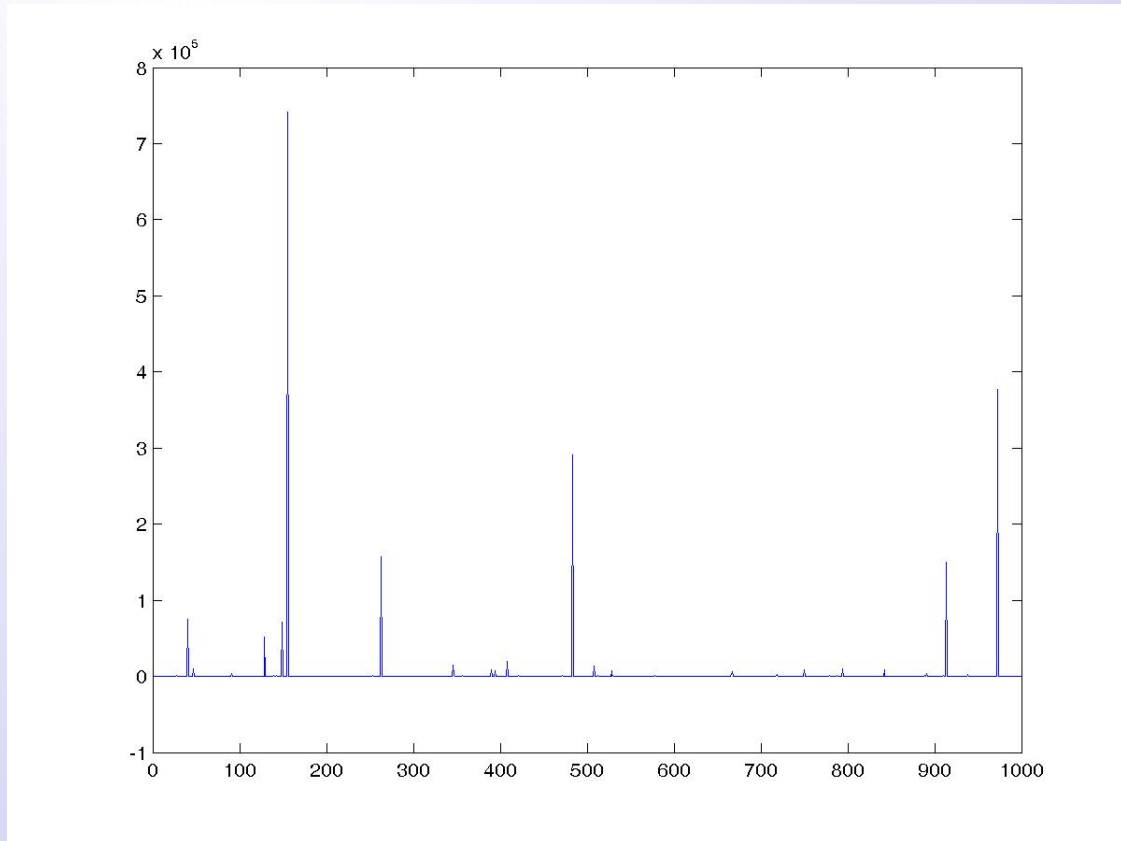


Фиг. 11.

(12)

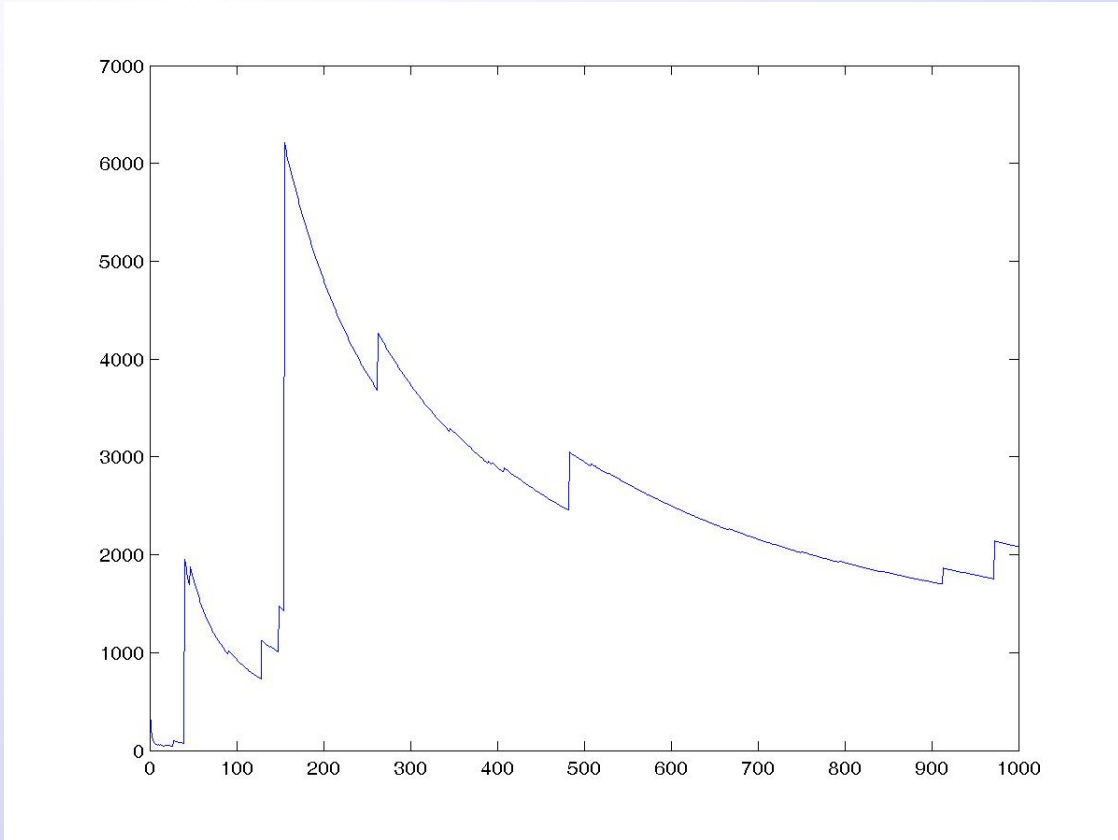


Фиг. 12.

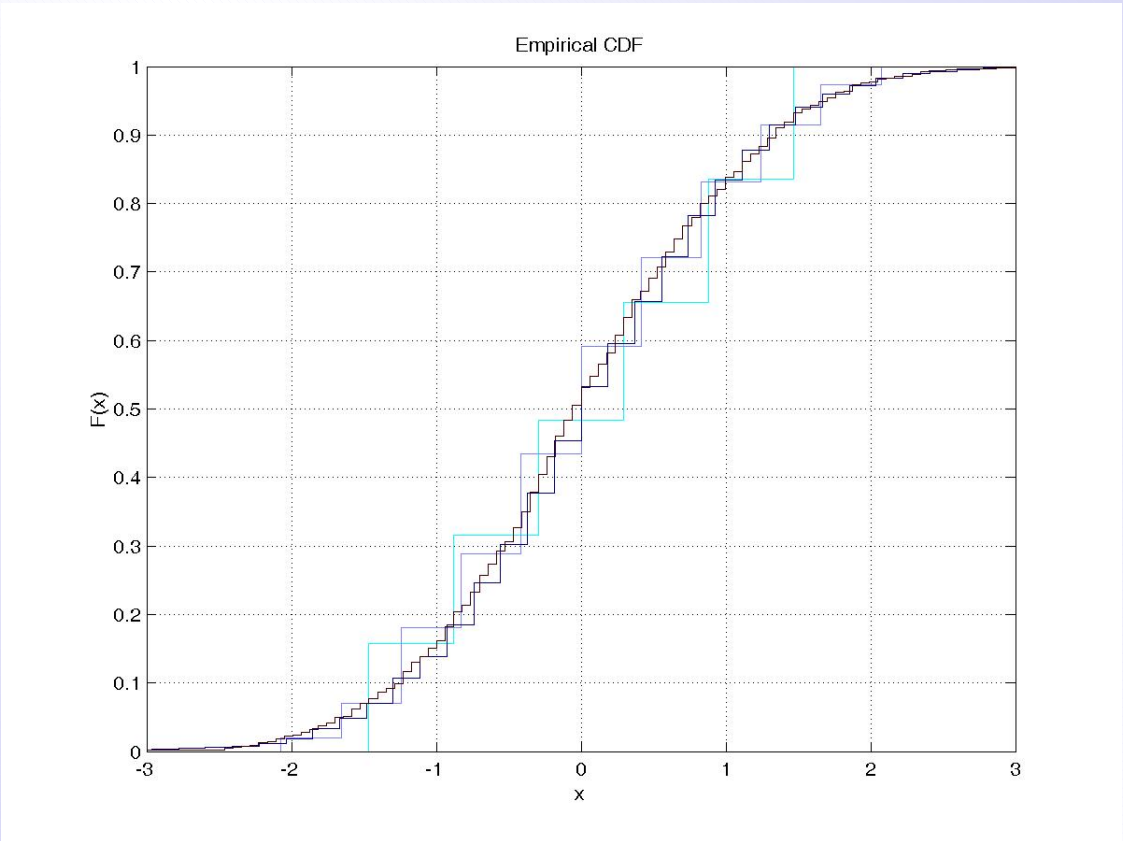


Фиг. 13. 1000 реализации на $\xi \sim S(0.5, 1)$.

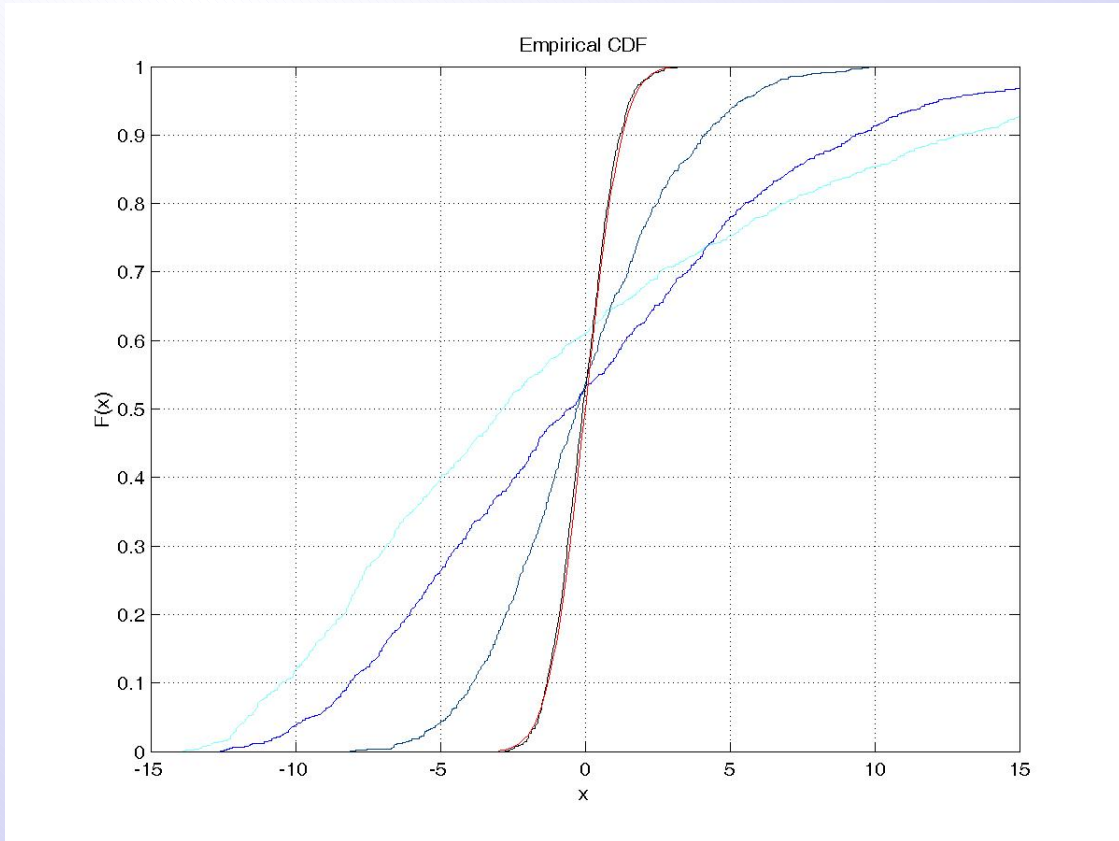
(14)



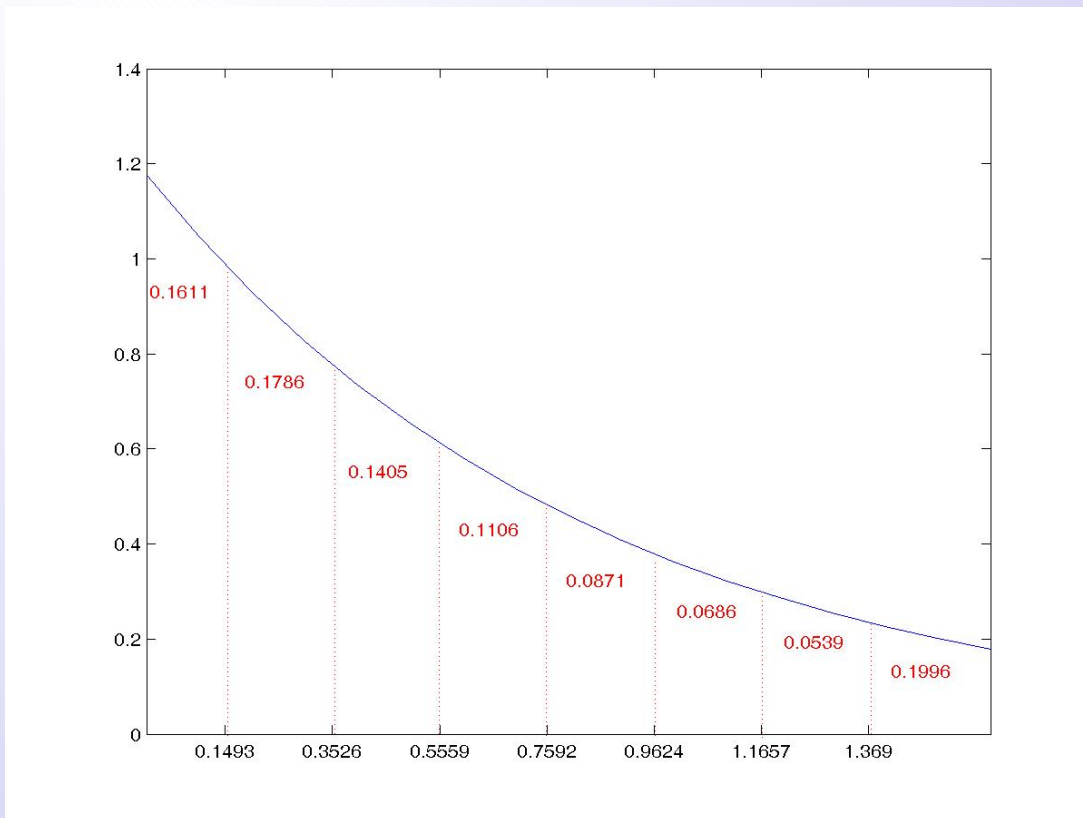
Фиг. 14.



Фиг. 15.



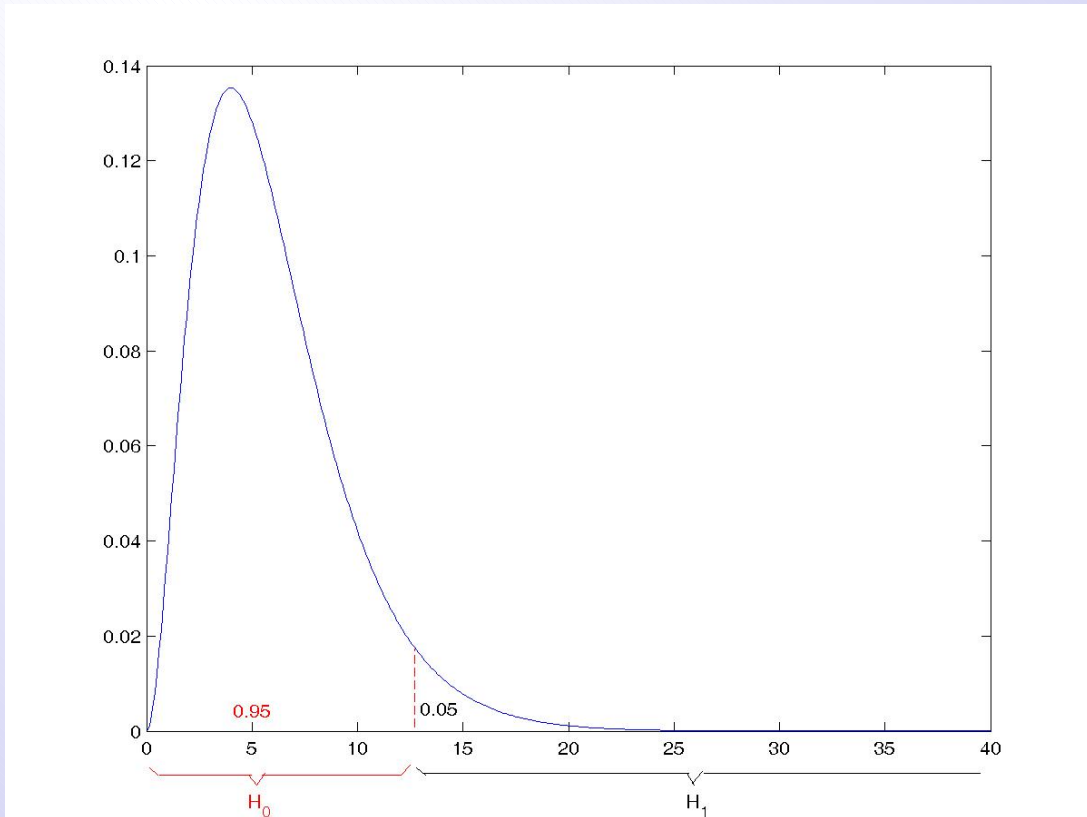
Фиг. 16.



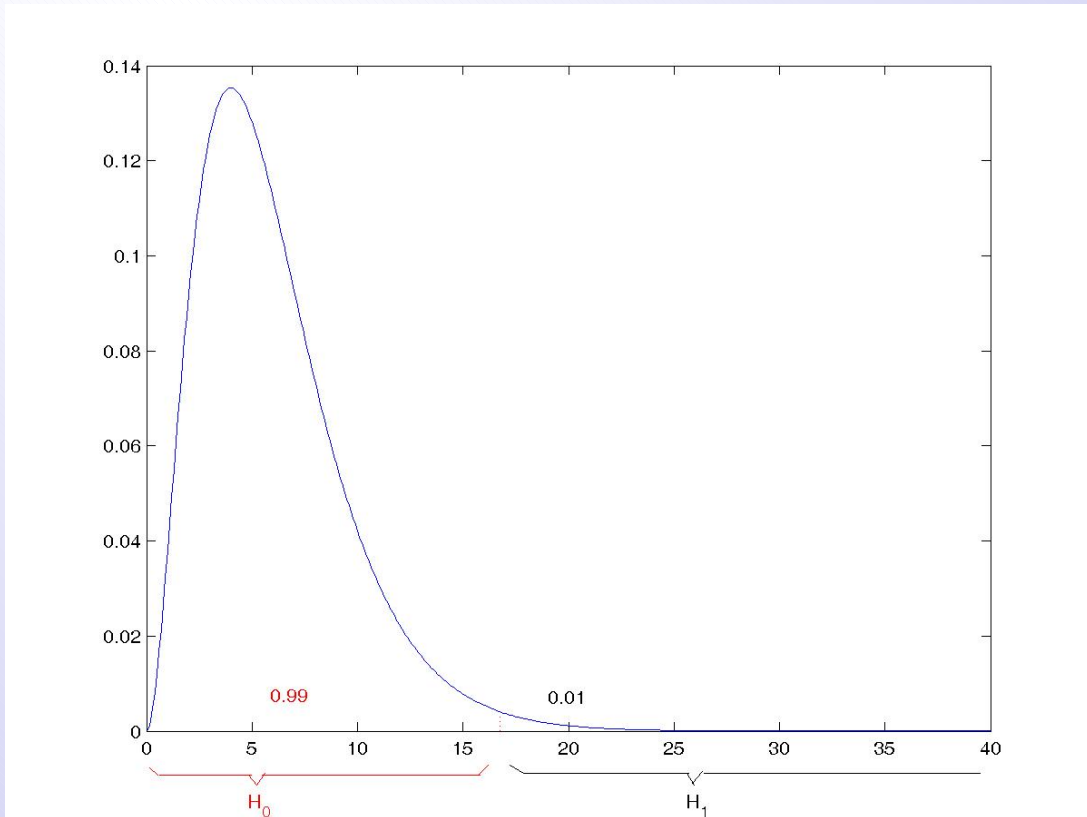
Фиг. 16.

Виж техническото изпълнение на 52

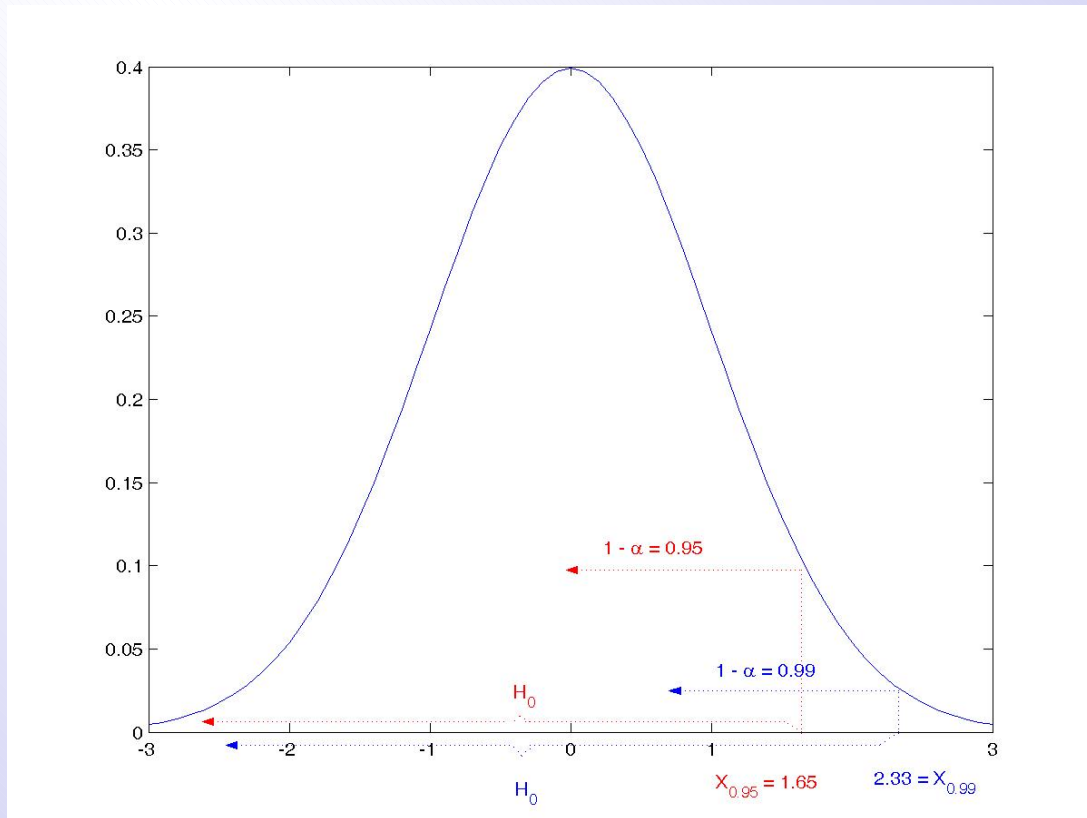
(18)



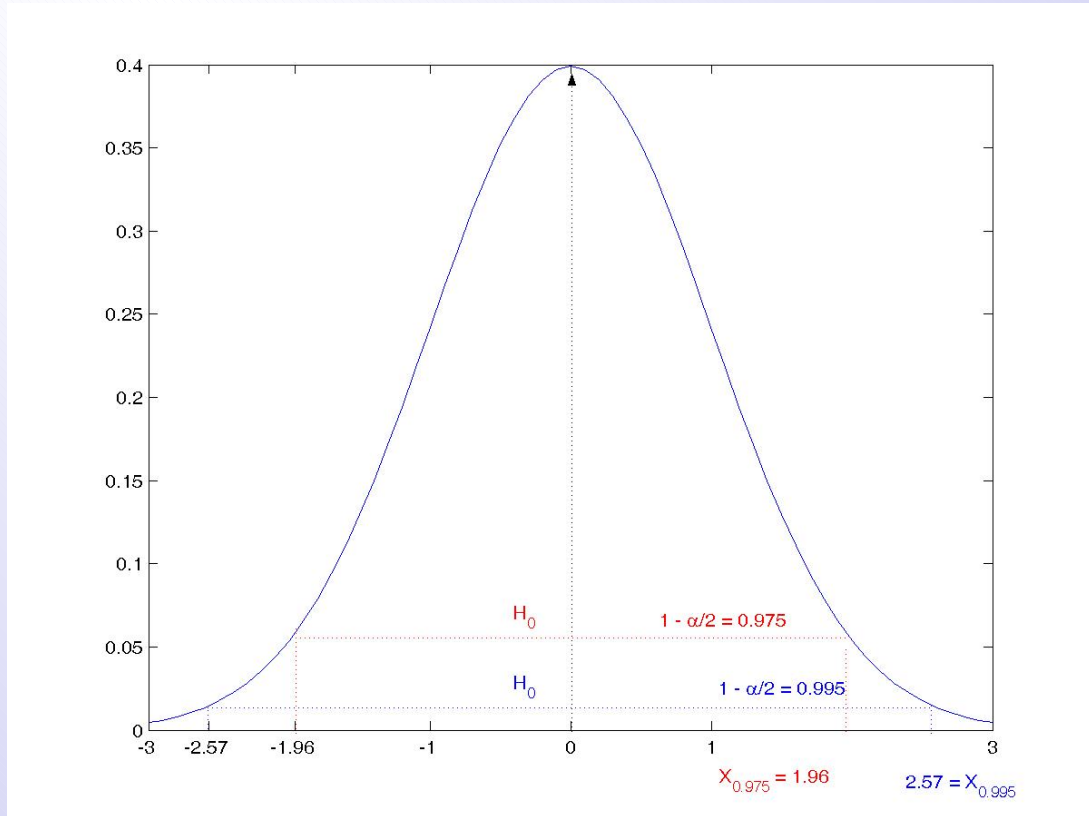
Фиг. 17.



Фиг. 18.

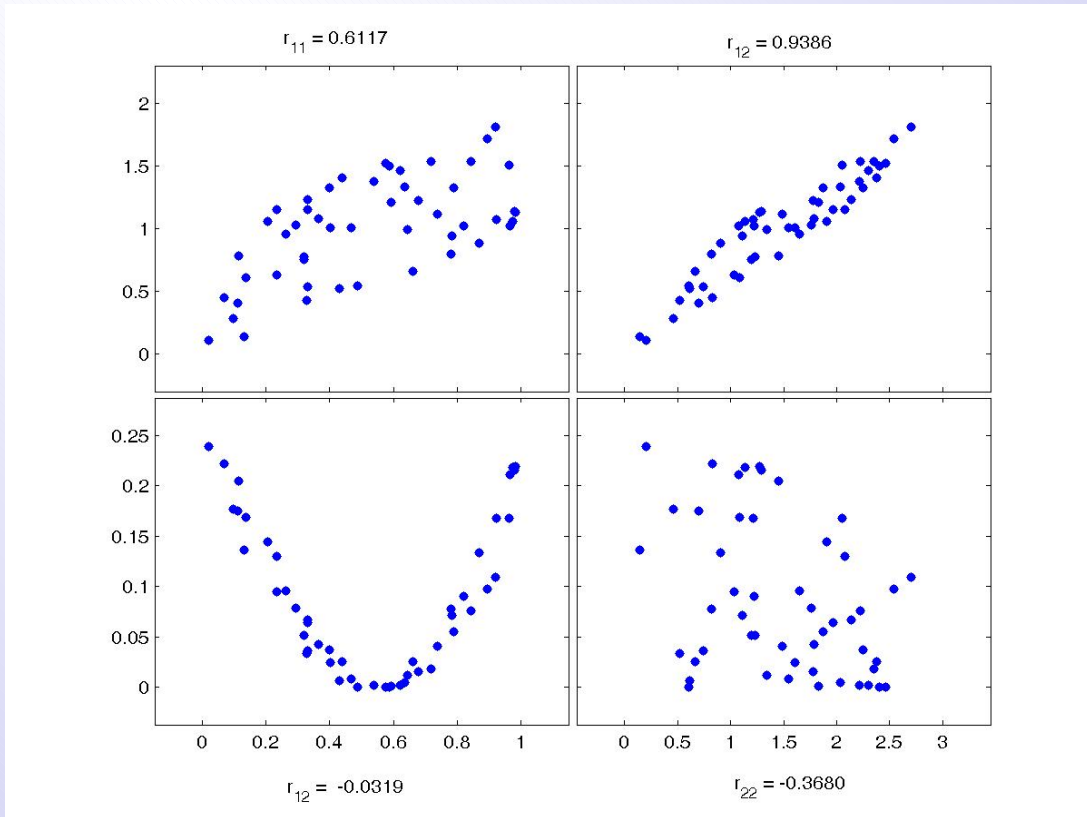


Фиг. 19. Сравняване на критичните области при $\alpha = 0.05$ и при $\alpha = 0.01$.

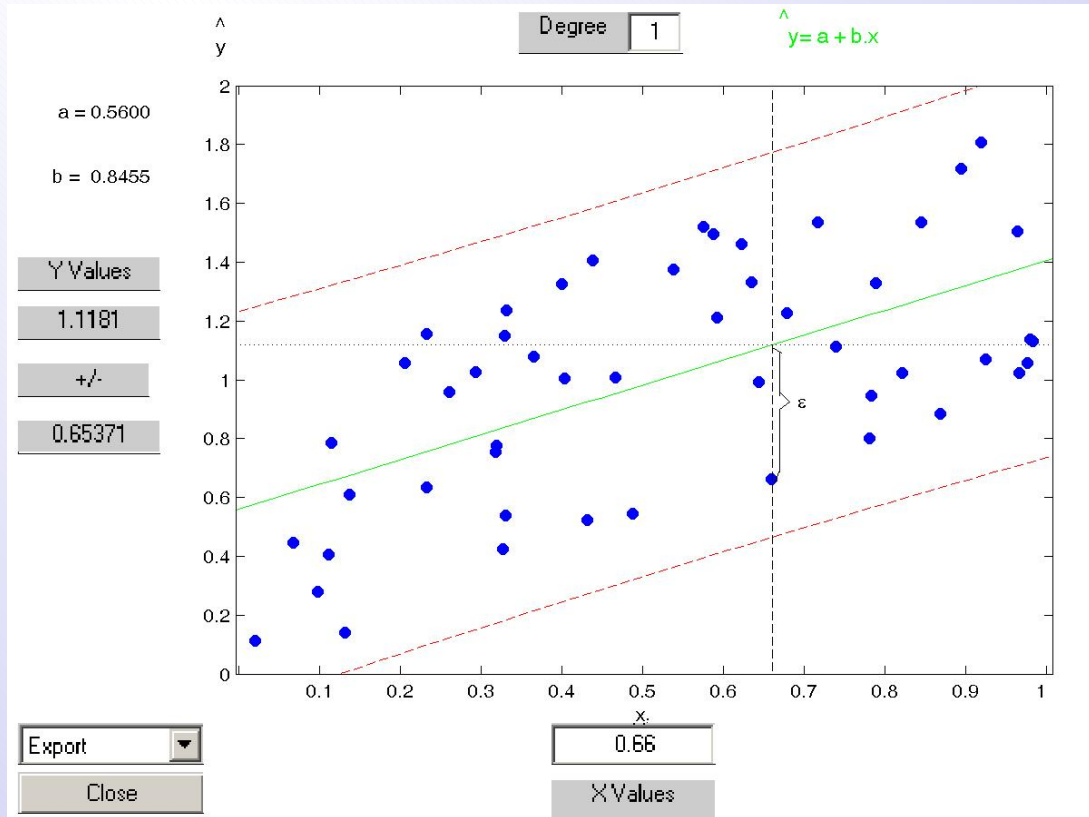


Фиг. 20. Сравняване на критичните области при $\alpha = 0.05$ и при $\alpha = 0.01$.

(22)

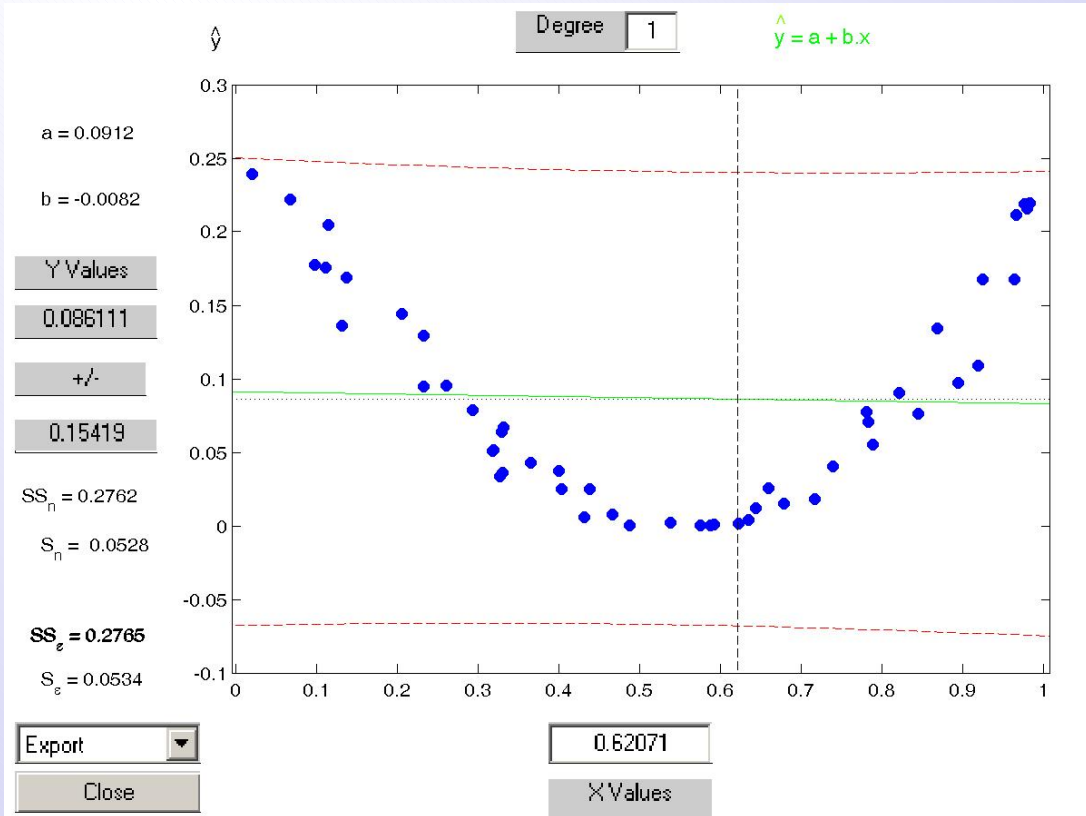


За техническото изпълнение виж (49).

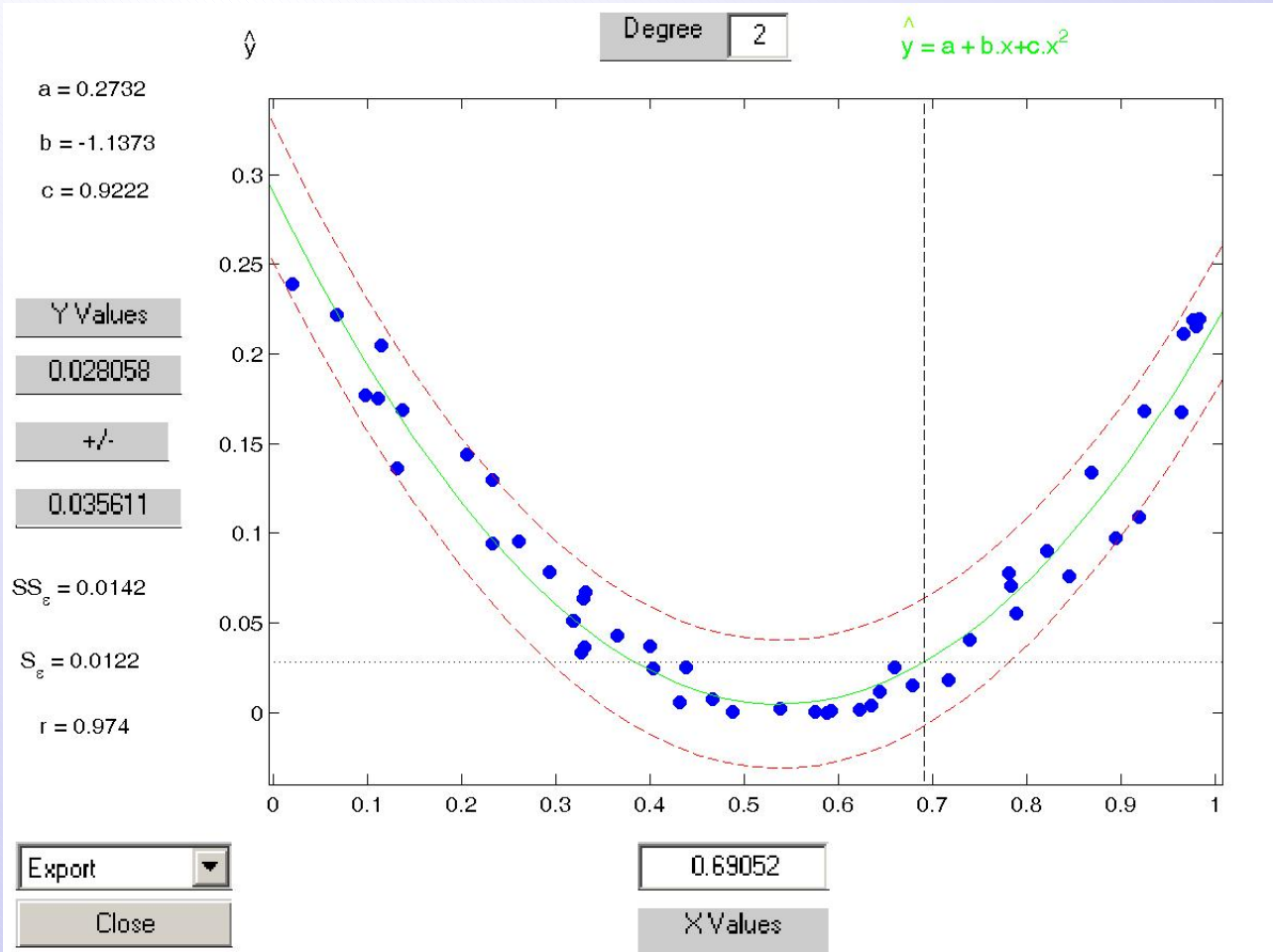


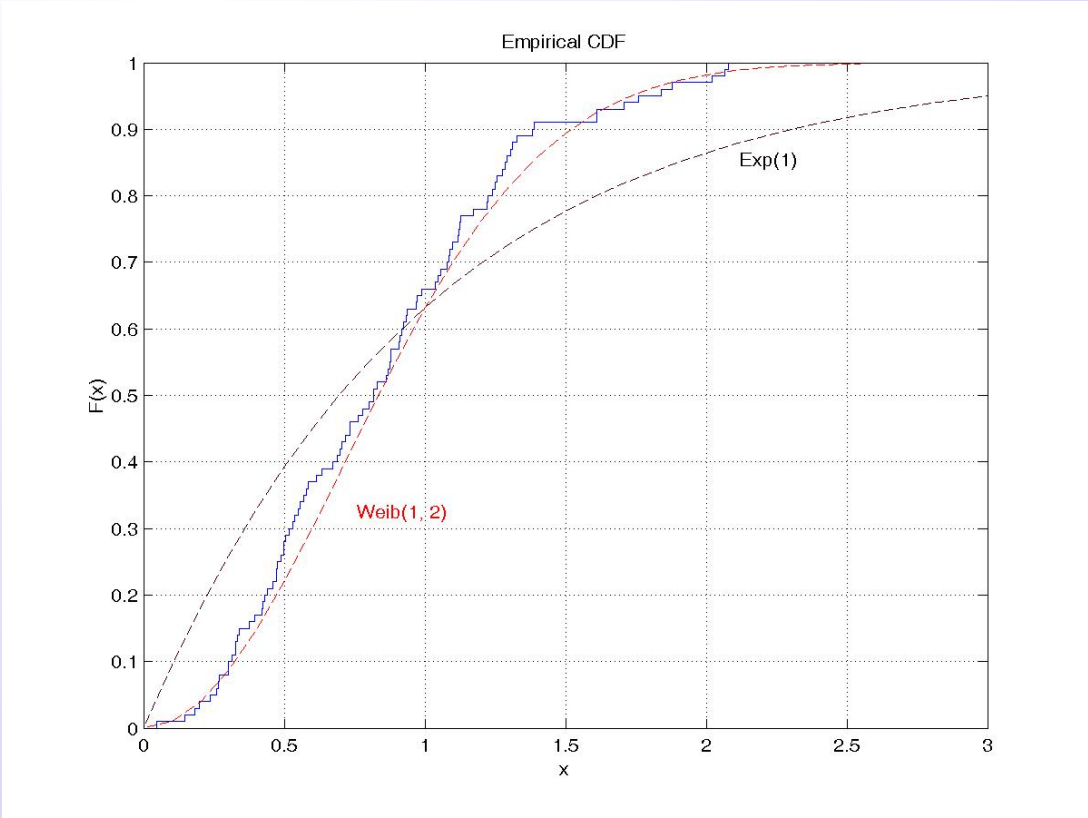
За техническото изпълнение виж (50).

(24)

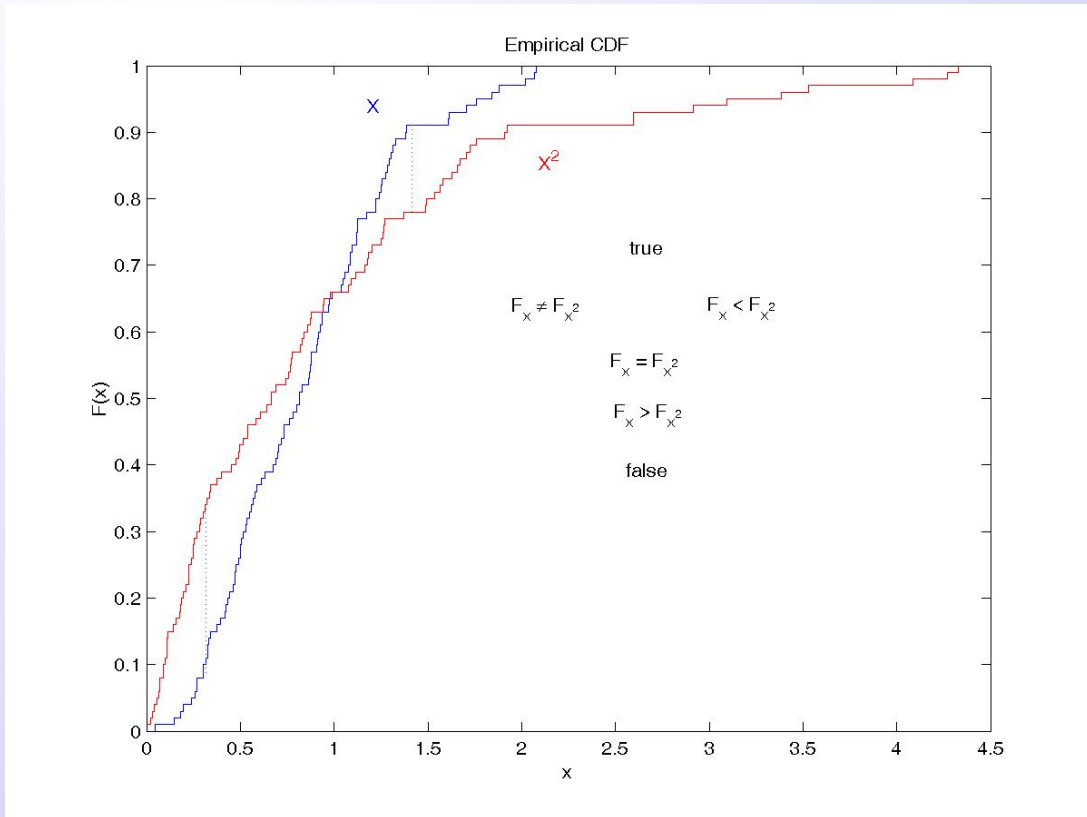


Виж още (25).





Виж техническото изпълнение на (51)



Виж техническото изпълнение на (53)

(28)

На Фиг. 9 с все по-тъмен син цвят са изобразени полигоните на разпределение съответно на

$$Bi(4, \frac{1}{2}),$$

$$Bi(5, \frac{4}{10}),$$

$$Bi(10, \frac{2}{10}),$$

$$Bi(20, \frac{1}{10}),$$

$$Bi(100, \frac{2}{100})$$

случайни величини. С червен цвят е изобразен полигонът на $Po(2)$.

Техническото изпълнение виж в (36).

(29)

На Фиг. 10 с все по-тъмен син цвят са изобразени плътностите на разпределение съответно на

$$t(1),$$

$$t(5),$$

$$t(10),$$

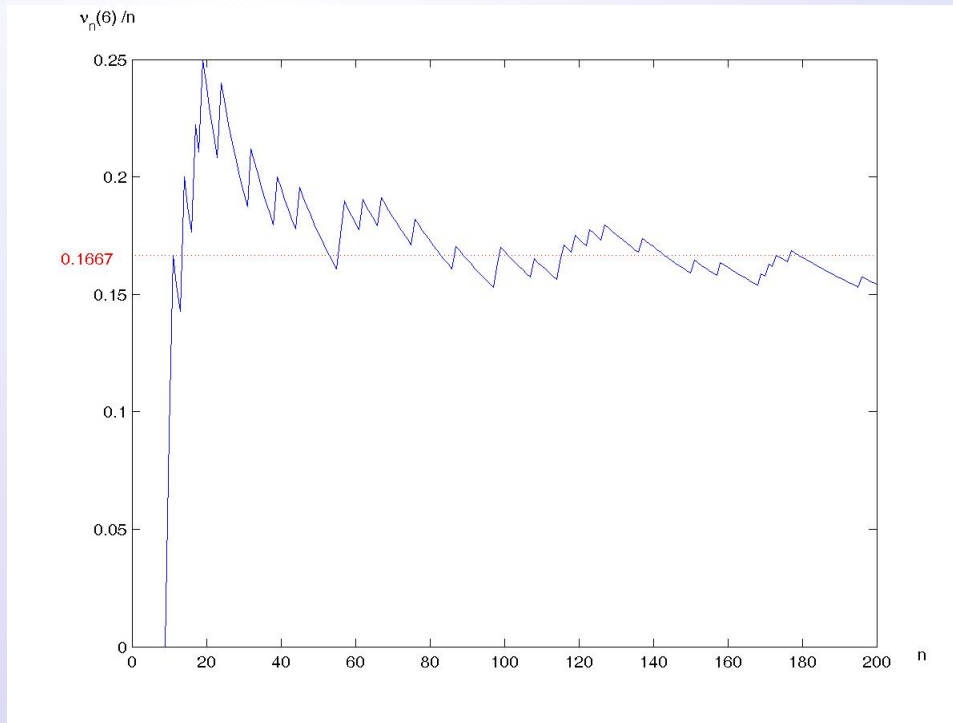
$$t(30)$$

случайни величини. С червен цвят е изобразена плътността на $N(0, 1)$.

Техническото изпълнение виж в (37).

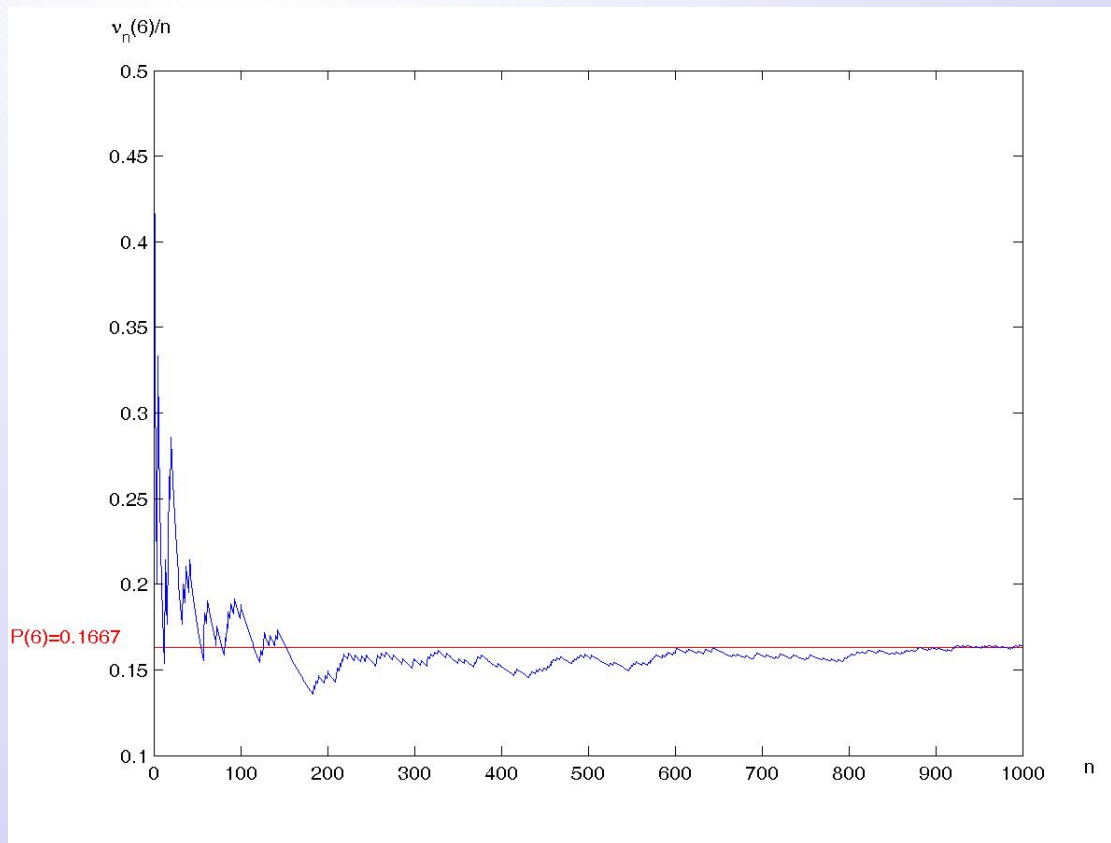
(30)

Пример 2: Подхвърля се симетричен зар 200 пъти. При всяко подхвърляне, се пресмята относителната честота на шестите до момента. Получените резултати са изобразени на следната графика.



Виж реализациите в (38)

При 1000 подхвърляния



На Фиг. 11 с все по-тъмен син цвят са изобразени функциите на разпределение

$$\frac{\nu_n - n\frac{1}{6}}{\sqrt{n\frac{1}{6}(1 - \frac{1}{6})}} = \frac{\frac{\nu_n}{n} - \frac{1}{6}}{\sqrt{\frac{1}{6n}(1 - \frac{1}{6})}}, \quad (31)$$

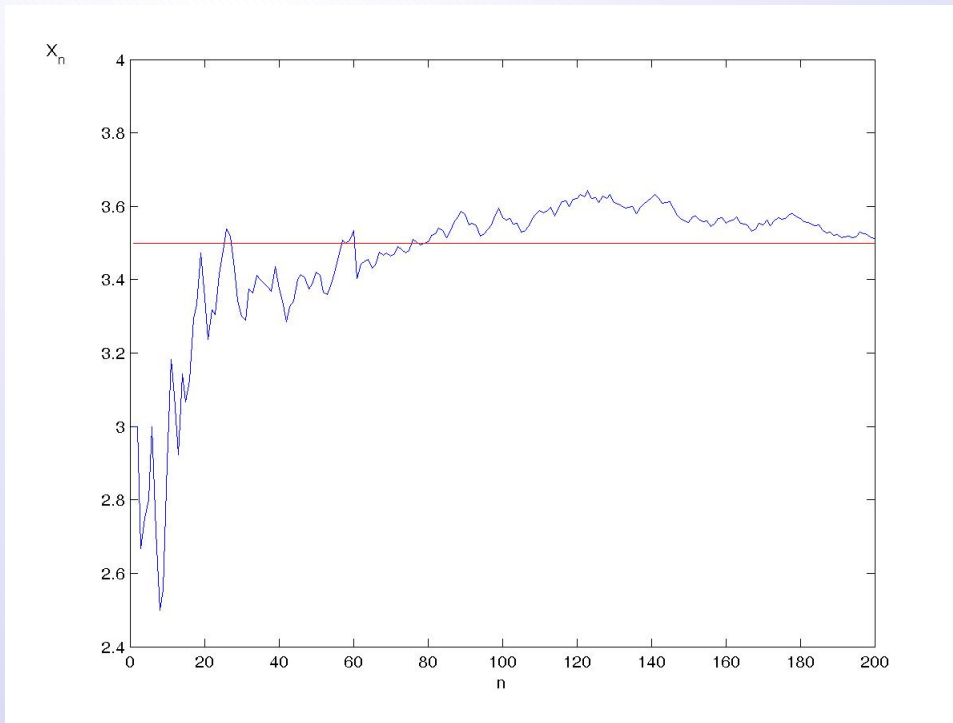
където $\nu_n \sim Vi(n, \frac{1}{6})$, съответно за $n = 10, 50, 100, 1000$.

С червен цвят е изобразена функцията на разпределение на $N(0, 1)$ случайна величина.

Техническото изпълнение при $n = 1000$ виж в (46).

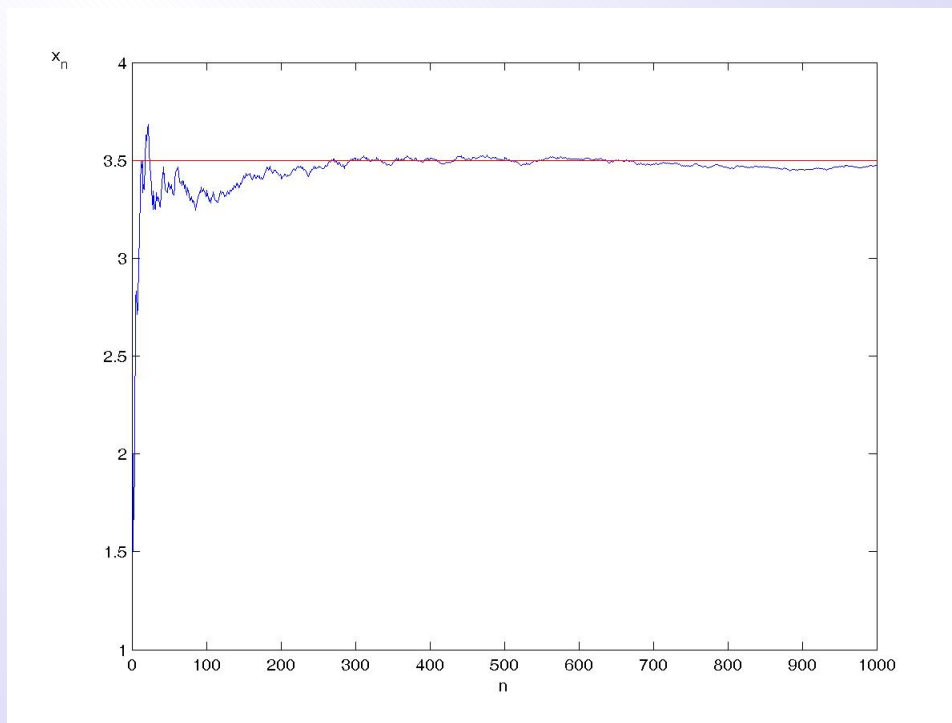
(32)

Пример 3: Подхвърля се симетричен зар 200 пъти. При всяко подхвърляне, се пресмята средният брой точки до момента. Получените резултати са изобразени на следната графика.



Виж реализациите в (38)

При 1000 подхвърляния



Техническото изпълнение виж в (47).

Виж на Фиг. 13 и Фиг. 14 какво се случва, когато не съществува м.о.

Пример 4. Нека X_1, \dots, X_n са н.е.р. сл. величини и X_1 е броят на точките, които са се паднали при едно подхвърляне на симетричен зар, тогава \bar{X}_n е средният брой точки, които са се паднали върху зара при n подхвърляния. На Фиг. 15 с все по-тъмен син цвят са изобразени функциите на разпределение ф.р. на

$$\frac{X_1 + \dots + X_n - n.a}{\sigma\sqrt{n}} = \frac{\bar{X}_n - a}{\frac{\sigma}{\sqrt{n}}}, \quad (33)$$

съответно при $n = 1, 2, 10$ и 100 където $a = 3.5$ и $\sigma = 1.707$.

Графиката на ф.р. на $\psi \sim N(0, 1)$ виж на Фиг. 12.

Техническото изпълнение при $n = 10$ виж в (48).

Пример 5. Нека X_1, \dots, X_n са н.е.р. сл. величини и $X_1 \sim \Gamma(2, 5)$. На Фиг. 16 с все по-тъмен син цвят са изобразени функциите на разпределение на

$$\frac{X_1 + \dots + X_n - n.a}{\sigma\sqrt{n}} = \frac{\bar{X}_n - a}{\frac{\sigma}{\sqrt{n}}}, \quad (34)$$

съответно при $n = 1, 2, 10$ и 100 където $a = 10$ и $\sigma^2 = 50$. С червен цвят е изобразена графиката на ф.р. на $\psi \sim N(0, 1)$.

(35)

Пример 6. Извършени са 100 нез. наблюдения върху ξ . След групировка, с десни краища на интервалите

0.1493 0.3526 0.5559 0.7592 0.9624 1.1657 1.3690 1.5723.

честотите в интервалите са се оказали 5, 16, 18, 13, 17, 14, 10, 7, а средната аритметична на резултатите от наблюдението е била 0.8496. Да се провери дали разпределението на наблюдаваната величина е експоненциално.

$$emp = \sum_{i=1}^8 \frac{(100 * p_i - f_i)^2}{p_i * 100} = 36.9836$$

$$\chi_{0.95}^2(6) = 12.5916, \quad \chi_{0.99}^2(6) = 16.8119.$$

- Виж геометричната интерпретация на p_i на Фиг. 17;
- На Фиг. 18 и 19 виж геометричната интерпретация на критичната област и влиянието на изменението на α .

(36)

```
k = 0:1:8  
y4 = binopdf(k, 4, 2/4)  
y5 = binopdf(k, 5, 2/5)  
y10 = binopdf(k, 10, 2/10)  
y20 = binopdf(k, 20, 2/20)  
y100 = binopdf(k, 100, 2/100)  
pois=poisspdf(k, 2)  
  
plot(k,y4,k,y5,k,y10,k,y20,k,y100,k,pois).
```

(37)

```
x=-5:0.1:5
y1=tpdf(x,1)
y5 = tpdf(x,5)
y10=tpdf(x,10)
y30=tpdf(x,30)
z = normpdf(x,0,1)
plot(x, y1,x, y5, x, y10, x, y30,x,z,'-.','color','r')
```


(38)

Резултати от проведения опит:

3 3 2 3 3 4 1 1 3 6 6 2 1 6 2 4 6 4 6 1 1 5 3 6 5 5 3 1 1 2
3 6 3 5 3 3 3 3 6 1 2 1 5 4 6 4 3 2 4 5 3 1 3 5 5 6 6 3 4 5
2 6 4 4 2 4 6 3 4 3 4 5 3 3 4 6 3 3 4 4 5 4 5 3 2 5 6 5 5 3
1 4 3 1 4 5 5 6 6 1 3 4 2 4 1 4 5 6 5 5 3 4 5 1 6 6 4 2 6 4
5 3 6 1 4 2 6 3 5 1 3 3 3 4 4 1 6 5 5 5 5 2 2 4 4 1 1 2 3 3
6 4 2 3 4 1 5 6 4 1 5 4 5 1 3 3 1 5 6 3 6 1 6 5 3 4 6 4 2 3
2 3 3 3 4 1 2 4 2 4 2 4 4 3 4 6 3 3 2 3

Относно симулирането на подхвърлянето на зар виж (39)

Можете да симулирате числата, които са се паднали при n подхвърляния на един зар като се обърнете към функцията

$$\text{unidrnd}(k, 1, n)$$

Тя генерира n реализации на равномерно разпределена върху $\{1, 2, \dots, k\}$ случайна величина.

За да симулирате резултатите от 200 подхвърляния е достатъчно в командния прозорец на MATLAB да напишете

$$\text{zar} = \text{unidrnd}(6, 1, 200). \quad (39)$$

Тук zar е името на матрица 1×200 , която получавате като резултат.

(40)

- Тест на Колмогоров (Една извадка); Виж 26.
- Избор на двустранна, лявостранна или дясностранна критична област при работа с теста на Смирнов (Две извадки). Виж 27.

- Средна на генерална съвкупност и константа

- с известна и крайна дисперсия;
- с неизвестна дисперсия;

Виж 42.

- Средни на две генерални съвкупности

- от независими и нормално разпределени съвкупности;

Виж 43.

- от нормално разпределени съвкупности;

Виж 42.

- Непараметрични методи;

Виж 44.

- Средни на повече от две генерални съвкупности

- Дисперсионен анализ;
- Тест на Kruskal-Wallis.

(42)

Извеждат се поотделно критериите за работа с алтернативните хипотези и се демонстрира графично зависимостта на критичната област от α .

- $H_1 : m.o. > a$;

Виж 20.

$$W_\alpha = \{(x_1, x_2, \dots, x_n) : \frac{\bar{X}_n - a}{\frac{\sigma}{\sqrt{n}}} \geq x_{1-\alpha}\}$$

- $H_1 : m.o. < a$;

$$W_\alpha = \{(x_1, x_2, \dots, x_n) : \frac{a - \bar{X}_n}{\frac{\sigma}{\sqrt{n}}} \geq x_{1-\alpha}\}$$

- $H_1 : m.o. \neq a$;

Виж 21.

$$W_\alpha = \{(x_1, x_2, \dots, x_n) : \left| \frac{a - \bar{X}_n}{\frac{\sigma}{\sqrt{n}}} \right| \geq x_{1-\frac{\alpha}{2}}\}$$

(43)

- С изнестни дисперсии;
- С неизвестни дисперсии.

(44)

- Критерий на значите;
- Независими извадки. Критерий на Wilkokson.

```
z10 = unidrnd(6,1000,10);
```

(45)

```
function [y]= c6(a)
```

```
[m,n]=size(a);
```

```
for i=1:m
```

```
    br=0
```

```
    for j=1:n
```

```
        if a(i,j) == 6,
```

```
            br = br + 1;
```

```
        else br = br;
```

```
        end
```

```
        y(i) = br;
```

```
    end
```

```
end y
```

```
c10 = c6(z10);
```

```
x10 =(c10/10-1/6)/sqrt(1/6/10*(1-1/6));
```

```
cdfplot(x10)
```

```
hold on
```



```
>>i_1000 = 0:1:1000
>> k_1000 = (i_1000-1000*1/6)/sqrt(1000*1/6*5/6)
>> a_1000=binocdf(i_1000,1000,1/6)
>>y(1)=0
>>y(2)=a_1000(1)
>>x(1)=k_1000(1)
>>x(2)=k_1000(1)
>>for s=2:1001
x(2*s)=k_1000(s)
x(2*s-1)=k_1000(s)
y(2*s)=a_1000(s)
y(2*s-1)=a_1000(s-1)
end
>>plot(x,y)
>>x = -3:.1:3
>>normcdf(x)
>>hold on
>>plot(x,ans)
```



```
k = unidrnd(6,1,1000);
```

(47)

```
function [y]= kolm(m)
y(1) = 0;
[p,q]=size(m)
for i=1:q
    y(i) = mean(m(1:i))
end
```

```
sr=kolm(k);
plot(sr)
```

```
z10=unidrnd(6,1000,10)
```

(48)

```
function [y]= clt(a)
[m,n]=size(a);
for i=1:m
    y(i)=(mean(a(i,[1:n]))-3.5)/1.707*sqrt(n)
end

y10=clt(z10)
cdfplot(y10)
hold on
```

(49)

```
X2 = unifrnd(0,1,1,50)
```

```
e=unidrnd(0,1,50,1)
```

```
plotmatrix([X2, X2+2*e], [X2+e (X2-e/10-0.5).^2])
```

```
c = [X2 X2+2*e X2+e (X2-e/10-0.5).^2]
```

```
S = corrcoef(c)
```

Получаваме матрица с отговорите

S =

1.0000	0.3013	0.6117	-0.0319
0.3013	1.0000	0.9386	-0.3680
0.6117	0.9386	1.0000	-0.3169
-0.0319	-0.3680	-0.3169	1.0000

(50)

```
X2 = unifrnd(0,1,1,50)
e=unidrnd(0,1,50,1)
polytool(X2, X2+e)
s = sqrt(sum(residuals)/(50-2-1))
```

```
sY = std(X2(:,4)+e(:,4))
```

```
sqrt(1-s/sY)
```

(51)

```
x = weibrnd(1, 2, 100, 1);  
» cdfplot(x)  
» xx = 0:.1:3  
» hold on  
» plot(xx,weibcdf(xx, 1, 2))  
» plot(xx,expcdf(xx, 1))  
» kstest(x, [x weibcdf(x, 1, 2)], 0.5)  
h = 0  
» kstest(x, [x expcdf(x, 1)], 0.5)  
h = 1
```

(52)

```
x = weibrnd(1, 2, 100, 1);
```

```
» [f,y1] = hist(x)
```

```
» xx = 0:.1:3
```

```
» plot(xx, exppdf(xx, 0.8496))
```

```
» e = expcdf(y1, 0.8496)
```

```
e = 0.1611 0.3397 0.4801 0.5908 0.6779 0.7464 0.8004 0.8429
```

```
» p = [e(1) e(2:7)-e(1:6) 1-e(7)]
```

```
p = 0.1611 0.1786 0.1405 0.1106 0.0871 0.0686 0.0539 0.1996
```

```
» emp = sum((100 * p - f).^2./(p * 100))
```

```
»  $\chi_{0.95}^2(8 - 1 - 1) = \text{chi2inv}(1 - 0.05, 6)$ 
```

```
»  $\chi_{0.99}^2(8 - 1 - 1) = \text{chi2inv}(1 - 0.01, 6)$ 
```

» $[h,p,k] = \text{kstest2}(x,y,0.05,0)$

$h = 1$

$p = 0.0050$

$k = 0.2400$

» $[h,p,k] = \text{kstest2}(x,y,0.05,1)$

$h = 0$

$p = 0.1728$

$k = 0.1300$

» $[h,p,k] = \text{kstest2}(x,y,0.05,-1)$

$h = 1$

$p = 0.0025$

$k = 0.2400$