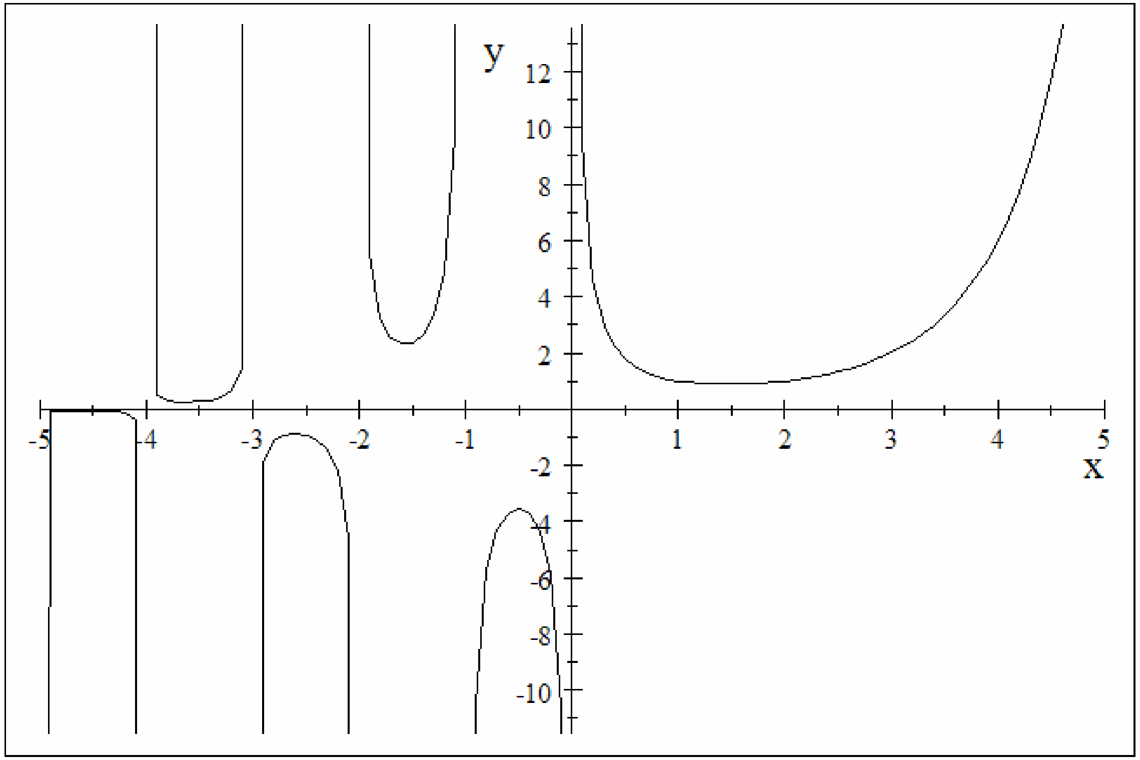
Daži svarīgi varbūtību sadalījumi.

Gamma sadalījums. Gamma funkcija $\Gamma(t) = \int_0^\infty e^{-x} x^{t-1} dx$

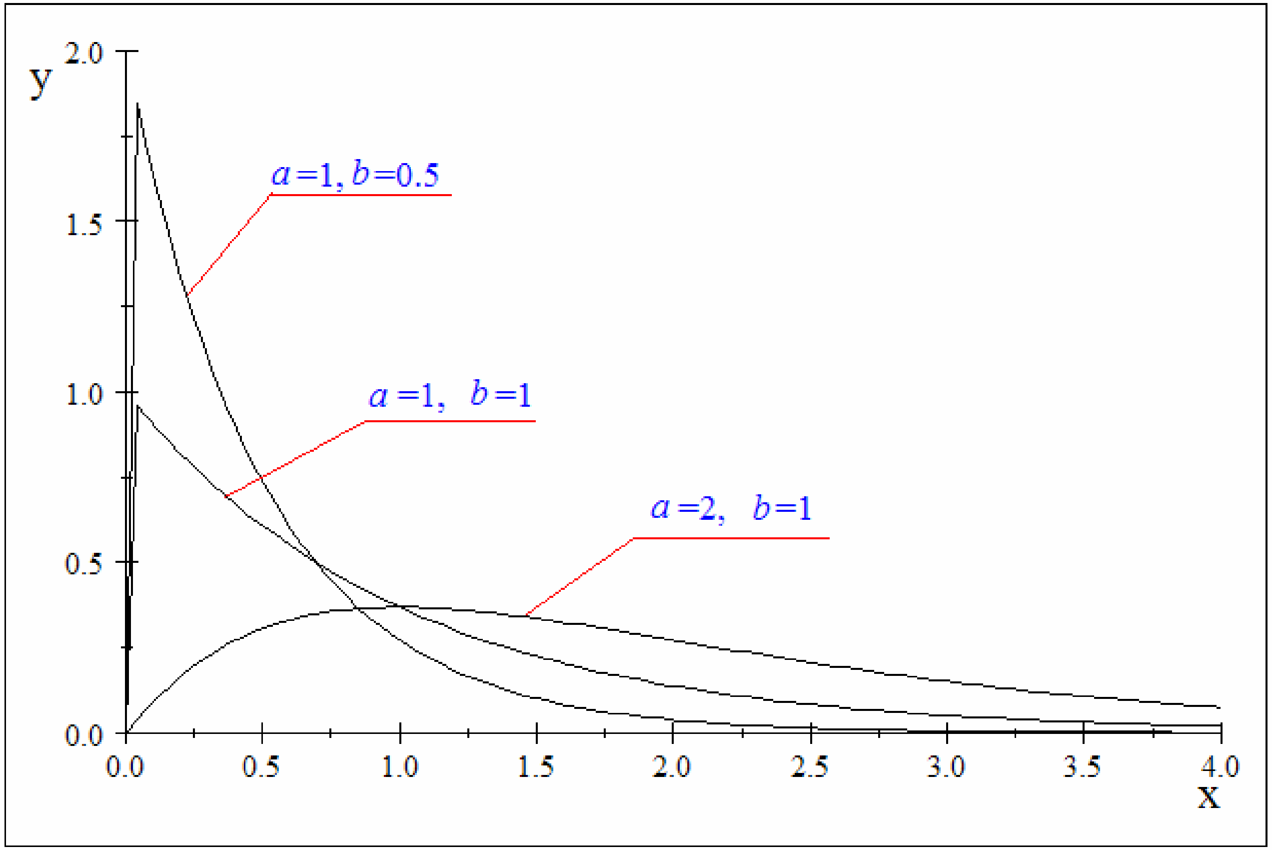


$\Gamma(1) = 1$, $\Gamma(t+1) = t\Gamma(t)$, veselām, pozitīvām argumenta vērtībām $\Gamma(k+1) = k!$

Gamma sadalījuma blīvuma funkcija $p_\xi(x) = \dfrac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}}$, $x > 0$

Parametri $a > 0$ – formas parametrs, $b > 0$ – mēroga parametrs.

Skaitliskie raksturotāji: $M\xi = ab$, $D\xi = ab^2$



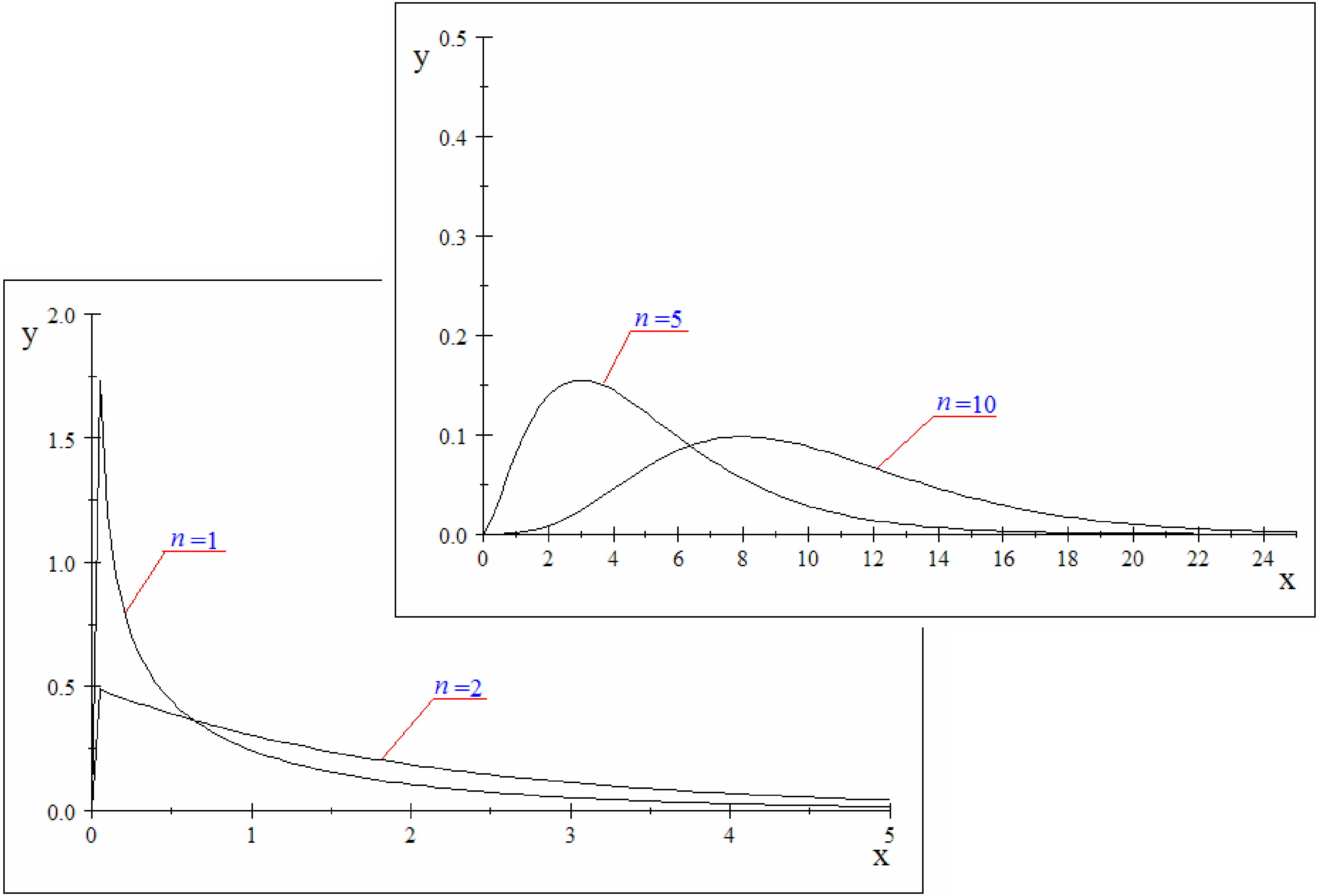
Hi – kvadrātā (χ^2) sadalījums. $$p_{\eta_n}(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \quad x > 0 .$$

Parametru $n > 0$ sauc par brīvības pakāpju skaitu (praktiski lieto sadalījumus ar veselām n vērtībām).

Ja gadījuma lielumi $\xi_1, \xi_2, \ldots, \xi_n$ neatkarīgi un visi standartnormāli sadalīti ($\xi_j \sim N(0,1)$), tad gadījuma lielumam $\eta_n = \sum_{j=1}^{n} \xi_j^2$ ir χ^2 sadalījums ar n brīvības pakāpēm ($\eta_n \sim \chi_n^2$).

$$M\eta_n = M\sum_{j=1}^{n} \xi_j^2 = \sum_{j=1}^{n} M\xi_j^2 = \sum_{j=1}^{n} D\xi_j = n$$

$$D\eta_n = D\sum_{j=1}^{n} \xi_j^2 = \sum_{j=1}^{n} M\xi_j^4 - \sum_{j=1}^{n}\left(M\xi_j^2\right)^2 = 3n - n = 2n$$


y
x
n =1
n =2
n =5
n =10
0.0
0.5
1.0
1.5
2.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5
0 2 4 6 8 10 12 14 16 18 20 22 24

Stjudenta jeb t sadalījums. $$p_{\tau_n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}}\left(1+\frac{1}{n}x^2\right)^{-\frac{n+1}{2}}, \quad -\infty < x < \infty$$

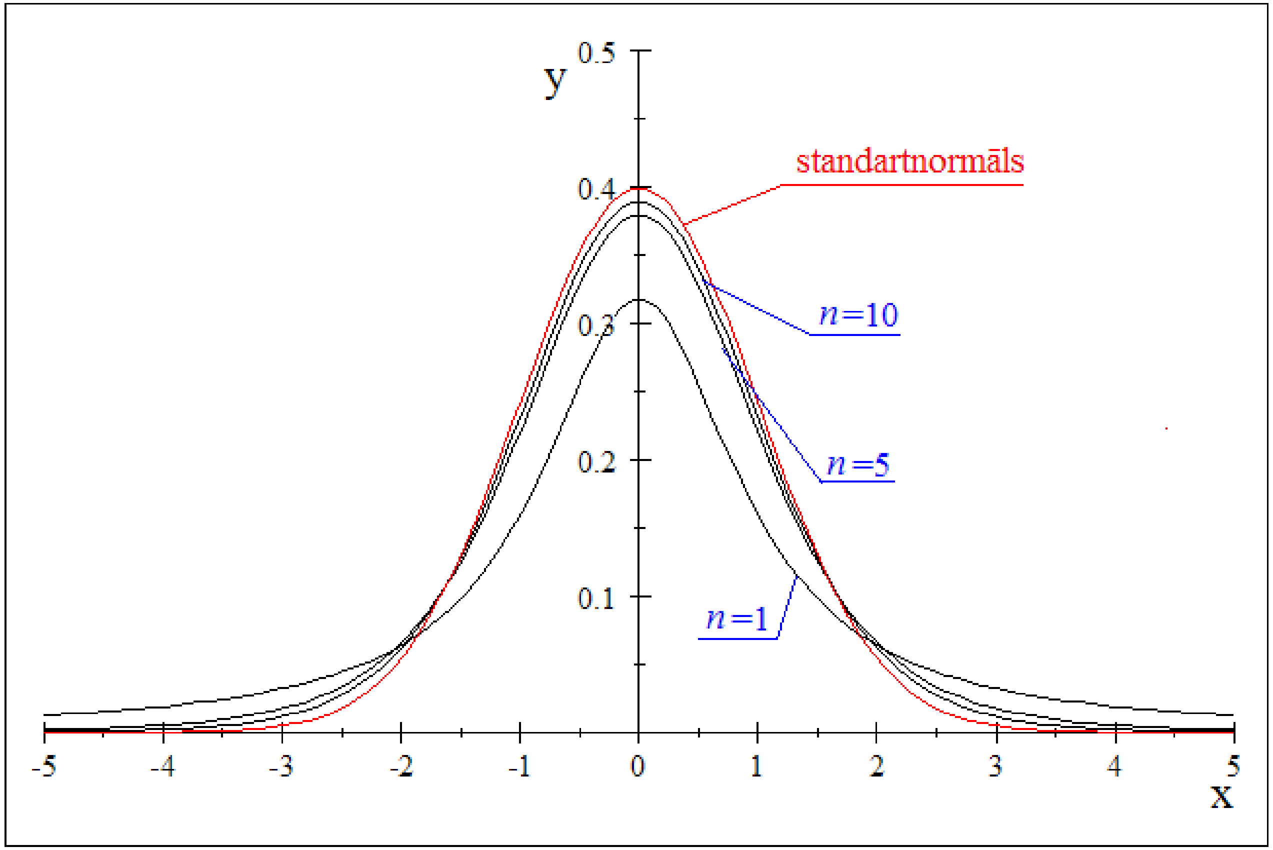
Parametru $n > 0$ sauc par brīvības pakāpju skaitu (praktiski lieto sadalījumus ar veselām n vērtībām).

Ja gadījuma lielumi $\zeta, \xi_1, \xi_2, \dots, \xi_n$ neatkarīgi un visi standartnormāli sadalīti ($\zeta, \xi_j \sim N(0,1)$), tad gadījuma lielumam $$\eta_n = \frac{\zeta}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}\xi_j^2}} = \frac{\zeta}{\sqrt{\frac{1}{n}\eta_n}}$$ ir

t sadalījums ar n brīvības pakāpēm ($\tau_n \sim t_n$).

$$M\tau_n = M\frac{\zeta}{\sqrt{\frac{1}{n}\eta_n}} = 0 \qquad D\tau_n = \begin{cases} \frac{n}{n-2}, & n > 2 \\ \infty, & n \le 2 \end{cases},$$

$$\lim_{n\to\infty} p_{\tau_n}(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$ Stjudenta sadalījums ir asimptotiski standartnormāls.



y
0.5
0.4
0.3
0.2
0.1
standartnormāls
n=10
n=5
n=1
-5
-4
-3
-2
-1
0
1
2
3
4
5
x

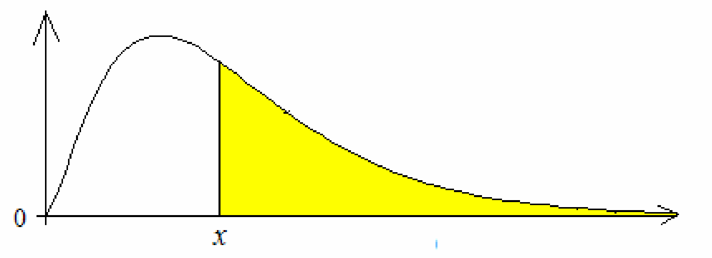
EXCEL iebūvētās statistikas funkcijas

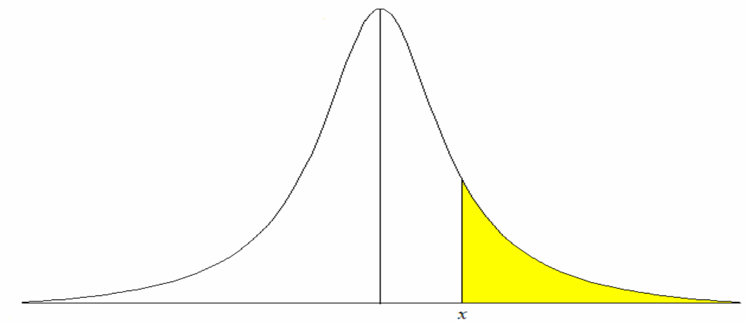
p = NORMDIST(x; a; σ;{1 $\to P(\xi < x)$ vai 0 $\to p(x)$ }),
x = NORMINV(p; a; σ),

p = NORMSDIST(x)
x = NORMSINV(p)

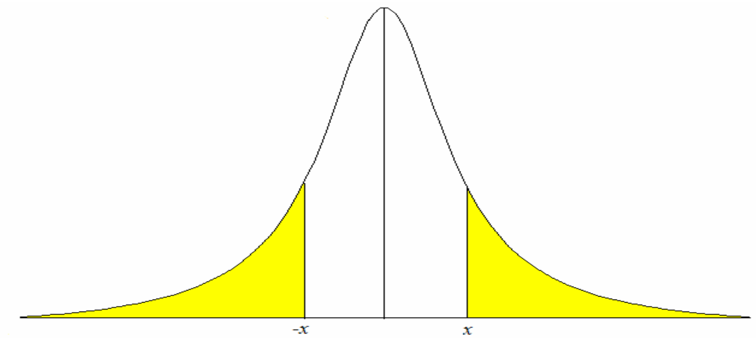
p = CHIDIST(x; *brīvības pakāpju skaits*)

x = CHIINV(*varbūtība*; *brīvības pakāpju skaits*)





p = TDIST(x; *brīvības pakāpju skaits*; 1)
nav inversās funkcijas



p = TDIST(x; *brīvības pakāpju skaits*; 2)

x = TINV(*varbūtība*; *brīvības pakāpju skaits*)

Matemātiskās statistikas elementi.

<u>Mūsu kursa programmas jautājumi:</u> *Novērojumu dati, ģenerālkopas un izlase. Statistiskie novērtējumi. Punktveida novērtējumi. Normāli sadalītas ģenerālkopas parametru ticamības intervāli. Statistisko hipotēžu pārbaude. Pirmā un otrā veida kļūdas. Kritērija jauda un nozīmības līmenis. Neimana-Pīrsona lemma. Baijesa un minimaksa kritēriji.*

Varbūtību teorijas uzdevums: Dota varbūtību telpa $\{\Omega, \mathcal{F}, P\}$. Tajā definēts gadījuma lielums $\xi = \xi(\omega)$, kas attēlo kopu Ω par skaitļu asi $\mathbf{R}^1$ vai kādu tās apakškopu $(\xi : \Omega \to \mathbf{R}^1)$. Nepieciešams, lai $\{\omega : \xi(\omega) < x\} \in \mathcal{F}$. ξ sadalījums uzdots ar tā sadalījuma funkciju $F(x)$ vai sadalījuma blīvuma funkciju (nepārtrauktā gadījumā) $p(x)$, kas ļauj aprēķināt jebkādu ar gadījuma lielumu ξ saistītu notikumu iestāšanās varbūtības. Varbūtību teorija apskata tās matemātiskās metodes, kas ļauj, pilnībā zinot ξ sadalījumu, aprēķināt jebkādu ar ξ saistītu prognožu varbūtības.

Statistiskā nenoteiktība: Dabā praktiski nekad ξ sadalījuma likums nav pilnībā zināms. Var būt zināma sadalījuma klase (piemēram, normālais, eksponenciālais, vienmērīgais, Binomiālais, Puasona, gamma, hi – kvadrātā, Stjudenta utt. sadalījums), bet nav zināmi tā parametri (parametriskais statistikas uzdevums). Var nebūt zināma arī sadalījuma klase (neparametriskais statistikas uzdevums).

Matemātiskās statistikas uzdevums: Balstoties uz gadījuma lieluma ξ novērojumu datiem, mazināt statistisko nenoteiktību – iegūt ziņas par sadalījuma klasi, ja tā nav zināma, vai iegūt ziņas par sadalījuma parametriem klases ietvaros, ja klase jau ir zināma.

Pieņemsim, ka mūsu rīcībā ir gadījuma lieluma ξ, kura sadalījuma likums nav zināms, novērojumu dati, kas iegūti n neatkarīgos novērojumos, kas izdarīti vienādos apstākļos. Par ξ zināms: $\xi \in X$. Zināmi novērojumu dati: $\vec{x} = \{x_1, x_2, .., x_n\} \in X^n$ – izlase no ģenerālkopas, n – izlases apjoms.

Piemēri.

1. Apdrošināšanas sabiedrība, apdrošinot klienta dzīvību, aprēķina apdrošināšanas prēmiju, novērtējot savus vidējos zaudējumus (risku), kas iestājas apdrošināšanas gadījumā: varbūtība × maksājums + administratīvās izmaksas + peļņa. Zināms, ka sabiedrības indivīdu dzīves laiku labi apraksta eksponenciālais sadalījums, balstoties uz ilgstošiem novērojumu datiem, sastādītas mirstības tabulas.

   $$\xi \sim E(\lambda) \text{ – } \lambda \text{ nav zināms}$$

2. Fizikālu lielumu mērījumos kļūda rodas daudzu neatkarīgu vai vāji atkarīgu nejaušību summārās iedarbes rezultātā. Saskaņā ar centrālo robežteorēmu mērījumu kļūdai jāpakļaujas normālajam sadalījumam.

   $$\xi \sim N(a, \sigma^2) \text{ – } a \text{ un } \sigma \text{ nav zināmi}$$

<u>Apzīmējumi</u>: $p(x) = \begin{cases} F'(x), & \text{ja } \xi \text{ nepārtraukts} \\ P(\xi = x), & \text{ja } \xi \text{ diskrēts} \end{cases}$ Nezināmos sadalījuma parametrus apzīmēsim $\theta_1, \theta_2, .., \theta_k$ un rakstīsim kā blīvuma funkcijas argumentus $p(x, \theta)$, kur $\theta = (\theta_1, \theta_2, .., \theta_k)$.

Mazliet vēstures.

Jakobs Bernulli (*Jakob Bernoulli 1654–1705*),

Pjērs Laplass (*Pierre-Simon, Marquis de Laplace 1749 – 1827*),

Karls Gauss (*Johann Carl Friedrich Gauß 1777 – 1855*) mērījumu kļūdu teorija,

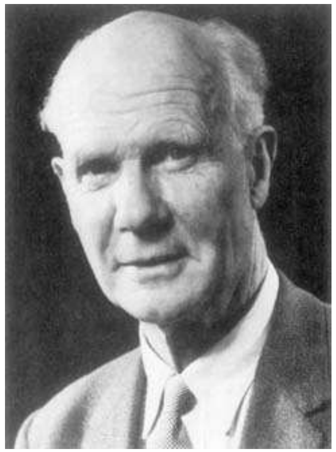
Pafnutijs Čebiševs (*Пафнутий Львович Чебышёв 1821 –1894*),

Andrejs Markovs (*Андрей Андреевич Марков 1856 – 1922*) Markova procesi,

Aleksandrs Ļapunovs (*Александр Михайлович Ляпунов 1857 – 1918*),

Karls Pīrsons (*Karl Pearson* 1857 – 1936) saskaņas kritēriji, pasaulē pirmā matemātiskās statistikas katedra University College Londonā 1911.

Ježi Neimans (*Jerzy Neyman 1894 –1981*),
Egons Pīrsons (*Egon Sharpe Pearson 1895 –1980*) optimāla statistiska kritērija esamības pierādījums. Neimana – Pīrsona lemma



Ronalds Fišers (*Sir Ronald Aylmer Fisher, 1890 – 1962*) maksimālās ticamības metode

ĀbramsValds (*Abraham Wald 1902 – 1950*) pakāpeniskā analīze

Haralds Kramērs (*Harald Cramér 1893 – 1985*)
Kaliampudi Radakrišna Rao (*Calyampudi Radhakrishna Rao 1920*) Rao – Kramēra nevienādība

Džordžs Bokss (*George Edward Pelham Box 1919*)
Gvilims Dženkinss (*Gwilym Meirion Jenkins 1933 – 1982*) Boksa – Dženkinsa metode

trīs "japāņu" guru
Geniki Taguči (*Genichi Taguchi* , 田口 玄一 *1924*) ,

Edvards Demings (*William Edwards Deming 1900 –1993*),
statistiska procesu vadība, Deminga prēmija

$$Quality = \frac{Results\ of\ work\ efforts}{Total\ costs}$$

Džozefs Džurans (*Joseph Moses Juran 1904*)

Deivids Kokss (*Sir David Roxbee Cox 1924*) pētījumi regresiju analīzē

Empīriskā sadalījuma funkcija.

Saranžēsim izlasi $\vec{x} = \left\{ x_1, x_2, .., x_n \right\}$ augošā kārtībā $\vec{x} = \left\{ x(1), x(2), .., x(n) \right\}$, kur $x(1) \le x(2) \le \ldots \le x(n)$. Ranžētu izlasi sauksim par variāciju rindu.

Katram reālam skaitlim x uzdosim gadījuma lielumu $\mu_n(x) = \left| \left\{ j : x_j < x \right\} \right|$. Šeit $\left| \left\{ \cdot \right\} \right|$ apzīmē kopas apjomu (elementu skaitu).

__**Definīcija.**__ Funkciju $F_n(x) = \dfrac{\mu_n(x)}{n}$ sauksim par izlases jeb empīrisko sadalījuma funkciju.

Šī funkcija katram fiksētam x ir gadījuma lielums, kas pieņem vērtības: $0, \dfrac{1}{n}, \dfrac{2}{n}, \ldots, \dfrac{n-1}{n}, 1$

__**Teorēma.**__ Ja $n \to \infty$, tad empīriskā sadalījuma funkcija tiecas uz gadījuma lieluma ξ sadalījuma funkciju pēc varbūtības. $F_n(x) \to F(x)$.

__**Empīriskās sadalījuma funkcijas īpašības**__

1. $F_n(x) \in [0,1]$
2. $F_n(x)$ – gabaliem konstanta nedilstoša funkcija
3. $F_n(x) = 0$, ja $x \le x(1) = x_{\min}$, $F_n(x) = 1$, ja $x > x(n) = x_{\max}$

Histogramma.

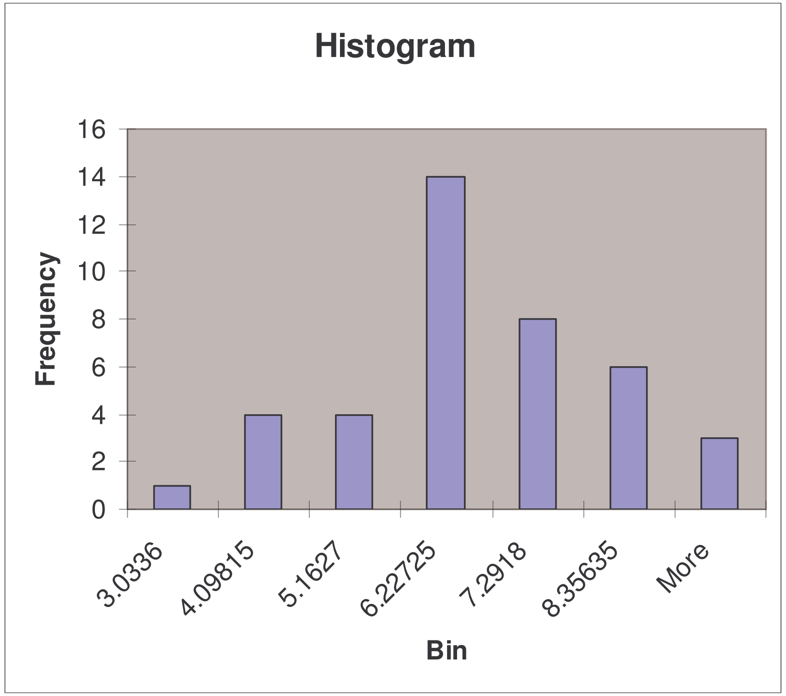
Neranžēta izlase $\vec{x} = \{x_1, x_2, .., x_n\}$, augošā kārtībā ranžēta izlase $\vec{x} = \{x(1) = x_{\min}, x(2), .., x(n) = x_{\max}\}$. Intervālu $[x_{\min}, x_{\max}]$ sadala k vienādās daļās ar garumu $h = \dfrac{x_{\max} - x_{\min}}{k}$. Katram apakšintervālam atrod tajā trāpījušo izlases elementu skaitu n_i. Uz katra intervāla kā pamata konstruē taisnstūri ar augstumu, proporcionālu n_i.

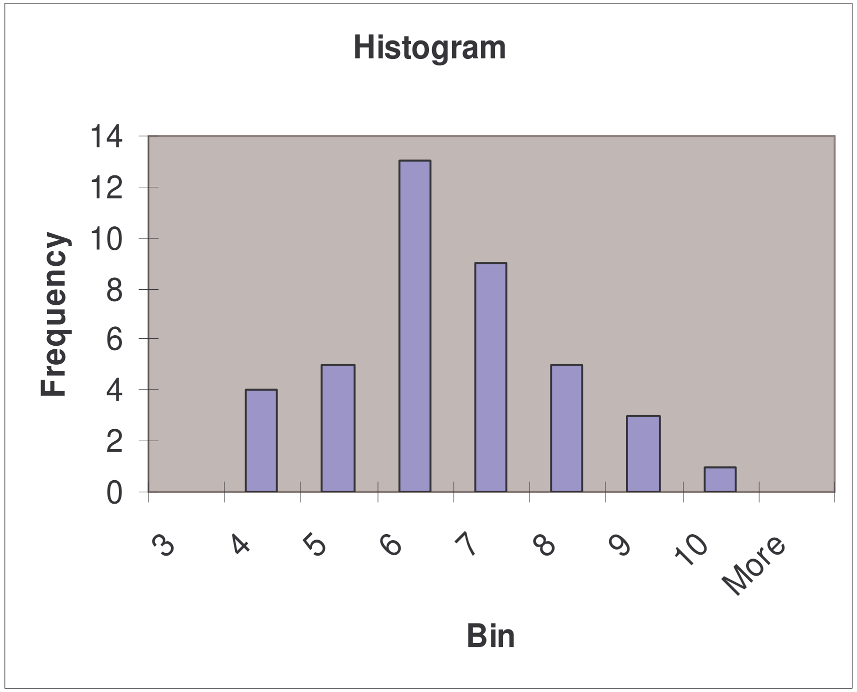
Piezīme. Apakšintervālu skaita k izvēle ir patvaļīga. To nepieciešams izvēlēties tā, lai atklātos izlases raksturīgās īpašības (parasti nelielām izlasēm, lai būtu 5 – 10 novērojumi apakšintervālā, lielām izlasēm – var izvēlēties diezgan brīvi).

Piemērs.

3.03	**4.29**	**5.29**	**5.83**	**5.97**	**6.24**	**6.51**	**7.89**
3.43	**4.36**	**5.70**	**5.83**	**5.97**	**6.24**	**7.33**	**8.16**
3.56	**4.62**	**5.83**	**5.97**	**6.10**	**6.24**	**7.61**	**8.44**
3.96	**4.76**	**5.83**	**5.97**	**6.24**	**6.38**	**7.75**	**8.72**
4.09	**5.29**	**5.83**	**5.97**	**6.24**	**6.51**	**7.89**	**9.42**

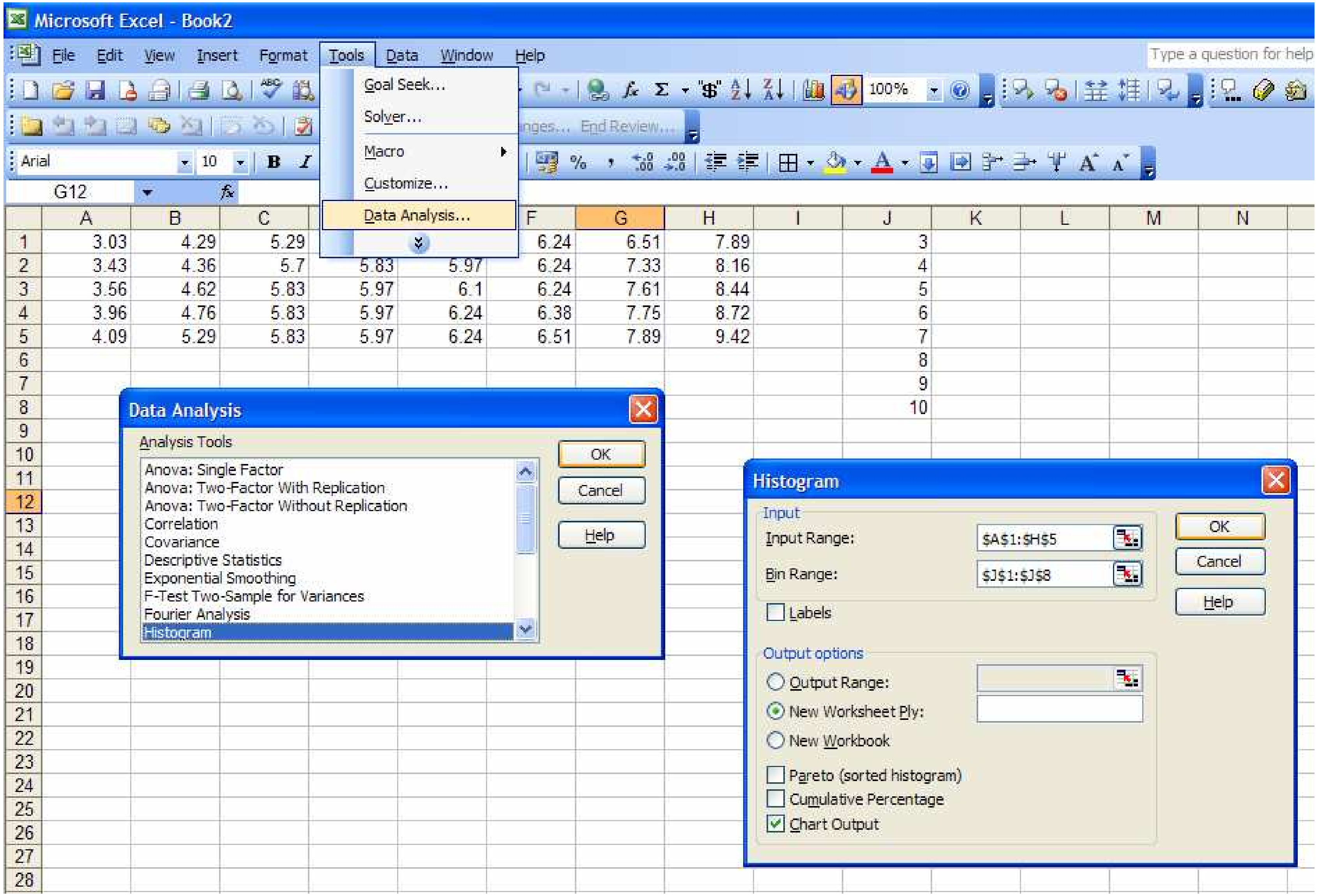
$n = 40$, $x_{\min} = 3.03$, $x_{\max} = 9.42$





Bin	*Frequency*
3.0336	1
4.09815	4
5.1627	4
6.22725	14
7.2918	8
8.35635	6
More	3

Bin	*Frequency*
3	0
4	4
5	5
6	13
7	9
8	5
9	3
10	1
More	0


Microsoft Excel - Book2
File Edit View Insert Format Tools Data Window Help
Type a question for help
Goal Seek...
Solver...
Macro
Customize...
Data Analysis...
Arial 10
G12
A B C F G H I J K L M N
1 3.03 4.29 5.29 6.24 6.51 7.89 3
2 3.43 4.36 5.7 5.83 5.97 6.24 7.33 8.16 4
3 3.56 4.62 5.83 5.97 6.1 6.24 7.61 8.44 5
4 3.96 4.76 5.83 5.97 6.24 6.38 7.75 8.72 6
5 4.09 5.29 5.83 5.97 6.24 6.51 7.89 9.42 7
8
9
10
Data Analysis
Analysis Tools
Anova: Single Factor
Anova: Two-Factor With Replication
Anova: Two-Factor Without Replication
Correlation
Covariance
Descriptive Statistics
Exponential Smoothing
F-Test Two-Sample for Variances
Fourier Analysis
Histogram
OK
Cancel
Help
Histogram
Input
Input Range: $A$1:$H$5
Bin Range: $J$1:$J$8
Labels
Output options
Output Range:
New Worksheet Ply:
New Workbook
Pareto (sorted histogram)
Cumulative Percentage
Chart Output
OK
Cancel
Help

Parametriskās statistikas uzdevumu klasifikācija.

Gadījuma lielums ξ, blīvuma funkcija $p(x,\theta)$, nezināmie parametri $\theta = (\theta_1, \theta_2, .., \theta_k)$.

Dota izlase $\vec{x} = \{x_1, x_2, .., x_n\}$.

1. <u>Nezināmo sadalījuma parametru punktveida novērtējumi.</u>

Atrast tādas izlases $\vec{x}$ funkcijas $\hat{\theta} = \hat{\theta}(\vec{x})$, kuras, ievietojot sadalījuma blīvuma funkcijā, nodrošina vislabāko gadījuma lieluma ξ sadalījuma likuma atbilstību novērojumu datiem.

Statistiskā nenoteiktība formāli novērsta, jo nezināmā θ vietā ir ievietots zināmais $\hat{\theta}$.

<u>Problēma:</u> novērtējums $\hat{\theta}$ un līdz ar to arī $p(x,\hat{\theta})$ ir gadījuma lielums, kura realizācijas atkarīgas no konkrētās izlases.

Svarīgas novērtējumu īpašības:

a) nenovirzītība: $M\hat{\theta} = \theta$ (novērtējuma matemātiskā cerība sakrīt ar nezināmo patieso parametra vērtību),

b) efektivitāte: novērtējumam $\hat{\theta}$ ir mazākā dispersija $D\hat{\theta}$, salīdzinot ar jebkuru citu iespējamu parametra θ novērtējumu,

a) + b) nenovirzītu efektīvu novērtējumu sauc par optimālu,

c) saturīgums: ja $\hat{\theta}_n$ ir parametra θ novērtējums no izlases ar apjomu n, tad $\lim\limits_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| < \varepsilon\right) = 1$ katram $\varepsilon > 0$ (ja $n \to \infty$, $\hat{\theta}_n \to \theta$ pēc varbūtības).

<u>Izlases vidējā vērtība</u> $\overline{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$ ir nenovirzīts matemātiskās cerības novērtējums.

$M\overline{x} = M\ \frac{1}{n}\sum\limits_{i=1}^{n} x_i = \frac{1}{n}\sum\limits_{i=1}^{n} Mx_i = \frac{1}{n}\sum\limits_{i=1}^{n} M\,\xi = \frac{1}{n} nM\,\xi = M\,\xi$. Var pierādīt, ka $\overline{x}$ ir arī efektīvs, tātad optimāls matemātiskās cerības novērtējums.

Var pierādīt, ka <u>izlases dispersija</u> $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ ir novirzīts dispersijas novērtējums, ka $MS^2 = \frac{n-1}{n} D\xi$, tātad $MS^2 \neq D\xi$. No šejienes izriet, ka nenovirzīts būs novērtējums $s^2 = \frac{n}{n-1} S^2$, jeb $s^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ - izlabotā izlases dispersija.

Savukārt redzams, ka $DS^2 < Ds^2$. Tātad s^2 nav efektīvs novērtējums. T.i., dispersijai neeksistē optimāls novērtējums.

Vat parādīt, ka gan $\overline{x}$, gan s^2, gan S^2 ir saturīgi novērtējumi.

Ja nezināmais parametrs nav ne matemātiskā cerība, ne dispersija, eksistē universālas punktveida novērtējumu iegūšanas metodes – <u>momentu metode</u> un <u>maksimālās ticamības metode.</u>

Momentu metode.

Gadījuma lielums ξ, blīvuma funkcija $p(x,\theta)$, nezināmie parametri $\theta=(\theta_1,\theta_2,..,\theta_k)$.

Sadalījuma teorētiskie momenti $M\xi^m$ un $M\left(\xi - M\xi\right)^m$, ja θ nezināms, nav skaitlis,

bet θ funkcija. Pierakstīsim: $M_\theta \xi^m$ un $M_\theta\left(\xi - M\xi\right)^m$. Atbilstošie empīriskie

(izlases) momenti ir skaitļi: $m_m^* = \frac{1}{n}\sum_{i=1}^{n} x_i^m$ $\quad\quad \mu_m^* = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^m$

Momentu vienādojumus iegūst, ja pielīdzina atbilstošos teorētiskos un empīriskos

momentus: $\quad M_\theta \xi^m = m_m^* \quad\quad M_\theta\left(\xi - M\xi\right)^m = \mu_m^*$

<u>Apgalvojums.</u> Ja sadalījumu klase satur k nezināmus parametrus, tad jebkura patvaļīga k dažādu momentu vienādojumu sistēma ir saderīga, tai eksistē viens vienīgs atrisinājums (viens un tas pats jebkurai sistēmai).

<u>Piemērs.</u> Eksponenciālais sadalījums $p(x,\theta)=\begin{cases}\theta e^{-\theta x}, & x>0,\\ 0, & x<0\end{cases}$. Zinām, ka $M_\theta \xi = \frac{1}{\theta}$.

Dota izlase $\vec{x} = \{15, 28, 18, 9, 36\}$. $\quad \overline{x} = 21.2$

Vajadzīgs viens momentu vienādojums. Piemēram: $M_\theta \xi = m_1^* = \overline{x}$

$$\frac{1}{\theta} = 21.2 \quad\quad \hat{\theta} = \frac{1}{21.2} \approx 0.047$$

<u>Piezīme.</u> Ar momentu metodi iegūtais novērtējums var būt novirzīts un neefektīvs, bet tas ir saturīgs.

<u>Maksimālās ticamības metode.</u>

Gadījuma lielums ξ , blīvuma funkcija $p(x,\theta)$, nezināmie parametri $\theta=(\theta_1,\theta_2,..,\theta_k)$.

Definēsim t.s. ticamības funkciju $p(\vec{x},\theta)=\prod_{i=1}^{n} p(x_i,\theta)$. Diskrētā gadījumā tā ir izlases varbūtība, nepārtrauktā – izlases blīvuma funkcija.

<u>Definīcija.</u> Par parametra $\theta=(\theta_1,\theta_2,..,\theta_k)$ maksimālās ticamības novērtējumu sauc to parametra vērtību, pie kuras ticamības funkcija $p(\vec{x},\theta)$ sasniedz globālo maksimumu.

Ekstrēma eksistences nepieciešamais nosacījums: $\frac{\partial p(\vec{x},\theta)}{\partial\theta_j}=0$. Reizinājuma atvasinājums – sarežģīts. Sistēmai $\frac{\partial \ln p(\vec{x},\theta)}{\partial\theta_j}=0$ ir tas pats atrisinājums, jo ln*x* ir monotona funkcija. Šos vienādojumus sauc par ticamības vienādojumiem.

Būtu jāpārbauda pietiekamie nosacījumi, bet ir spēkā apgalvojums – ticamības vienādojumu sistēmai eksistē viens vienīgs atrisinājums, un tas ir ticamības funkcijas globālais maksimums.

Maksimālās ticamības novērtējumi ir saturīgi un asimptotiski efektīvi. Ja parametram θ eksistē optimāls novērtējums, tas tiek iegūts ar maksimālās ticamības metodi.

Piemērs. $\xi \sim N(\theta_1, \theta_2{}^2)$, $\vec{x} = \{x_1, x_2, .., x_n\}$, $p(x_i, \theta_1, \theta_2) = \dfrac{1}{\sqrt{2\pi}\theta_2} e^{-\frac{(x_i-\theta_1)^2}{2\theta_2^2}}$

$$p(\vec{x}, \theta_1, \theta_2) = \left(\frac{1}{\sqrt{2\pi}\theta_2}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{(x_i - \theta_1)^2}{2\theta_2^2}\right)$$

$$\ln p(\vec{x}, \theta_1, \theta_2) = -\frac{n}{2}\ln 2\pi - n\ln\theta_2 - \frac{1}{2\theta_2^2}\sum_{i=1}^{n}(x_i - \theta_1)^2$$

$$\begin{cases} \dfrac{\partial \ln p(\vec{x}, \theta_1, \theta_2)}{\partial \theta_1} = \dfrac{1}{\theta_2^2}\displaystyle\sum_{i=1}^{n}(x_i - \theta_1) = 0 \\ \dfrac{\partial \ln p(\vec{x}, \theta_1, \theta_2)}{\partial \theta_2} = -\dfrac{n}{\theta_2} + \dfrac{1}{\theta_2^3}\displaystyle\sum_{i=1}^{n}(x_i - \theta_1)^2 = 0 \end{cases}$$

$$\begin{cases} \displaystyle\sum_{i=1}^{n} x_i = n\theta_1 \qquad \rightarrow \qquad \hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x} \\ \dfrac{1}{\theta_2^2}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2 = n \qquad \rightarrow \qquad \hat{\theta}_2^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = S^2 \end{cases}$$

Galvenais punktveida novērtēšanas trūkums – strādājot ar sadalījumu $p(x,\hat{\theta})$ patiesā sadalījuma $p(x,\theta)$ vietā, nepieciešams novērtēt kļūdas lielumu. Jo vairāk darbību tiek izdarīts, izmantojot $p(x,\hat{\theta})$, jo lielākas kļūdas, līdz beidzot rezultātiem vairs nav nekādas ticamības.

2. **<u>Nezināmo sadalījuma parametru intervālie novērtējumi.</u>** Uzdevums atrast optimālu intervālu uz skaitļu ass, kas ietvertu sevī patieso nezināmo parametra θ vērtību ar iepriekš fiksētu, no θ neatkarīgu, pietiekami lielu varbūtību.

Uzdevums – atrast tādas izlases $\vec{x}$ funkcijas $\underline{\theta}(\vec{x})$ un $\overline{\theta}(\vec{x})$, lai $P(\underline{\theta}(\vec{x}) < \theta < \overline{\theta}(\vec{x})) = 1-2\alpha$. Intervālu $\left[\underline{\theta}(\vec{x}), \overline{\theta}(\vec{x})\right]$ sauc par parametra θ ticamības intervālu ar ticamības varbūtību (ticamības līmeni, drošumu) $1-2\alpha$. Šī varbūtība ir konstante, kas nav atkarīga no nezināmās parametra vērtības θ. Ar vienu un to pašu drošumu var izveidot bezgalīgi daudz dažādu intervālu. Ticamības intervāla optimalitāti saprot tādā nozīmē, ka intervālam ir minimālais garums pie uzdotā līmeņa $1-2\alpha$.

<u>Piezīme.</u> Ne visiem sadalījumiem eksistē optimāli parametru ticamības intervāli.

Normālā sadalījuma matemātiskās cerības novērtējums, ja tā dispersija ir zināma.

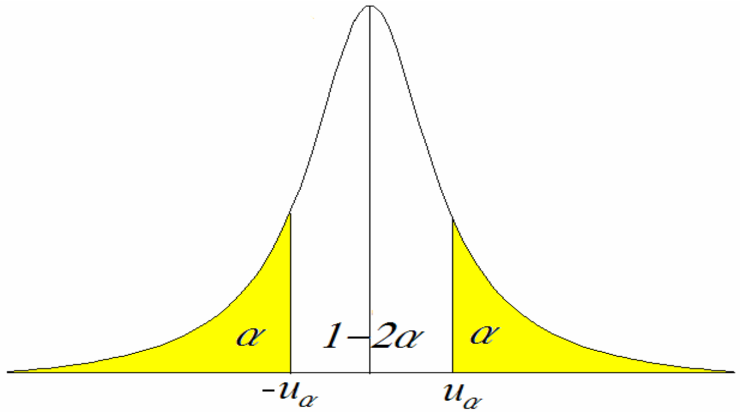
$\xi \sim N(\theta_1, \sigma^2)$, θ_1 nezināmā matemātiskā cerība, σ zināmā vidējā kvadrātiskā novirze.

Jau zināms, ka izlases vidējās vērtības matemātiskā cerība vienāda ar ξ matemātisko cerību: $M\overline{x} = \theta_1$. Atradīsim

$$D\overline{x} = D\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n^2}D\sum_{i=1}^{n} x_i = \frac{1}{n^2}\sum_{i=1}^{n} Dx_i = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.$$

Gadījuma lielums $\dfrac{\overline{x}-\theta_1}{\sigma/\sqrt{n}} \sim N(0,1)$

$$P\left(\left|\frac{\overline{x}-\theta_1}{\sigma/\sqrt{n}}\right| < u_\alpha\right) = 1-2\alpha$$



$$1-2\alpha = P\left(-u_\alpha < \frac{\overline{x}-\theta_1}{\sigma/\sqrt{n}} < u_\alpha\right) = P\left(-u_\alpha\frac{\sigma}{\sqrt{n}} < \overline{x}-\theta_1 < u_\alpha\frac{\sigma}{\sqrt{n}}\right) =$$

$$= P\left(-\overline{x}-u_\alpha\frac{\sigma}{\sqrt{n}} < -\theta_1 < -\overline{x}+u_\alpha\frac{\sigma}{\sqrt{n}}\right) = P\left(\overline{x}-u_\alpha\frac{\sigma}{\sqrt{n}} < \theta_1 < \overline{x}+u_\alpha\frac{\sigma}{\sqrt{n}}\right)$$

Piemērs. $\xi \sim N(\theta_1, 4)$, izvēlēsimies $1-2\alpha = 0.9$

izlase $\vec{x} = \{1.2, -2.7, 1.8, -3.0, -1.5, -0.5, 0.2, 0.1, -0.4, -0.9\}$,

$\overline{x} = -0.57 \qquad \alpha = 0.05 \qquad u_{0.05} = 1.645$

$$P\left(-0.57 - 1.645\frac{2}{\sqrt{10}} < \theta_1 < -0.57 + 1.645\frac{2}{\sqrt{10}}\right) = P\left(-1.61 < \theta_1 < 0.47\right) = 0.9$$

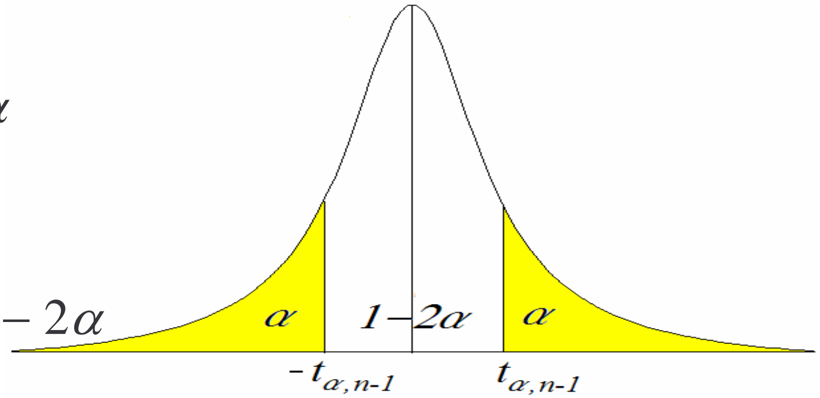
Normālā sadalījuma matemātiskās cerības novērtējums, ja tā dispersija nav zināma.

$\xi \sim N(\theta_1, \theta_2^{\,2})$, θ_1 nezināmā matemātiskā cerība, θ_2 nezināmā vidējā kvadrātiskā novirze.

Var parādīt, ka gadījuma lielumam $\tau = \dfrac{\overline{x} - \theta_1}{s/\sqrt{n}}$ ir Stjudenta sadalījums ar $n-1$ brīvības pakāpi.

$$P\left(-t_{\alpha, n-1} < \frac{\overline{x} - \theta_1}{s/\sqrt{n}} < t_{\alpha, n-1} \right) = 1 - 2\alpha$$

$$P\left(\overline{x} - t_{\alpha, n-1} \frac{s}{\sqrt{n}} < \theta_1 < \overline{x} + t_{\alpha, n-1} \frac{s}{\sqrt{n}} \right) = 1 - 2\alpha$$



Piemērs. $\xi \sim N(\theta_1, \theta_2^2)$, izvēlēsimies $1 - 2\alpha = 0.9 \qquad \alpha = 0.05$

Tā pati izlase $\vec{x} = \{1.2, -2.7, 1.8, -3.0, -1.5, -0.5, 0.2, 0.1, -0.4, -0.9\}$,

$$\overline{x} = -0.57 \qquad s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2} \approx 1.54 \qquad t_{0.05,\,9} = 1.83$$

$$P\left(-0.57 - 1.83 \frac{1.54}{\sqrt{10}} < \theta_1 < -0.57 + 1.83 \frac{1.54}{\sqrt{10}} \right) = P\left(-1.46 < \theta_1 < 0.32\right) = 0.9$$

Normālā sadalījuma dispersijas novērtējums, ja tā matemātiskā cerība ir zināma.

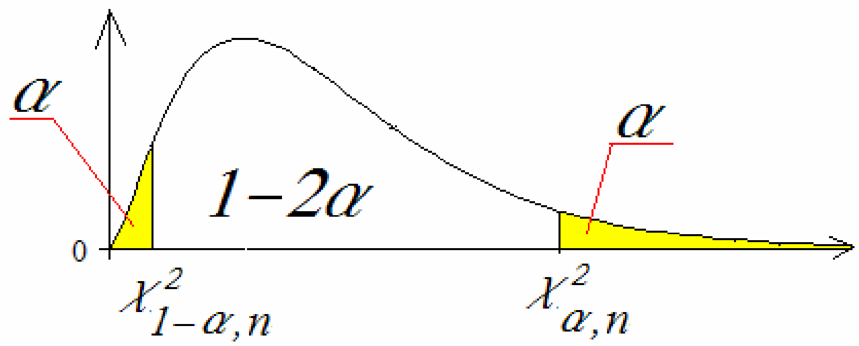
$\xi \sim N(a, \theta_2^{\,2})$, a zināmā matemātiskā cerība, θ_2 nezināmā vidējā kvadrātiskā novirze.

Nav jēgas lietos izlaboto izlases vidējo kvadrātisko novirzi s^2, jo $\bar{x}$ vietā ir a.

Lietosim $\hat{s}^2 = \sum_{i=1}^{n} (x_i - a)^2$. Gadījuma lielumam $\dfrac{\hat{s}^2}{\theta_2^2}$ ir χ^2 sadalījums ar n brīvības pakāpēm.

$$P\left(\chi^2_{1-\alpha,n} < \frac{\hat{s}^2}{\theta_2^2} < \chi^2_{\alpha,n} \right) = 1 - 2\alpha$$

$$P\left(\frac{\hat{s}}{\chi_{\alpha,n}} < \theta_2 < \frac{\hat{s}}{\chi_{1-\alpha,n}} \right) = 1 - 2\alpha$$



Piezīme. Šis ticamības intervāls patiesībā nav optimāls, jo sadalījuma blīvuma funkcija nav simetriska. Eksistē mazliet mazāka garuma intervāls, bet tā meklēšanas „spēle nav sveču vērta”.

Piemērs. $\xi \sim N(a, \theta_2^2)$, $a = -0.5$, $\quad 1 - 2\alpha = 0.9 \quad \alpha = 0.05$

Tā pati izlase $\vec{x} = \{1.2, -2.7, 1.8, -3.0, -1.5, -0.5, 0.2, 0.1, -0.4, -0.9\}$,

$\hat{s} = \sqrt{\hat{s}^2} \approx 4.61$, $\chi_{0.95,10} = \sqrt{\chi^2_{0.95,10}} \approx \sqrt{3.94} \approx 1.99$, $\chi_{0.05,10} = \sqrt{\chi^2_{0.05,10}} \approx \sqrt{18.31} \approx 4.28$

$$P\left(\frac{4.61}{4.28} < \theta_2 < \frac{4.61}{1.99} \right) = P\left(1.08 < \theta_2 < 2.32\right) = 0.9$$

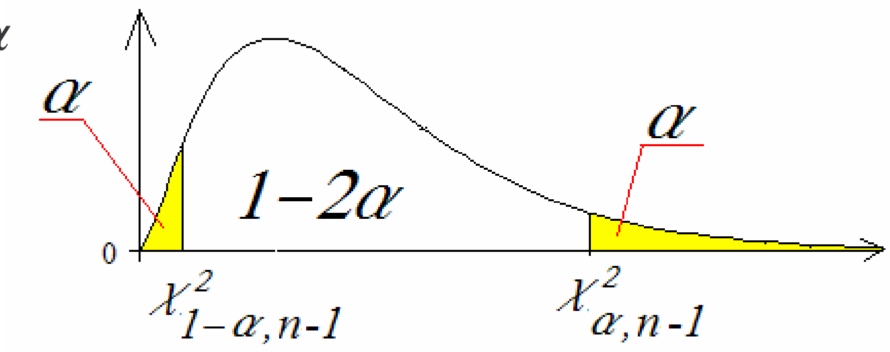
Normālā sadalījuma dispersijas novērtējums, ja tā matemātiskā cerība nav zināma.

$\xi \sim N(\theta_1, \theta_2^{\,2})$, θ_1 nezināmā matemātiskā cerība, θ_2 nezināmā vidējā kvadrātiskā novirze.

Gadījuma lielumam $\dfrac{(n-1)s^2}{\theta_2^2}$ ir χ^2 sadalījums ar $n-1$ brīvības pakāpi.

$$P\left(\chi^2_{1-\alpha\,,\,n-1} < \frac{(n-1)s^2}{\theta_2^2} < \chi^2_{\alpha\,,\,n-1} \right) = 1 - 2\alpha$$

$$P\left(\frac{\sqrt{n-1}\;s}{\chi_{\alpha\,,\,n-1}} < \theta_2 < \frac{\sqrt{n-1}\;s}{\chi_{1-\alpha\,,\,n-1}} \right) = 1 - 2\alpha$$



Piezīme. Arī šis ticamības intervāls nav optimāls.

Piemērs. $\xi \sim N(\theta_1, \theta_2^2)$, $\bar{x} = -0.57$, $1 - 2\alpha = 0.9$ $\alpha = 0.05$

Tā pati izlase $\vec{x} = \{1.2, -2.7, 1.8, -3.0, -1.5, -0.5, 0.2, 0.1, -0.4, -0.9\}$,

$s = \sqrt{s^2} \approx 1.54$, $\chi_{0.95,9} = \sqrt{\chi^2_{0.95,9}} \approx \sqrt{3.33} \approx 1.82$, $\chi_{0.05,10} = \sqrt{\chi^2_{0.05,10}} \approx \sqrt{16.92} \approx 4.11$

$$P\left(\frac{3\cdot 1.54}{4.11} < \theta_2 < \frac{3\cdot 1.54}{1.82} \right) = P\left(1.12 < \theta_2 < 2.54\right) = 0.9$$