

STATISTILINE ANDMETÖÖTLUS MÕÕTMISTULEMUSTE TÖÖTLEMINE LOENGUD 1997

TEHNILISED MÄRKUSED:

Tekst peab olema hõlpsalt 1.5 või 2 korda vähendatav. Selleks reeglid:

1. Reavahe defineeritud ainult kui “multiple”. Standard = 1.
2. Kirja mõõdud ainult $N*6$: 12, 18, 24, 30.

Normaalkiri: Times 18.

Pealkirjad: Times 24.

Dokument STAT97FK hargneb kaheks: STAT97F ja STAT97K.

STATISTILINE ANDME- TÖÖTLUS

MÕÕTMIS- TULEMUSTE TÖÖTLEMINE

SISSEJUHATUS

Viis esimest peatükki on mõlema kursuse jaoks ühised ja nende sisuks on *matemaatilise statistika* mõisted ning meetod.

1672. a. *statistika* = riigiõpetus, poliitiline aritmeetika.

Mereste, U. Statistika üldteooria. 496 lk.

Massnähtuste teooria.

1. Andmehõive. 2. Andmestik. 3. Andmetöötlusõpetus.

Vene matemaatikaentsüklopeedias statistika = statistik.

Tiit, E. jt. Tõenäosusteooria ja matemaatiline statistika. 472 lk.

Oluline lähtekoht: Gaussi mõõtmisvigade teooria 1809.

Üldine ülesanne:

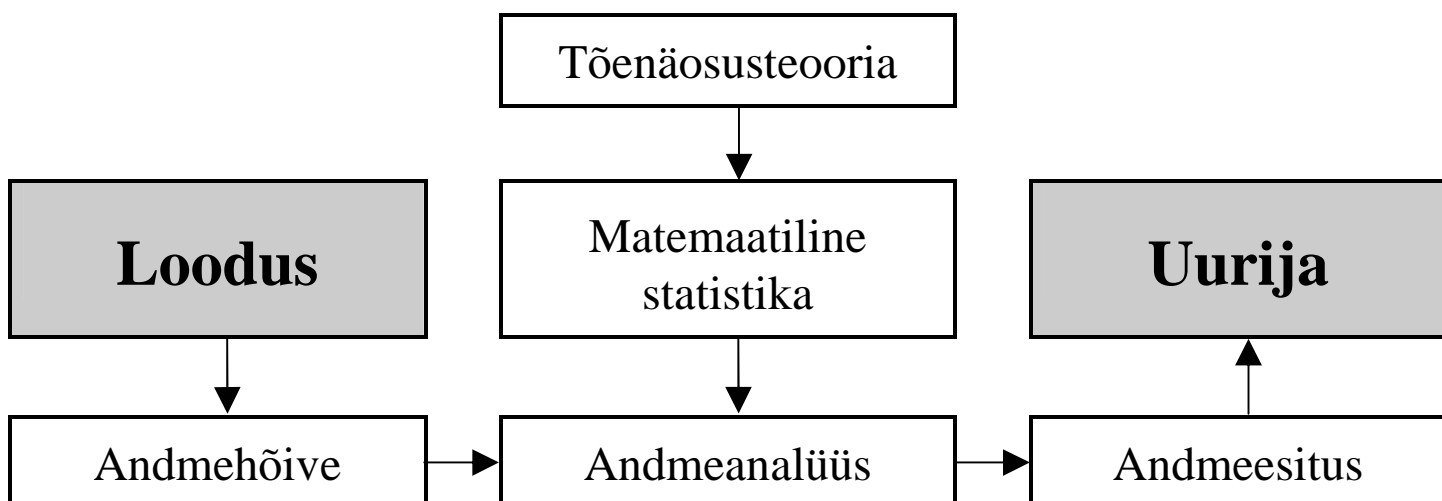
antud – mõõtmised, vaatlused, küsitlused,
leida – järeldused, otsustused.

Osaülesannete meetod:

mõõtmised \Rightarrow arvud \Rightarrow tuletatud arvud \Rightarrow järeldused
(Võrdlus: Tartu \Rightarrow Põltsamaa \Rightarrow Paide \Rightarrow Tallinn)

Osaülesande absolutiseerimise oht ja juhtmõte:

arvutamise eesmärgiks ei ole arvud, vaid arusaamine.



1. Statistilise analüüsi lähteandmed

Objektid, tunnused ja väärtused. Mõõtmine.

Mõõtskaala: tunnuse võimalike väärtuste&nimede hulk.

Nimeskaala ja järjeskaala.

Aditiivne tunnus ja meetriline skaala. Skaala metriseerimine.
(näide: koolihinnete aritmeetika)

Potentsiaalid ja vaheskaala

(näited: kellaaeg, Celsiuse skaala, energia).

Skalaarandmed ja tinglikud skalaarandmed.

Mitmemõõtmelised andmed:

andmetabel,
geomeetiline interpretatsioon,
koordinaaditeisendused,
meetrika.

Vektorandmed ja tinglikud vektorandmed

(probleemi näide: amplituudspekter ja võimsusspekter)

Lünklikud andmed.

Pidev ja diskreetne skaala. Väärtusklassifitseerimine.

Aegread.

Andmete esitus arvutis:

fikseeritud formaat ja mastabeerimine,
eraldajatega jada.

Andmestruktuurid: jadafail ja massiiv.

On-line andmed.

Andmetabeli näidis

Average solar radiation (watts/sq.m.) for selected cities

Month:	jul	aug	sep	oct	nov	dec	jan	feb	mar	apr	may	jun	ave	lat
Atlanta, GA	257	246	201	166	30	102	106	140	184	236	258	271	192	34.0
Barrow, AK	208	123	56	20	0	0	0	18	87	184	248	256	100	71.0
Bismark, ND	296	251	185	132	78	60	76	121	170	217	267	284	178	47.0
Boise, ID	324	275	221	152	88	60	69	113	164	235	284	309	191	43.5
Boston, MA	240	206	165	115	70	58	67	96	142	176	228	242	150	42.5
Caribou, ME	246	218	161	102	53	51	66	111	178	194	229	232	153	47.0
Cleveland, OH	267	239	182	127	68	56	60	87	151	182	253	271	162	41.5
Dodge City, KS	311	287	239	184	138	113	123	153	202	256	275	315	216	38.0
El Paso, TX	324	309	278	224	178	151	160	209	266	317	346	353	260	32.0
Fresno, CA	323	293	243	182	117	77	90	143	212	264	308	337	216	37.0
Greensboro, NC	263	235	197	156	118	95	97	134	171	227	257	273	185	36.0
Honolulu, HI	305	293	271	245	208	176	175	200	234	262	300	297	247	21.0
Little Rock, AR	270	250	214	167	118	91	96	127	173	220	256	272	188	35.0
Miami, FL	260	246	216	188	171	154	166	201	238	263	267	257	219	26.0
New York, NY	251	238	175	127	77	62	71	102	151	183	220	255	159	41.0
Omaha, NE	275	252	192	142	96	80	99	134	172	224	248	272	182	21.0
Rapid City, SD	288	262	208	152	99	76	90	135	193	235	259	287	190	44.0
Seattle, WA	242	209	150	84	44	29	34	60	118	174	216	228	132	47.5
Tucson, AZ	304	286	281	216	172	144	151	195	264	322	358	343	253	41.0
Washington, DC	267	190	196	145	75	64	101	124	153	182	215	247	163	39.0

2. Matemaatilise statistika üldmõisted, ülesanded ja meetodid

Üldkogum ja valim.

Lõplik ja lõpmatu üldkogum.

Juhuslik valim.

(7, 2, 11, 5, 12, 6, 8, 7, 5)

Variatsioonrida.

(2, 5, 5, 6, 7, 7, 8, 11, 12)

Statistikud.

Empiirilised momendid $\sum_{i=1}^n x_i^k / n$

$\sum_{i=1}^n (x_i - \mu)^k / n$???

Järkstatistikud.

Tõenäosusteoreetilised mudelid.

Ühtlane jaotus.

Poissoni jaotus

Normaaljaotus

Cauchy jaotus.

$$P(x) = \frac{\mu^x}{x!} e^{-\mu}$$

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

$$f(x) = \frac{\lambda}{\pi(\lambda^2 + (x - x_0)^2)}$$

Matemaatilise statistika ülesanded.

Hüpoteeside kontrollimine.

Tõenäosusjaotuse karakteristikute mudelivaba hindamine.

Tõenäosusjaotuse mudeli valimine.

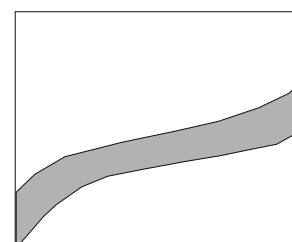
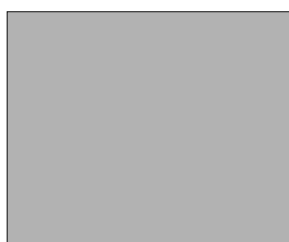
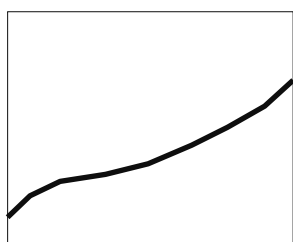
Parameetrite hindamine.

Mudelite konstrueerimine.

Lihtne probleem: Kuhjas on 1 tonn kartuleid. Mitu kartulit on kuhjas?

Parameetrilised, mitteparameetrilised ja robustsed meetodid. (1964/1981)

Jaotuste
ruum:



Statistikute tõenäosusjaotuste meetod.

Näide: aritmeetiline keskmine

Program Näide;

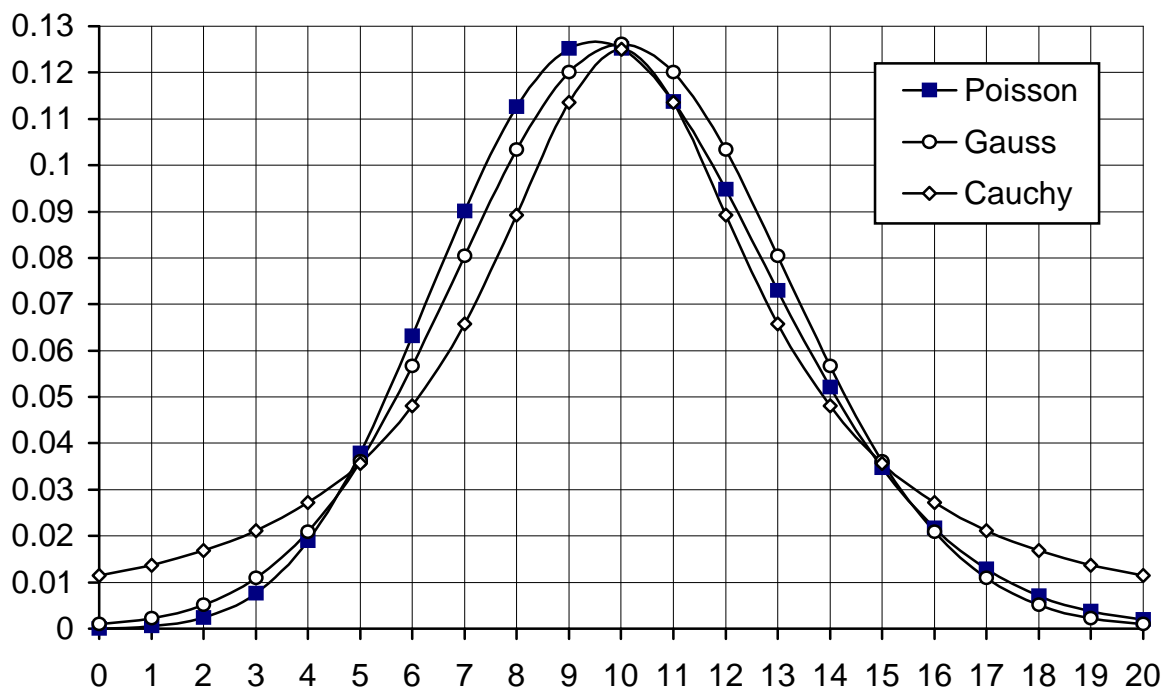
```

var i : integer;
    x, y, f, a, s, ss, g, c, y0, g0 : real;
    t : text;

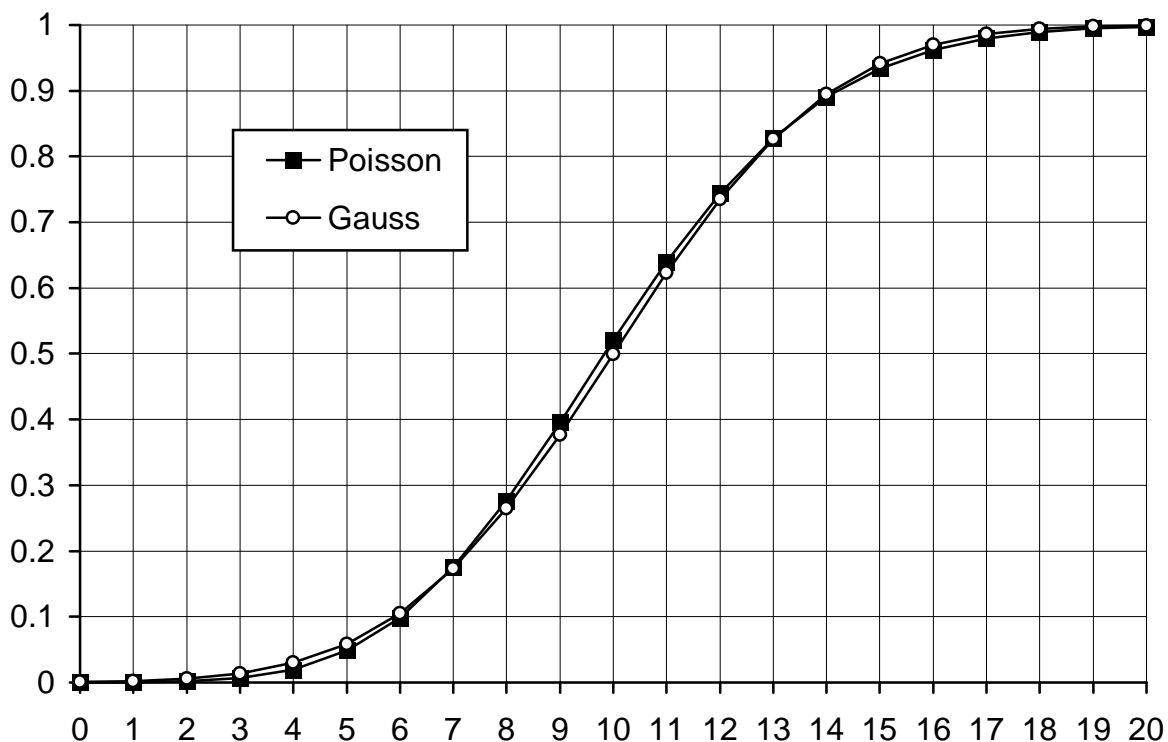
begin
assign (t, 'c:\b\poisson.tab'); rewrite (t);
f := 1; a := 1; y := 0; g := 0;
s:= 0; ss := 0.0004; y0 := 0; g0 := 0;
for i := 0 to 20 do begin x := i;
    if x > 0 then begin a := a * 10; f := f * x end;
    y0 := y; y := (a/f)/exp(10);
    s := s + (y + y0)/2;
    g0 := g; g := exp(-(x - 10)*(x - 10)/20) / 7.926655;
    ss := ss + (g + g0)/2;
    c := 10 / (8*(10 + (x - 10)*(x - 10)));
    writeln (t, i:5, y:10:4, s:10:4, g:10:4, ss:10:4, c:10:4);
end;
close (t);
readln;
end.

```

x	Poisson		Gauss		Cauchy
	P(x)	F(x)	f(x)	F(x)	f(x)
0	0.0000	0.0000	0.0009	0.0008	0.0114
1	0.0005	0.0003	0.0022	0.0023	0.0137
2	0.0023	0.0016	0.0051	0.0060	0.0169
3	0.0076	0.0066	0.0109	0.0140	0.0212
4	0.0189	0.0198	0.0209	0.0299	0.0272
5	0.0378	0.0482	0.0361	0.0584	0.0357
6	0.0631	0.0986	0.0567	0.1048	0.0481
7	0.0901	0.1752	0.0804	0.1734	0.0658
8	0.1126	0.2765	0.1033	0.2652	0.0893
9	0.1251	0.3954	0.1200	0.3769	0.1136
10	0.1251	0.5205	0.1262	0.5000	0.1250
11	0.1137	0.6399	0.1200	0.6231	0.1136
12	0.0948	0.7442	0.1033	0.7347	0.0893
13	0.0729	0.8280	0.0804	0.8266	0.0658
14	0.0521	0.8905	0.0567	0.8951	0.0481
15	0.0347	0.9339	0.0361	0.9415	0.0357
16	0.0217	0.9621	0.0209	0.9700	0.0272
17	0.0128	0.9793	0.0109	0.9859	0.0212
18	0.0071	0.9893	0.0051	0.9939	0.0169
19	0.0037	0.9947	0.0022	0.9976	0.0137
20	0.0019	0.9975	0.0009	0.9991	0.0114



Tihedusfunktsioon $f(x)$ (Poissoni jaotuse puhul tinglik)



Jaotusfunktsioon $F(x)$

3. Punkthinnangud

Hinnangufunktsiooni omadused.

Valim $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, hinnatav ξ , hinnangufunktsioon $T(\xi)$.

Hinnangu nihe $ET(\mathbf{x}) - \xi$.

Hinnangu mõjususe ehk konsistentsuse.

Hinnangu hajuvus ja efektiivsus $E\{T(\mathbf{x}) - \xi\}^2$.

Hinnangu arvutamise kulud.

Parima hinnangufunktsiooni probleem.

Suurima tõepära meetod.

Juhusliku suuruse tõenäosustihedus $f_x(x; \xi)$.

Valimi tõenäosustihedus $f_x(\mathbf{x}; \xi) = \prod_{i=1}^n f_x(x_i; \xi)$.

Konkreetse hinnangu tõepärasus $L(\xi; \mathbf{x}) = f_x(\mathbf{x}; \xi)$.

Ülesanne: valida ξ niiviisi, et $L = L_{\max}$, kus \mathbf{x} on fikseeritud.

Võte: $l(\xi; \mathbf{x}) = \ln L(\xi; \mathbf{x}) = \sum_{i=1}^n \ln f_x(x_i; \xi)$.

Näide: Poissoni jaotuse parameetri μ hinnang λ

$$l(\lambda; \mathbf{x}) = \sum_{i=1}^n \ln \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) = \sum_{i=1}^n (x_i \ln \lambda - \ln x_i - \lambda),$$

lahendada võrrand $dl/d\lambda = 0$.

Vähimruutude meetod.

Näide: Suuruse ξ korduva mõõtmise tulemused on $\{x_1, x_2, \dots, x_n\}$.

Üksikmõõtmise hälve $(x_i - \xi)$. Hälvete hulga mõõt $S = \sum_{i=1}^n (x_i - \xi)^k$

Antud \mathbf{x} , leida ξ nii, et $S = S_{\min}$. Lahendada juhul $k = 2$.

Põhimõte: hinnang leitakse nii, et hälvete ruutude summa \Rightarrow min.

Keskvärtuse, dispersiooni ja standardhälbe hindamine.

Keskvärtuse efektiivne hinnang m normaaljaotuse korral ja normaaljaotuse eriomadus. Hinnangu dispersioon.

Keskvärtuse efektiivne hinnang ühtlase jaotuse korral.

Dispersiooni ja standardhälbe hinnangud normaaljaotuse korral.

$\sigma(m)$ hinnang w/n ja n++ kompensatsioon.

4. Statistilised hüpoteesid

Hüpoteeside formuleerimine.

Sisuline hüpotees, formaliseerimine ja statistiline hüpotees.

- Näide:* * aiateibad (kasvavad/kahanevad/muutuvad) kuuvalguse mõjul,
 * aiateivaste pikkuste (keskväärtus/mediaan/jaotusseadus) muutub kuuvalgetel öödel (pikemaks/...../.....),
 * aiateivaste hommikuste mõõtmistulemuste hulk on pärit teisest üldkogumist kui õhtuste mõõtmistulemuste hulk, (kusjuures

Hüpoteeside tüübid.

- Näited:* a) normaaljaotusega suuruse $\mu = m$ ja $\sigma = s$,
 b) normaaljaotusega suuruse $\mu = m$,
 c) suuruse jaotusseadus on normaaljaotusseadus,
 d) kaks suurust alluvad samale jaotusseadusele
 e) normaaljaotusega suuruse $\mu > m$ ja $\sigma = s$.

Parameetrilised (a, b, e) ja mitteparameetrilised (c, d) hüpoteesid.

Lihtsad (a) ja ühendhüpoteesid (b, e).

Valimiruum ja testipiirkonnad.

Kriitiline piirkond $w - x \in w \Rightarrow$ hüpotees kummutatakse, vastuvõtmise piirkond $W-w$, hüpoteesi tunnustatakse.

Alternatiivhüpoteesid

Pideva jaotuse puhul on “täpse tabamuse” tõenäosus null, milline mõte on hüpoteesidel (a–e)?

Praktilisem ülesanne: teha valik kahe mudeli vahel.

Tüüpsituatsioon: mitteformaalsest probleemist otseselt tulenev primaarhüpotees väidab mingi efekti olemasolu ja osutub ühendhüpoteesiks. Alternatiiv väidab efekti puudumist ja see formuleeritakse võimalikult lihtsalt kontrollitavana. Sellisel juhul nimetatakse alternatiivi nullhüpoteesiks.

Näide:

Sisuline hüpotees: magnetorm mõjustab inimeste enesetunnet.

Eksperiment: n vaatlusalust ei informeerita magnetormidest ja nad hindavad igal õhtul enda subjektiivset enesetunnet 100 pallises skaalas.

Defineerime x_i = hinne magnetormi päeval – hinne eelmisel päeval.

Aprioorne eeldus: mõju puudumise korral oleks x negatiivse ja positiivse väärtuse tõenäosused võrdsed.

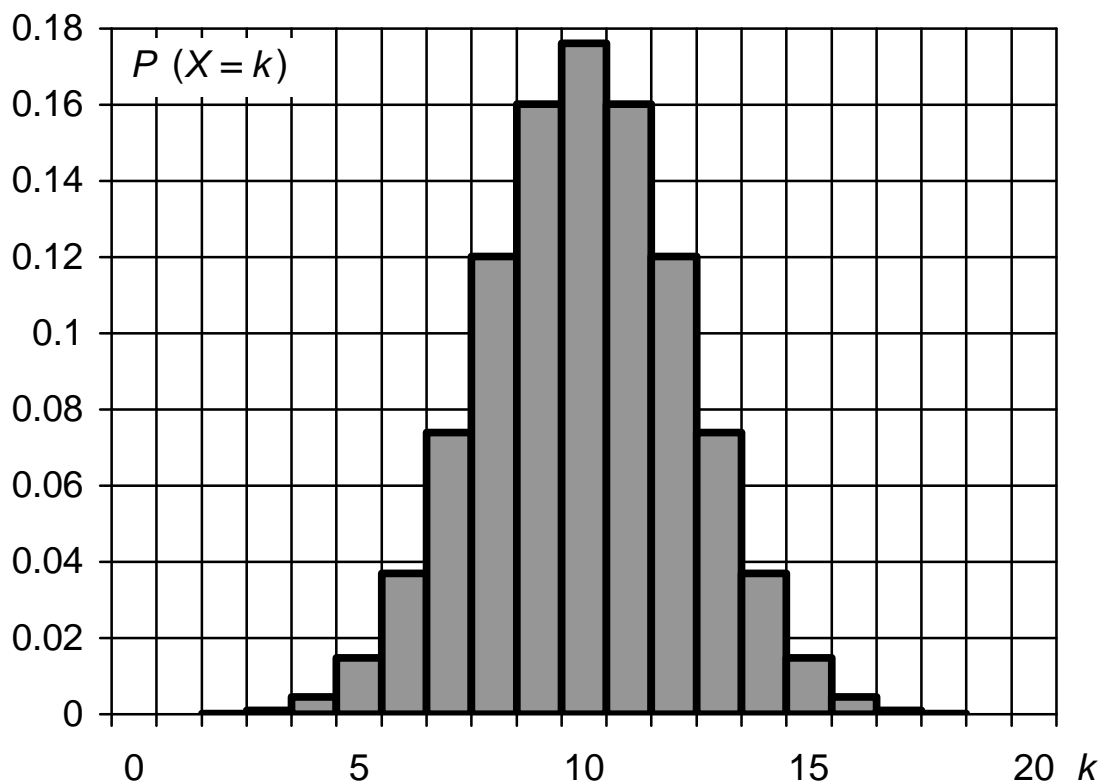
Statistiline primaarhüpotees H_1 : x mediaan erineb nullist.

Alternatiiv H_0 : x mediaan on null (nullhüpotees, mõju puudub).

H_0 on lihthüpotees.

n – valimi maht, X – positiivsete nihete arv valimis. Jaotusseadus:

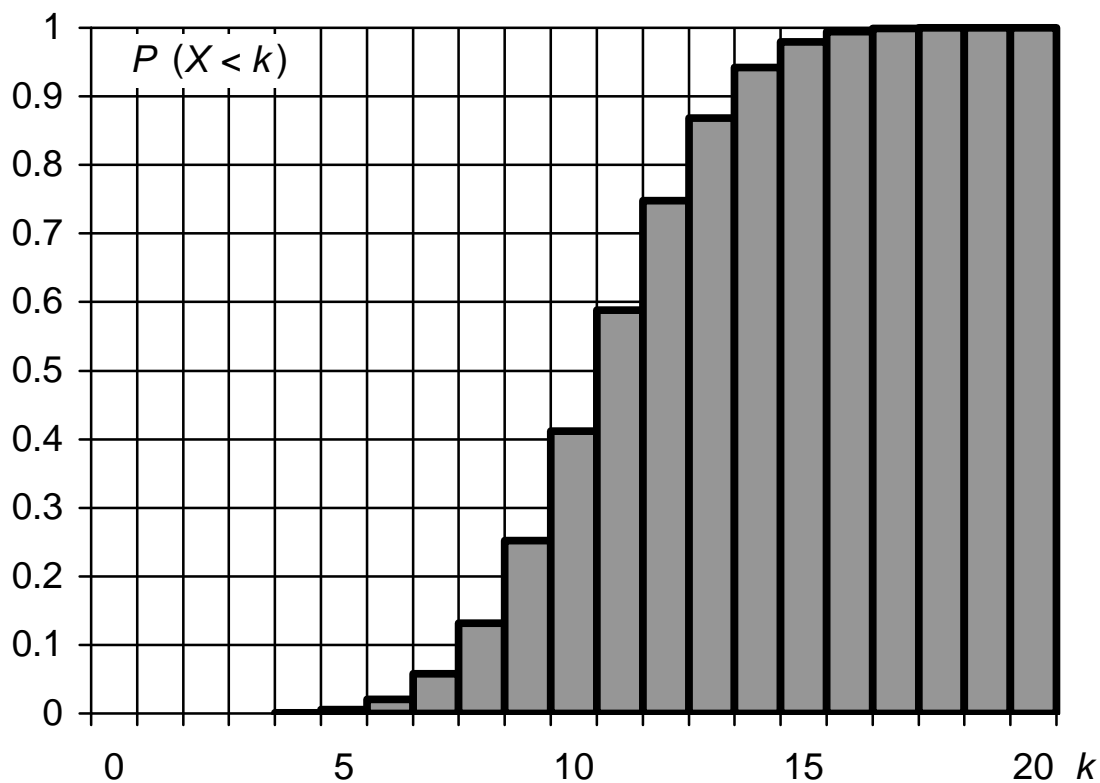
$$P(X = k) = \frac{n!}{k!(n-k)!2^n}$$



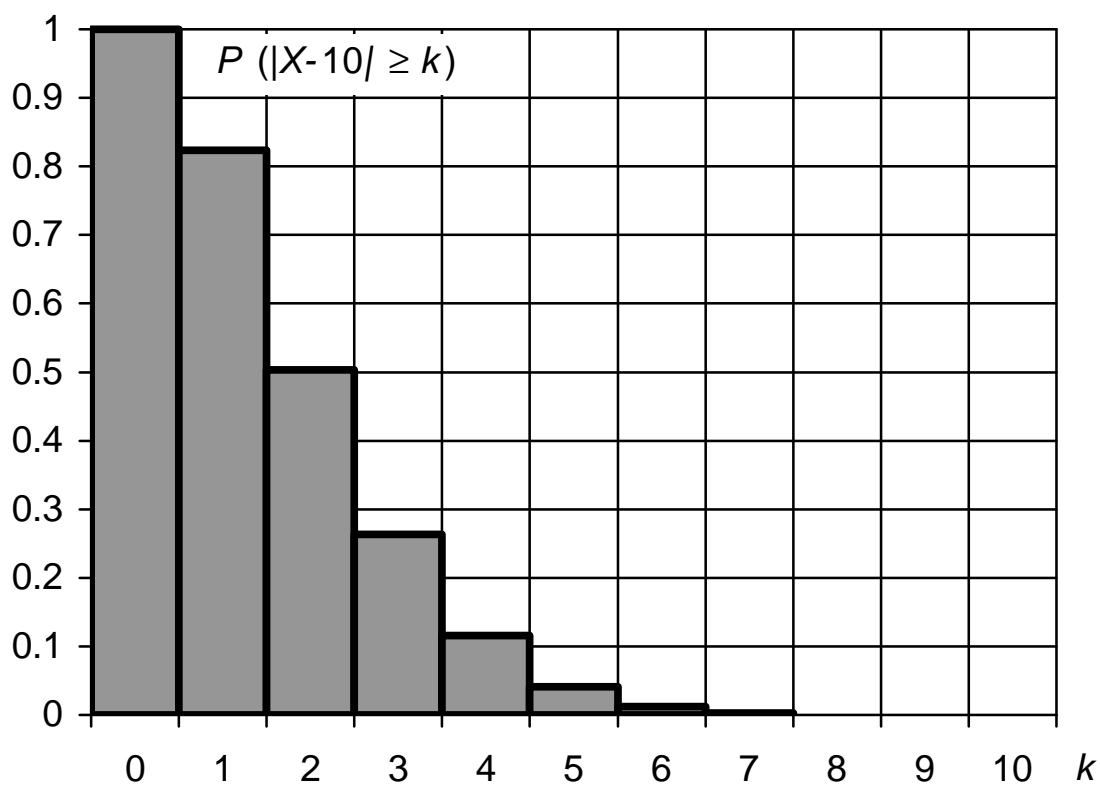
Binoomjaotuse tõenäosused erijuhul $n = 20$, $p = q = 0.5$

Binoomjaotus erijuhul $n = 20, p = q = 0.5$

Jaotusfunktsioon diskreetsel skaalal



Tõenäosus, et absoluutne tsentraalhälve pole väiksem ettantud arvust k



k	$P (X-10 \geq k)$
0	1.000000
1	0.823803
2	0.503445
3	0.263176
4	0.115318
5	0.041389
6	0.011818
7	0.002577
8	0.000402

Testipiirkonna konstrueerimine ühemõõtmelise valimifunktsiooni abil.

Kahepoolsed ja ühepoolsed testid.

Hüpoteeside kontrollimise standardmeetod ja võimalikud vead.

Hüpoteesile H_1 Püstitatakse selline alternatiiv H_0 , mille korral mingi statistiku t tõenäosusjaotus on teoreetilise mudeli abil määratav.

Määratakse statistiku väärtusete piirkond S , mille puhul alternatiiv H_0 kuulutatakse uskumatuks ja H_1 tunnustatavaks.

Tegelikult on õige	Testi tulemus	
	$t \notin S \Rightarrow H_0$	$t \in S \Rightarrow H_1$
H_0	Tehakse õige otsus	Esimest liiki viga
H_1	Teist liiki viga	Tehakse õige otsus

Olulisuse nivoo q = esimest liiki vea suurim lubatud tõenäosus.

Usaldusnivoo $p = 1 - q$.

Testi võimsus = teist liiki vea vältimise tõenäosus.

5. Vahemikhinnangud

Gaussi ülesanne.

Normaalselt jaotatud mõõtmisvigadega suuruse mõõtmistulemused on $x_1 \dots x_n$. Leida vahemik (a, b) , mis $p\%$ tõenäosusega sisaldaks X kesk-
väärtust μ . Lisaeldus: n on suur.

Lahendus: Arvutame μ ja σ_m punkthinnangud m ja $s_m = s/\sqrt{n}$.

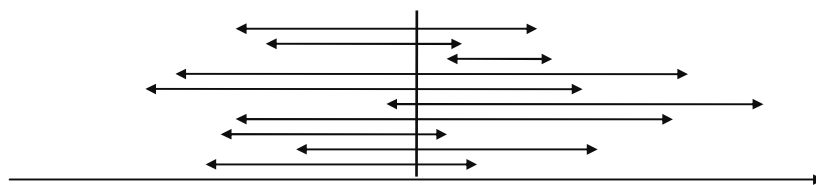
Tõenäosus, et m jääb vahemikku $\mu \pm k\sigma_m$ on $p_k = \Phi(k) - \Phi(-k)$.

$\Phi(k) = P\left(\frac{m - \mu}{\sigma_m} < k\right)$ – standardiseeritud normaaljaotusfunktsioon.

Valime k nii, et $p_k = p\%$. See on tõlgendatav kui tõenäosus, et m ja μ vahe ei ületa $k\sigma_m$ ehk $\mu = m \pm k\sigma_m$. Kui n on suur,

siis $\sigma_m \approx s_m$ ja $\mu \approx m \pm ks_m$.

p	90%	99%	99.9%	99.99%
k	1.64	2.58	3.29	3.89



Näide: 10 vahemikhinnangut ($p = 90\%$)

Studenti ülesanne.

Gaussi lahenduse puudulikkus lõpliku n puhul oli selge juba Gaussile endale. Lahenduse leidis aga alles 1908. a. Gosset.

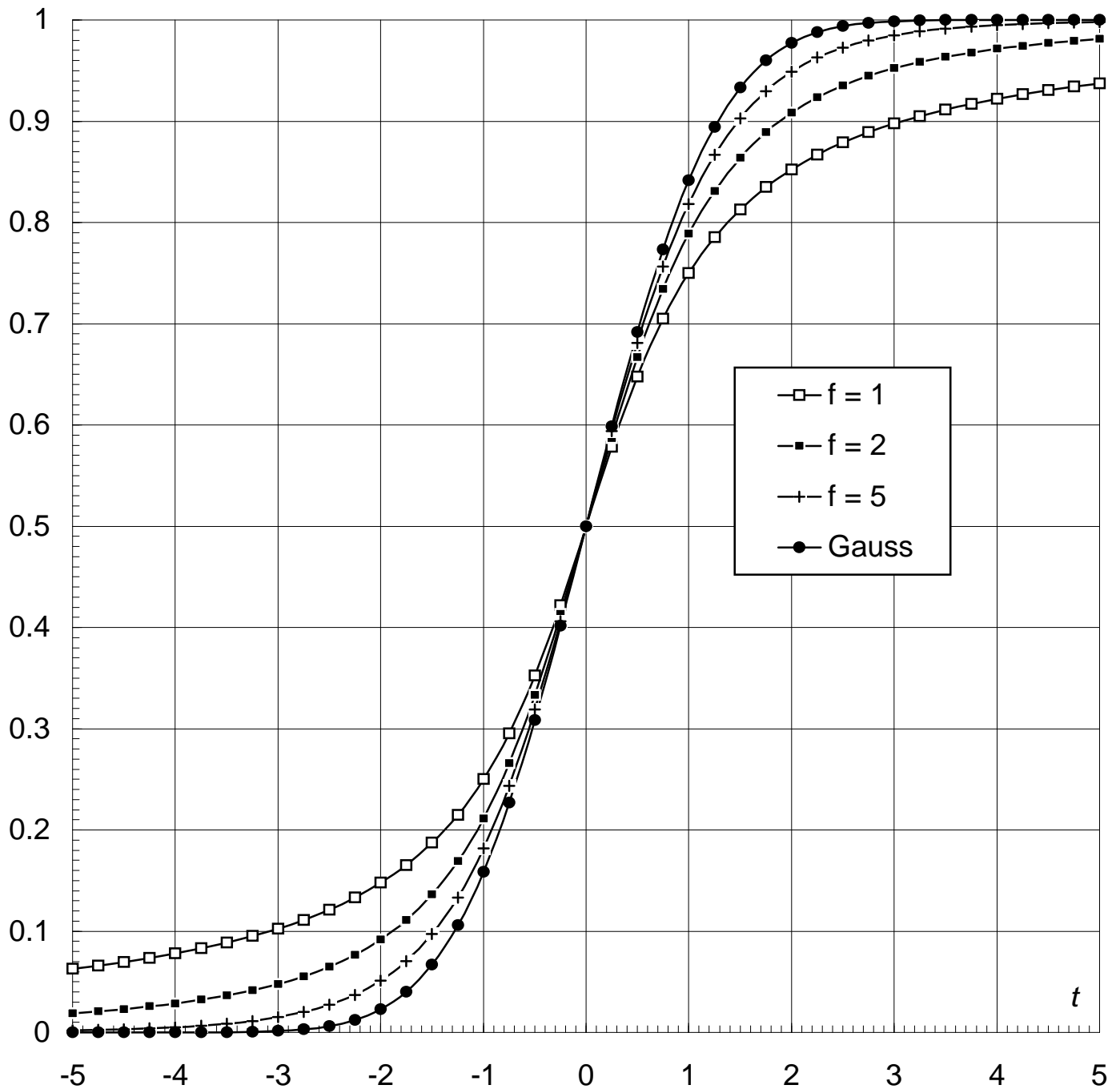
Idee: kuidas on jaotatud statistik $t_f = \frac{m - \mu}{s_m}$?

(n mõõtmist, $f = n - 1$ vabadusastet)

$$S_f(t) = P\left(\frac{m - \mu}{s_m} < t\right)$$

$$\mu = m \pm ts_m, \text{ kus } p = S_f(t) - S_f(-t)$$

Studenti jaotus



Vahemikhinnangute koostamise üldine meetod.

Statistiline hüpotees: hinnatav suurus on vahemikus (a, b) ,

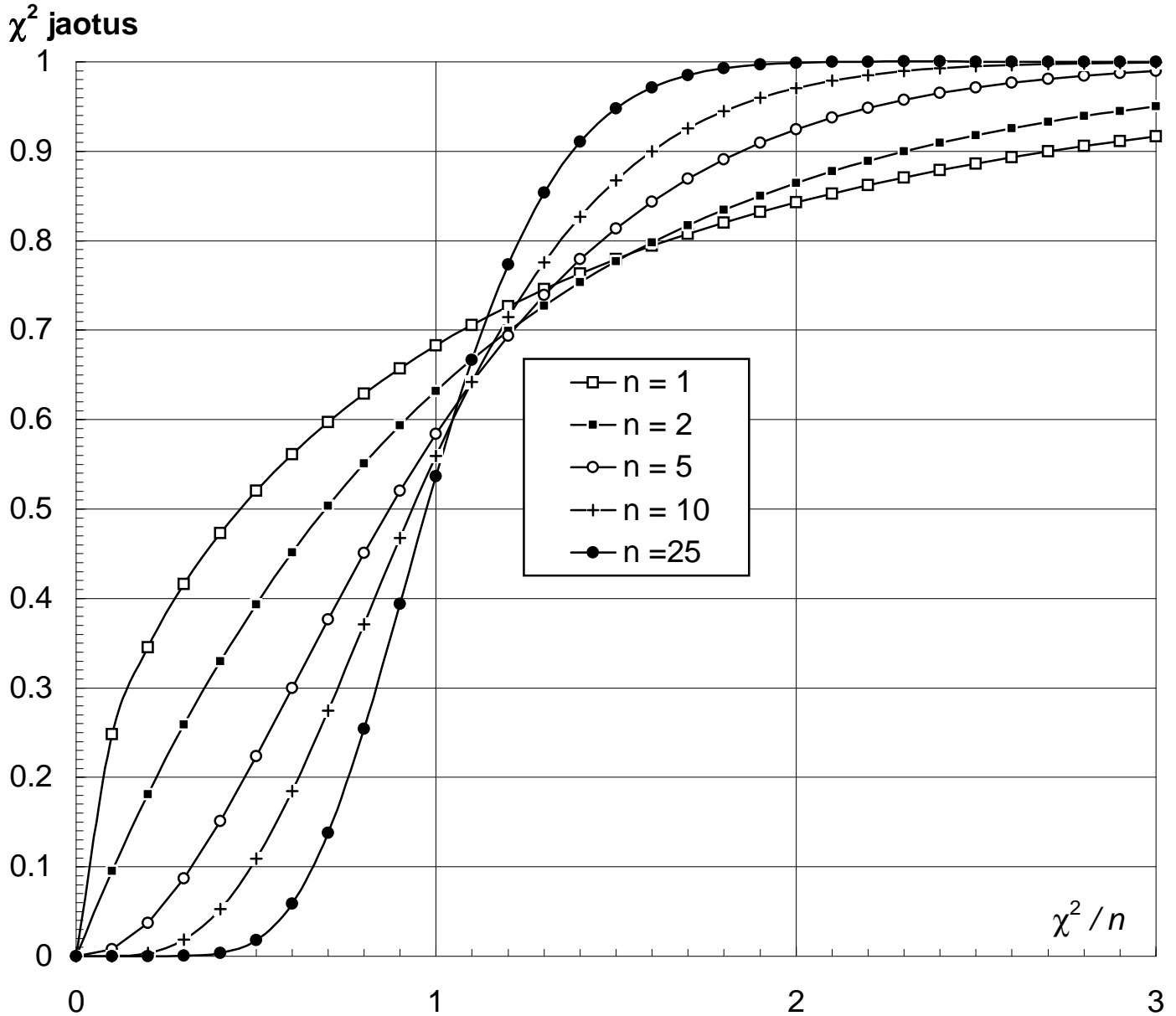
alternatiiv: hinnatav suurus on väljaspool vahemikku (a, b) .

Koostatakse hüpoteesi kontrollimise eeskiri ja leitakse minimaalne võimalik alternatiivi kummutamise olulisuse nivoo q . $p = 1 - q$.

Ülesanne: konstrueerida vahemik (a, b) nii, et etteantud p puhul $b - a \rightarrow \min$.

Hii-ruut jaotus ja dispersiooni vahemikhinnang.

$$X \sim N(0, 1), \chi_n^2 = \sum_{i=1}^n X_i^2, \text{ jaotusfunktsioon } F_n(x) = P(\chi_n^2 < x)$$



$$(n-1) \frac{s^2}{\sigma^2} \sim \chi_{n-1}^2$$

Mitteparameetrisised vahemikhinnangud.

Hinnangu $\lambda \in (x_{(1)}, x_{(n)})$ usaldustõenäosus.

Hinnangu $\lambda \in (x_{(k)}, x_{(n+1-k)})$ usaldustõenäosus.

```

function Fisher (m, n : integer; x : real) : real;
function Hii2 (m : integer; h2 : real) : real;
    Attention: a very rough algorithm!
function Student (n : integer; t : real) : real;
function Gauss (x : real) : real;
function Binom (m, n : integer; p : real) : real;

```

```
=====
```

```

function Student (n : integer; t : real) : real;
    {Probability of (Student statistic < t) using Fisher}
    var sign : real;
    begin
    if t < 0 then sign := -1 else sign := 1;
    Student := 0.5*(1+sign*Fisher (1, n, sqr (t)))
    end;

```

```
=====
```

```
Program StudTab;
```

```
{$I stat.pro}
```

```

const f : array [1..5] of integer = (1, 2, 3, 5, 30000);
var i, j : integer; t : real; s : text;

```

```

begin
assign (s, 'stud.tab'); rewrite (s);
for i := 0 to 40 do begin
    t := (i-20)/4;
    write (s, t:7:2);
    for j := 1 to 5 do write (s, student (f[j], t):7:4);
    writeln (s);
end;
close (s);
end.

```


6. Skalaarsete mõõtmiste eeltöötlus

Klassikaline lähenemisviis: keskmistamine vähendab juhuslikke mõõtmisvigu, järelikult tuleks mõõta nii palju kui jõuab ja niiviisi kogunev mõttetult suur andmestik hiljem keskmistamise teel kokku suruda.

Probleemid:

- müra normaaljaotus ei ole garanteeritud ja aritmeetiline keskmine võib osutuda halvaks hinnanguks,
- müra tegeliku jaotusseaduse kohta ei ole eelinformatsiooni,
- vaheandmete suur maht koormab arvutiressursse ja piirab sellega mõõtmisagedust.

Siit tulenevad soovitusel:

- kasutada ainult robustseid hinnanguid,
- teostada eeltöötlus samaaegselt mõõtmistega (“on-line” meetod).

Hinnangu robustsuse tagavad:

- võimalike jaotuste hulga kohta tehtud eeldustele tuginevad meetodid,
- klassikalised suurte hälvete elimineerimise meetodid
- mitteparameetrilised meetodid,
- totaalse tsensuuri meetodid.

Klassikalised on-line hinnangud

Ainult aritmeetilise keskmise arvutamisel pole tarvis vaheandmeid salvestada. Kui üksikmõõtmise tulemus on x , siis mõõtmiste käigus kogutakse Σx ja hiljem arvutatakse $\bar{x} = \Sigma x / n$.

Ülesanne: kuidas korraldada arvutused niiviisi, et koguda mitte summat vaid otse aritmeetilist keskmist?

Kui aga soovitakse ka standardhälbe hinnangut, siis peab arvutused korraldama erilisel viisil. Tõenäosusteooria käsiraamatutest võib leida kõigi tsentraalmomentide avaldised algmomentide kaudu. Iga algmomenti saab aga arvutada samuti kui aritmeetilist keskmist andmeid säilitamata.

Ülesanne: kuidas leida standardhälbe hinnang kolme summa $n = \Sigma 1$, Σx ja Σxx järgi?

Lineaarne regressioonimudel $y = \alpha x + \beta$

Tähistused: $x = \{x_i\}$, $y = \{y_i\}$, $i = 1..n$

$$\Sigma z = \sum_{i=1}^n z_i$$

Parameetrid:

Hinnangud:

Keskväärtused

$$\bar{x} \quad \bar{y}$$

Standardhälbed

$$\hat{s}_x \quad \hat{s}_y \quad s_x \quad s_y$$

Korrelatsioonikordaja

$$r$$

Regressiooniparameetrid

$$a \quad b$$

Mõõtmise ajal kogutakse 6 summat väärtustest:

1, x , y , xx , yy , xy

Pärast mõõtmisi arvutatakse:

$$n = \Sigma 1$$

$$\bar{x} = \Sigma x / n$$

$$\bar{y} = \Sigma y / n$$

$$\hat{s}_x^2 = \Sigma xx / n - \bar{x}^2$$

$$s_x^2 = n\hat{s}_x^2 / (n - 1)$$

$$\hat{s}_y^2 = \Sigma yy / n - \bar{y}^2$$

$$s_y^2 = n\hat{s}_y^2 / (n - 1)$$

$$r = (\Sigma xy / n - \bar{x}\bar{y}) / (\hat{s}_x \hat{s}_y)$$

$$a = rs_y / s_x$$

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

Tsensuuri tehnoloogia

Massiivi minimaalse ja maksimaalse elemendi leidmise algoritm.

Üheastmeline ekstremaalväärtuste elimineerimine.

Hierarhiline ekstremaalväärtuste elimineerimine.

Monte Carlo meetod

```

Program KESKMINE,
  {Monte Carlo meetodi demonstratsioon}

Const n = 25; {yksikmootmiste arv}
      m = 5;  {      sqrt (n)      }

Type andmed = array [1..n] of real;

Function mootmine (karbsed : boolean) : real;
  {Karbeste puudumisel Gaussi jaotusega m = 0, s = 1 suurus,
   karbeste esinemisel lisandub 1% toenaosusega lisaviga,
   mis on yhtlaselt jaotatud vahemikus [-100..+100]}
  Var x : real; i : integer;
  Begin
    x := 0;
    for i := 1 to 12 do x := x + random;
    x := x - 6; {Gauss m = 0, s = 1}
    if karbsed and (random < 0.01) then x := x + 200 * random - 100;
    mootmine := x;
  End;

Procedure mullsort (var x : andmed);
  Var OK : boolean;
      i : integer;
      a : real;
  Begin
    repeat OK := true;
      for i := 2 to n do if x [i-1] > x [i] then begin
        a := x [i-1]; x [i-1] := x [i]; x [i] := a;
        OK := false;
      end;
    until OK;
  End;

Procedure statistika (x : andmed;
  var
    km {mediaan},
    k0 {totaalkeskmine},
    k1 {valistatud ekstremaalvaartustega keskmine},
    k2 {kvartiilidevaheline keskmine},
    k5 {5-ste rühmade tsenseeritud keskmiste
        tsenseeritud keskmine}
  : real);

  Var      i, j, k : integer;
          a, b, s, min, max,
          vs, vmin, vmax : real;
  Begin

    {On-line rezhiimis teostatavad arvutused}

    {Totaalkeskmine}
    s := 0;
    for i := 1 to n do s := s + x [i];
    k0 := s / n;

    {Ekstreemumtsenseeritud keskmine}
    s := x [1]; min := s; max := s;
    for i := 2 to n do begin;

```

```

    s := s + x [i];
    if x [i] < min then min := x [i];
    if x [i] > max then max := x [i];
end;
k1 := (s - min - max) / (n - 2);

{5*5 tsenseeritud keskmine}
vs := 0; vmin := 1E33; vmax := -vmin;
i := 1;
for j := 1 to m do begin
    s := x [i]; i := i + 1; min := s; max := s;
    for k := 2 to m do begin;
        a := x [i];
        s := s + a;
        if a < min then min := a;
        if a > max then max := a;
        i := i + 1;
    end;
    b := (s - min - max) / (m - 2); {m mootmise tsenseeritud keskmine}
    vs := vs + b;
    if b < vmin then vmin := b;
    if b > vmax then vmax := b;
end;
k5 := (vs - vmin - vmax) / (m - 2);
{ylesanne 1 : kirjutada viimane algoritm ratsionaalsemalt
             kogudes kohe (m - 2)^2 arvu summa}
{ylesanne 2 : kirjutada algoritm m^k mootmise k-astmelise
             tsenseeritud keskmise arvutamiseks}

{On-line rezhiimis mitteteostatavad arvutused}
mullsort (x);

{mediaan}
km := x [13]; {mediaan NB! arvulised indeksid
               demonratsiooni lihtsuse huvides}

{kvartiilidevaheline keskmine}
s := 0;
for i := 7 to 19 do s := s + x [i];
k2 := s / 13;

End;
```

```

Procedure MonteCarlo (kordi : integer; karbsed : boolean; var f : text);
Var i, j : integer;
    km, k0, k1, k2, k5 : real;
    sm, s0, s1, s2, s5 : real;
    x : andmed;
Begin
randomize;
sm := 0; s0 := 0; s1 := 0; s2 := 0; s5 := 0;
for i := 1 to kordi do begin
    for j := 1 to 25 do x [j] := mootmine (karbsed);
    statistika (x, km, k0, k1, k2, k5);
    sm := sm + km*km;
    s0 := s0 + k0*k0;
    s1 := s1 + k1*k1;
    s2 := s2 + k2*k2;
    s5 := s5 + k5*k5;
end;
sm := sm / kordi;
s0 := s0 / kordi;
```

```

s1 := s1 / kordi;
s2 := s2 / kordi;
s5 := s5 / kordi;
writeln (f, sm:18:4, s0:9:4, s1:9:4, s2:9:4, s5:9:4);
End;

```

```

VAR i : integer;
     f : text;

```

```

BEGIN
assign (f, '\b\keskmine.tab'); rewrite (f);
writeln (f, 'Karbesteta mediaan totaal -minmax 7...19 5*5');
for i := 1 to 5 do MonteCarlo (10000, false, f);
writeln (f);
writeln (f, 'Karbestega mediaan totaal -minmax 7...19 5*5');
for i := 1 to 5 do MonteCarlo (10000, true, f);
close (f);
END.

```

{Arvutusaeg reaalarvuprotsessorita 25 Mhz masinal 22 minutit}

Karbesteta	mediaan	totaal	-minmax	7...19	5*5
	0.0645	0.0402	0.0414	0.0488	0.0532
	0.0630	0.0397	0.0407	0.0479	0.0529
	0.0647	0.0404	0.0416	0.0485	0.0538
	0.0625	0.0400	0.0410	0.0478	0.0528
	0.0634	0.0406	0.0417	0.0488	0.0531
Karbestega	mediaan	totaal	-minmax	7...19	5*5
	0.0637	1.3355	0.0924	0.0490	0.0534
	0.0646	1.3791	0.0868	0.0488	0.0525
	0.0638	1.3041	0.0926	0.0487	0.0538
	0.0646	1.3950	0.0955	0.0493	0.0541
	0.0643	1.3749	0.1020	0.0494	0.0536

7. Lineaarne spektraalaparaat

Spekter, kirje ja müra

Spekter on jaotusfunktsioon. Näide: kiirgusenergia jaotus sageduse järgi

$$f(\nu) = \frac{dE}{d\nu}$$

Matemaatiliselt saab kõiki jaotusi kirjeldada samalaadsed. Nimetust “spekter” kasutatakse ka jaotuste puhul, mis füüsiliselt ei sarnane valguse spektriga. Näide: aerosooliosakeste mõõtmespekter dn/dr .

Spektromeetri (üldisemas kontekstis aparadi) kirje on spektri mõõtmisel saadud arvutabel $y = \{y_i\}$. Ideaalne spektromeeter peaks mõõtma otse spektrit ja kirje peaks võrduma spektriga. Miks see aga nii olla ei saa?

Müra $\{\xi_i\}$ on kirje mõõtmisviga. Mudel $y = y_0 + \xi$.

Lõpmatumõõtmeline lineaarne mudel

Tähistused: spektri argument ν , spektraalfunktsioon x , kirje (ja müra) argument w , kirje y , müra ξ .

Üks isoleeritud kirje punkt on spektraalfunktsiooni väärtuste lineaarkombinatsioon, millele liitub müra:

$$y_0 = \int_{\nu_{\min}}^{\nu_{\max}} g_0(\nu)x(\nu) d\nu + \xi_0$$

Milline on funktsiooni g füüsikaline sisu?

Kokkulepe $\langle \xi_0 \rangle = 0$ – homogeenne aparaat.

Üldine võrrand:

$$y(w) = \int_{\nu_{\min}}^{\nu_{\max}} g(w, \nu)x(\nu) d\nu + \xi(w)$$

Aparaadifunktsiooni interpretatsioon.

Analoogia lineaarse filtriga (signaaliteooria).

Sidumaparaat

Ülesanded lihtsustuvad oluliselt juhul kui $g(w, v) = g(w-v)$:

$$y(w) = \int_{-\infty}^{\infty} g(w-v)x(v)dv + \xi(w) = \int_{-\infty}^{\infty} g(u)x(w-u)du + \xi(w)$$

Kui $\xi = 0$, siis on ülaltoodud teisendus konvolutsioon ehk sidum (vene k. “svertka”).

Tähistame Fourier pärioperaatori F ja pöördoperaatori $F' = F^{-1}$:

$$S(\omega) = F\{s(x)\} = c_1 \int_{-\infty}^{\infty} s(x)e^{-i\omega x} dx$$

$$s(x) = F^{-1}\{S(\omega)\} = c_2 \int_{-\infty}^{\infty} S(\omega)e^{i\omega x} d\omega$$

$$c_1 c_2 = \frac{1}{2\pi}, \text{ tavaliselt } c_1 = \frac{1}{2\pi} \text{ ja } c_2 = 1$$

Vahel kasutatakse argumendina ω asemel $f = \omega/(2\pi)$.

Küsimus: millised on ω ja Fourier kujutise mõõtühikud?

Sidumi omadus:

$$F\{y\} = F\{g\} F\{x\}$$

PS: matemaatikahariduseta raadioamatööri arutlus lineaarsest filtrist.

Koopia raamatust: valik Fourier kujutisi.

Lõplikumõõtmeline mudel

Lõpmatumõõtmelise mudeli tähistused: spektri argument ν , spektraalfunktsioon x , kirje (ja müra) argument w , kirje y , müra ξ .

Lõplikumõõtmelise mudeli tähistused: spektri argument $i = 1..n$, spektraalfunktsioon $\mathbf{x} = \{x_i\}$, kirje (ja müra) argument $j = 1..m$, kirje $\mathbf{y} = \{y_j\}$, müra $\xi = \{\xi_j\}$.

Spektri elemendid – fraktsiooniintensiivsused,

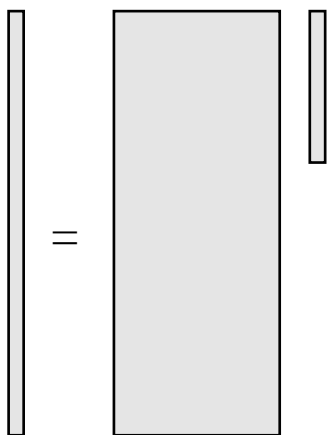
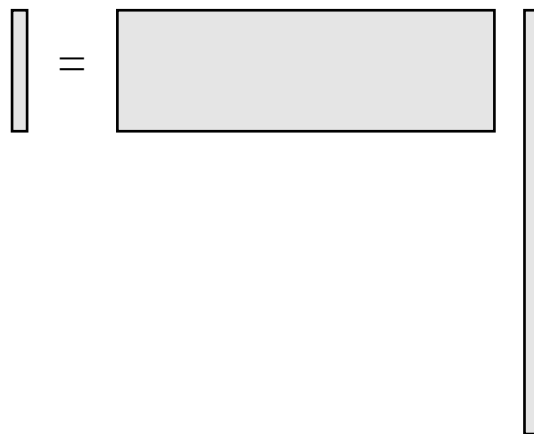
Kirje elemendid – kanalisignaalid.

Üks kanalisignaal $y_1 = g_{11}x_1 + g_{12}x_2 + g_{13}x_3 + \dots + \xi_1$

Aparaativõrrand

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \xi \quad \text{ehk} \quad y_j = \sum_{i=1}^n G_{ji}x_i + \xi_j$$

\mathbf{G} – aparaadimaatriks.



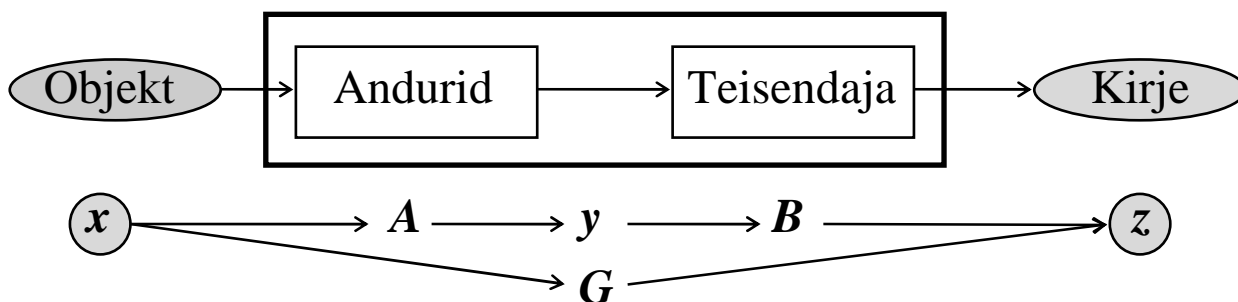
Näited erinevate n ja m kombinatsioonidega.

Koopiad A. Mirme dissertatsioonist:

- 1) aparaadimaatriksi näide,
- 2) kolmnurkbaasi kasutamine.

Mudeli universaalsus

Andmete teekond algab alati loodusest, mille olekut kirjeldab spekter x ja lõpeb mõõtmisprotokollis või artiklis, kus tulemusi kirjeldab vektor z . Teisenemine võib olla keerukas ja vahetult modelleeritav ainult mitmetapilises skeemis. Esimene näide: liitaparaadi mudel:



$$y = Ax \quad z = By \quad z = B(Ax) = (BA)x$$

$$\text{Järeldus: } z = Gx, \text{ kus } G = BA$$

Teine näide: tikutoosi mõõtmine.

Kolmas näide: kahe paberitüki kaalumine.

Mudeli kitsendused: lineaarsus, homogeensus.

Lõpmatumõõtmelise mudeli algebraiseerimine

Delta-summa meetod. Ristkülikfraktsioonide meetod.

Algebraiseerimine numbrilise integreerimise valemite abil.

Universaalne mudel: algebraiseerimine koordinaatesituse abil.

Baas $\{f^i(v)\}, i = 1..n.$ $x(v) = \sum_{i=1}^n x_i f^i(v)$

Erijuhtumid: δ -baas, tulp-baas, kolmnurkbaas, spline-baas, ...,
Fourier baas?

Spektromeetri päriülesanne ja pöördülesanne

Matemaatiline päriülesanne ja füüsikaline päriülesanne.

Valik eriomadustega aparate

- Ühikaparaat
- Diagonaalaparaat
- Toeplitzi aparaat
- Tsirkulatsioonaparaat
- Toeplitzi aparadi tsirkulantlaiend
- Ortogonaalaparaat
- Fourier aparaat
- Multipleksaparaadid ja Fellgeti (1951, 1958) võit
näide: kahe keha kaalumine.

8. Lineaarse spektraalaparaadi pöördülesanne.

Lõpmatuumõõtmeline ülesanne

Kuna tegelik lahendamine on paratamatult lõplikumõõtmeline, siis ainult probleemi lihtsustatud kirjeldus, mille juures mõõtmisvigu ei arvestata.

Aparaadivõrrand

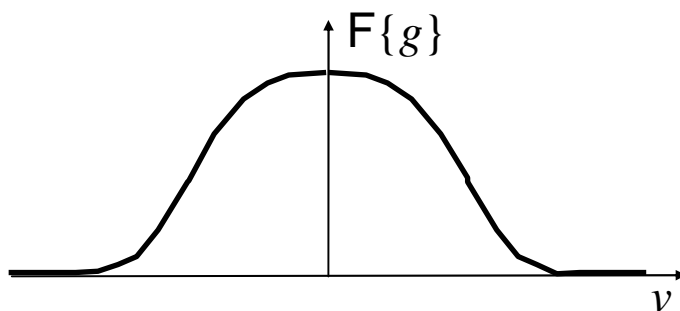
$$y(w) = \int_{v_{\min}}^{v_{\max}} g(w, v)x(v)dv$$

on formaalselt Fredholmi esimest liiki võrrand ja selle lahendamine Hadamard'i mõttes matemaatiliselt ebakorrektselt püstitatud ülesanne. 60-ndatel aastatel loodi regulariseerimise teooria, mis mõõtmisandmete töötlemise ülesannete puhul ei ole rahuldav.

Probleemi olemus on hästi läbi nähtav sidumaparaadi puhul

$$y(w) = \int_{-\infty}^{\infty} g(w-v)x(v)dv \quad F\{y\} = F\{g\} F\{x\}$$

Formaalne lahend on $F\{x\} = F\{y\} / F\{g\}$. Tüüpolukord:



Analoogia helisalvestise restaureerimise ülesandega.

Ruutvormid

$$R(u) = \sum_{i,j} A_{ij}u_iu_j = \sum_i u_i \left(\sum_j A_{ij}u_j \right) = (u, Au) = u^T Au$$

Sümmetriseerimine $A := (A + A^T) / 2$

Mittenegatiivne ja positiivne määratus ning A^{-1} olemasolu tingimus.

Mitmemõõtmeline normaaljaotus

Juhuslik vektor $\mathbf{v} = \{v_i\}$, $\langle \mathbf{v} \rangle = \{\langle v_i \rangle\}$, $\mathbf{u} = \mathbf{v} - \langle \mathbf{v} \rangle$

Dispersiooni- ehk kovariatsioonimaatriks $\mathbf{C} = \{c_{ij}\} = \{\langle u_i u_j \rangle\}$

Normaalne tihedusfunktsioon

$$f(\mathbf{u}) = \text{const} \times \exp\left(-\frac{1}{2}R(\mathbf{u})\right)$$

$$R(\mathbf{u}) = \mathbf{u}^T \mathbf{A} \mathbf{u} = (\mathbf{v} - \langle \mathbf{v} \rangle)^T \mathbf{A} (\mathbf{v} - \langle \mathbf{v} \rangle), \quad \text{const} = \frac{\sqrt{\det \mathbf{A}}}{(2\pi)^{n/2}}$$

$$f(\mathbf{v}) = \frac{\sqrt{\det \mathbf{A}}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(\mathbf{v} - \langle \mathbf{v} \rangle)^T \mathbf{A} (\mathbf{v} - \langle \mathbf{v} \rangle)\right)$$

$$\mathbf{C} = \mathbf{A}^{-1}$$

Kontrolliks: ühemõõtmeline erijuhtum.

Vektormõõtmise statistiline mudel

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \boldsymbol{\xi} \quad \text{ehk} \quad y_j = \sum_{i=1}^n G_{ji} x_i + \xi_j$$

Hüpotees: $\boldsymbol{\xi}$ tõenäosusjaotus on mitmemõõtmeline normaaljaotus.

$$\boldsymbol{\xi} = \mathbf{G}\mathbf{x} - \mathbf{y}$$

Eeldus: \mathbf{G} , \mathbf{y} ja $\mathbf{D} = \{\langle \xi_i \xi_j \rangle\}$ on tuntud suurused

Ülesanne: leida tundmatu parameetri \mathbf{x} jaoks parim hinnang $\hat{\mathbf{x}}$.

Gauss-Markovi algoritm

\mathbf{x} on tundmatu, kuid kindel suurus, $\hat{\mathbf{x}}$ on esialgu vabalt varieeritav.

Gauss vähimruutude meetodi lihtsustatud variandi puhul on ülesandeks valida $\hat{\mathbf{x}}$ niiviisi, et hälvete ruutude summa $\sum(\mathbf{G}\mathbf{x} - \mathbf{y})^2$ oleks vähim võimalik. Täiustatud variandi korral korrutatakse summeeritavad kaaludega, mis on pöördvõrdelised mõõtmisvea ruuduga. Selle ülesande lahendas Gauss ise. Veel enam täiustatud ülesanne võtab arvesse ka mõõtmisvea koordinaatide vahelisi korrelatsioone. Selle ülesande lahendas Markov.

Veel üldisem ja otsesemalt põhjendatud on suurima tõepära ülesanne. Teatavasti taandub see normaaljaotuse korral Gauss-Markovi vähimruutude ülesandeks.

Tõepärafunktsiooni võrrand on vormilt sama, kui tõenäosusjaotuse tihedusfunktsioon, siin aga on varieeritavaks parameetri hinnangu testväärus $\hat{\mathbf{x}}$:

$$L(\hat{\mathbf{x}}) = c_n \times \exp\left(-\frac{1}{2}\boldsymbol{\xi}^T \mathbf{D}^{-1}\boldsymbol{\xi}\right) = c_n \times \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{G}\hat{\mathbf{x}})^T \mathbf{D}^{-1}(\mathbf{y} - \mathbf{G}\hat{\mathbf{x}})\right)$$

Tõepära maksimiseerimiseks tuleb minimiseerida eksponendi argumendi moodul

$$R = (\mathbf{y} - \mathbf{G}\hat{\mathbf{x}})^T \mathbf{D}^{-1}(\mathbf{y} - \mathbf{G}\hat{\mathbf{x}}).$$

Ekstreemumülesande lahendamise eeskiri: lahendada võrrandisüsteem

$$\frac{\partial R}{\partial \hat{x}_k} = 0, \quad k = 1..n.$$

Lahenduskäik:

$$R = (\mathbf{y} - \mathbf{G}\hat{\mathbf{x}})^T \mathbf{D}^{-1} (\mathbf{y} - \mathbf{G}\hat{\mathbf{x}})$$

$$R = \mathbf{y}^T \mathbf{D}^{-1} \mathbf{y} - \mathbf{y}^T \mathbf{D}^{-1} \mathbf{G}\hat{\mathbf{x}} - (\mathbf{G}\hat{\mathbf{x}})^T \mathbf{D}^{-1} \mathbf{y} + (\mathbf{G}\hat{\mathbf{x}})^T \mathbf{D}^{-1} \mathbf{G}\hat{\mathbf{x}}$$

$$R = \text{const} + \hat{\mathbf{x}}^T (\mathbf{G}^T \mathbf{D}^{-1} \mathbf{G}) \hat{\mathbf{x}} - 2(\mathbf{y}^T \mathbf{D}^{-1} \mathbf{G}) \hat{\mathbf{x}}$$

Tähistame: $\mathbf{A} = \mathbf{G}^T \mathbf{D}^{-1} \mathbf{G}$ ja $\mathbf{b} = \mathbf{G}^T \mathbf{D}^{-1} \mathbf{y}$

Saame minimiseeritava: $S = R - \text{const} = \hat{\mathbf{x}}^T \mathbf{A} \hat{\mathbf{x}} - 2\mathbf{b}^T \hat{\mathbf{x}}$

Arvutame:

$$\frac{\partial S}{\partial \hat{x}_k} = \frac{\partial}{\partial \hat{x}_k} \left(\sum_{i \neq k} A_{ik} \hat{x}_k \hat{x}_i + \sum_{j \neq k} A_{kj} \hat{x}_j \hat{x}_k + A_{kk} \hat{x}_k^2 - 2b_k \hat{x}_k \right) = 2 \sum_j A_{kj} \hat{x}_j - 2b_k = 0$$

Minimiseerimisülesanne sai kuju $\mathbf{A}\hat{\mathbf{x}} - \mathbf{b} = 0$

Markovi lahend: $\hat{\mathbf{x}} = \mathbf{A}^{-1} \mathbf{b} = (\mathbf{G}^T \mathbf{D}^{-1} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{D}^{-1} \mathbf{y}$

Algoritm:

$$\mathbf{H} := \mathbf{G}^T \mathbf{D}^{-1}$$

$$\mathbf{C} := (\mathbf{H}\mathbf{G})^{-1}$$

$$\mathbf{x} := \mathbf{C}\mathbf{H}\mathbf{y}$$

Mõõtmisvigade teisenemine.

Juhuslikku viga sisaldava \mathbf{x} tõenäosustiheduse ja selle mittejuhusliku testväärtuse tõepärafunktsiooni avaldis on ühine. Normaalselt jaotatud $f(\mathbf{x})$ avaldist võib kirjutada kahel võrdväärset viisil

$$cn_1 \times \exp\left(-\frac{1}{2}(\mathbf{x} - \langle \mathbf{x} \rangle)^T \mathbf{C}^{-1} (\mathbf{x} - \langle \mathbf{x} \rangle)\right) = cn_2 \times \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x}\right)$$

Ruutliikmete kordajad vasakus ja paremas astmenäitajas peavad olema võrdsed. Kodune (ja eksami-) ülesanne: tõestada, et $\mathbf{C} = \mathbf{A}^{-1}$.

Järeldus: hinnangu $\hat{\mathbf{x}}$ mõõtmisvigu kirjeldab $\mathbf{C} = (\mathbf{G}^T \mathbf{D}^{-1} \mathbf{G})^{-1}$, see on sama \mathbf{C} , mis arvutati vahetulemusena välja ülaltoodud algoritmis.

Ülesanne: Kirjutada ülaltoodud algoritm välja ühemõõtmelise erijuhu jaoks!

Pöördülesanne tsirkulatsioonaparaadi puhul

Fourier teisendusel on lõplikumõõtmeline analoog. n -mõõtmelin vektor \mathbf{x} ja tema Fourier kujutis $\tilde{\mathbf{x}}$ on seotud järgnevalt:

$$\tilde{\mathbf{x}} = \mathbf{F}\mathbf{x}$$

$$\mathbf{x} = \frac{1}{n}\mathbf{F}^* \tilde{\mathbf{x}},$$

teisendusmaatriksi \mathbf{F} elemendid on:

$$F_{kl} = \exp\left(-\frac{2\pi i}{n}kl\right).$$

Praktilisel arvutamisel kasutatakse FFT algoritmi.

Olgu \mathbf{x} ja \mathbf{y} vektorid, \mathbf{G} tsirkulatsioonmaatriks ja \mathbf{g} selle esimene veerg, ning

$$\mathbf{y} = \mathbf{G}\mathbf{x}$$

Siis on rangelt algebraliselt

$$\tilde{y}_i = \tilde{g}_i \tilde{x}_i.$$

Kui \mathbf{g} Fourier kujutise elementide hulgas pole nulle, taandub võrrandi lahendamine sageduste ruumis n jagamistehtele.

Selleks, et pöördülesande lahendamisel saaks lõplikumõõtmelist Fourier teisendust rangelt kasutada, peavad \mathbf{G} ja \mathbf{D} olema mõlemad tsirkulatsioonmaatriksid. Ainult sel juhul võib rääkida tsirkulatsioonaparaadist.

Küsimus: millised füüsikalised-tehnilised kitsendused tulenevad sellest nõudest?

Ülesanne: kirjutage välja varasema üldise algoritmi analoog tsirkulatsioonaparaadi jaoks.

LS84-85 Tsirkulatsioonmaatriksid

LS86-87 Tsirkulatsioonmaatriksid

9. Mõõtmise stabiliseerimine.

Vektormõõtmise pöördülesande ebakorrektsus või ebastabiilsus.

Nagu eespool öeldud, lõpmatumõõtmeline aparaativõrrand

$$y(w) = \int_{v_{\min}}^{v_{\max}} g(w, v)x(v)dv$$

on matemaatiliselt ebakorrektselt püstitatud.

Algebraalse aparaativõrrandi

$$y = Gx$$

puhul ei saa ebakorrektsusest Hadamard'i mõttes rääkida, algebras sellist mõistet pole. Küll aga võib võrrandi lahend osutada ebastabiilseks.

Näide 1:

$$x_1 + x_2 = y_1$$

$$x_1 + 1.001x_2 = y_2$$

$$x_2 = 1000(y_2 - y_1)$$

$$\sigma(x_2) = 1414 \sigma(y)$$

Näide 2:

$$y(w) = \int_{v_{\min}}^{v_{\max}} g(w, v)x(v)dv + \xi(w)$$

Konkreetsed tingimused: $g(w, v) = e^{-(w-v)^2}$, valge müra, $\sigma = 1$.

Algebraiseerime: $v = ci, w = cj, dv = c$

$$y_j = \sum g_{ji}x_i \quad g_{ji} = ce^{-c^2(i-j)^2} \quad D = E$$

$$C = (G^T G)^{-1} \quad \text{Veategur} = \langle \sigma_y \rangle = \sqrt{\frac{\text{Spur}(C)}{n}}$$

Mis juhtub, kui hakata algebraiseeritud võrrandit täpsustama vähendades sammu c ?

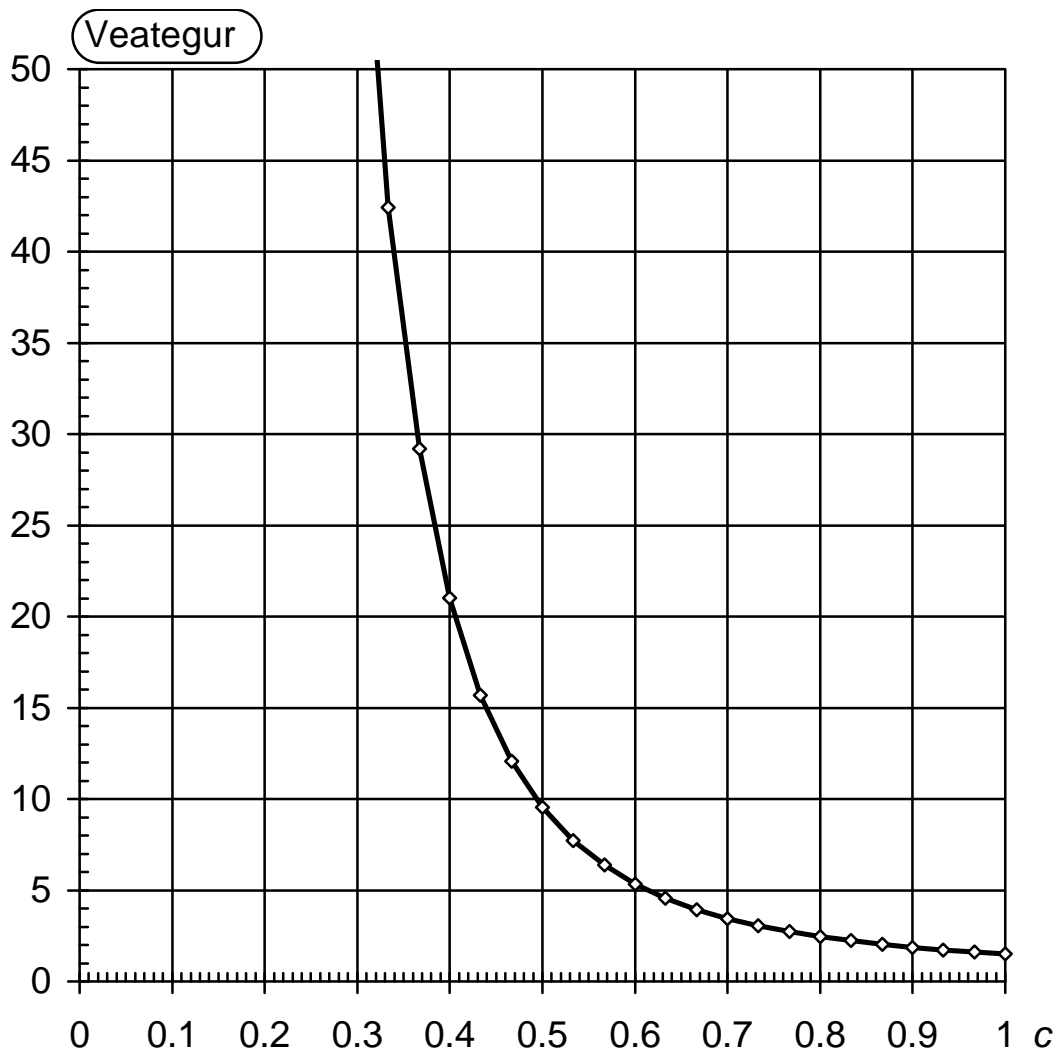
```

Program Demo;
Const nmatrix = 5;
{$I MATRIX.PRO}

Var i, j, n, k : integer;
    x, y, c : double;
    a, g : matrix;
    f : text;

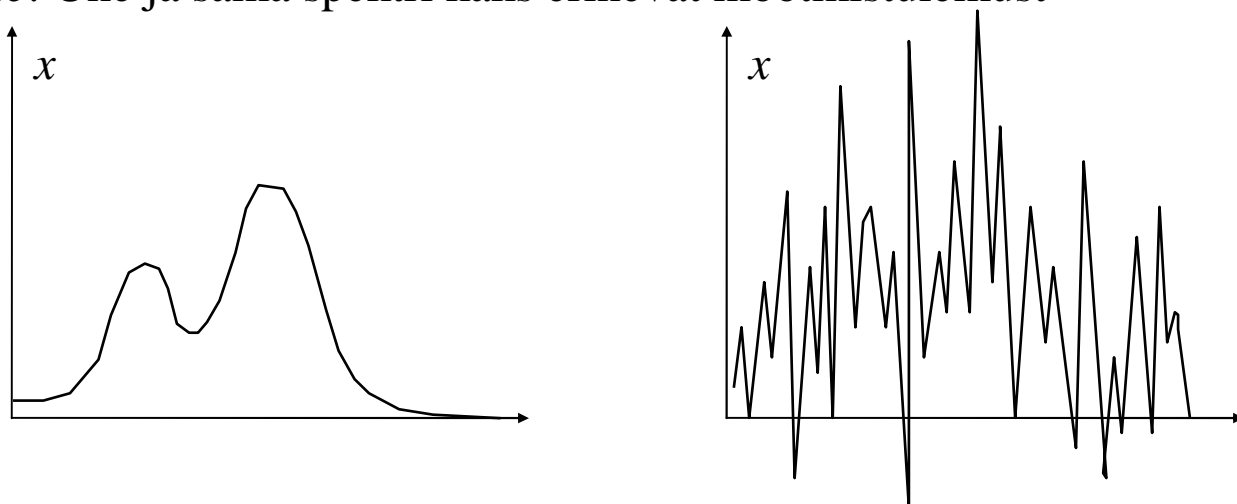
BEGIN
assign (f, '\b\demo.xls'); rewrite (f);
n := nmatrix;
for k := 9 to 30 do begin
  c := k / 30;
  for i := 1 to n do for j := 1 to n do
    g [i,j] := c*exp(-c*sqr(i-j));
  multm (g, g, a); x := invdet (a);
  y := 0; for i := 1 to n do y := y + a [i, i];
  writeln (f, c:9:3, sqrt (y/n):9:3);
end;
close (f);
END.

```



Mõõtmise stabiliseerimine aprioorse informatsiooni abil.

Näide: Ühe ja sama spektri kaks erinevat mõõtmistulemust



Objekti tundmata võib öelda, et teine spekter ei saa olla õige. Väite aluseks on aprioorne informatsioon, mille kohaselt spektrifunktsioonil ei saa olla negatiivseid väärtusi. Objekti tundes lisandub täiendav aprioorne informatsioon, mille kohaselt spektri graafik peaks olema suhteliselt sile kõver. Esimene aprioorne informatsioon ei ole lineaarses mudelis formaliseeritav, teine aga on.

Mõõtmisinformatsiooni kirjeldab spektri hinnangu tõenäosusjaotus

$$f_m(x) = cn \times \exp\left(-\frac{1}{2}(x - \hat{x})^T A_m (x - \hat{x})\right) \quad A_m = G^T D^{-1} G$$

ja mõõtmisvigu iseloomustavat aprioorset informatsiooni kirjeldab

$$f_p(x) = cn \times \exp\left(-\frac{1}{2}(x - \hat{x})^T A_p (x - \hat{x})\right)$$

Kui need informatsioonid on sõltumatud, siis arvutatakse ühendatud informatsioonil baseeruv tõenäosustihedus korrutamise teel

$$f(x) = f_p(x) f_m(x) \quad A = G^T D^{-1} G + A_p$$

Tulemuseks on modifitseeritud algoritm:

$$C = (G^T D^{-1} G + A_p)^{-1} \quad \hat{x} = CG^T D^{-1} y$$

Ebalineaarse aprioorse informatsiooni formaliseerimine ei ole nii lihtne.

Probleem: kuidas kirjutada aprioorne informatsioon üles maatriksi kujul?

Apriorsete võrrandite meetod.

Me alustasime lõplikumõõtmelise aparaadivõrrandi formuleerimist järgmiselt:

Üks kanalisignaali $y_1 = g_{11}x_1 + g_{12}x_2 + g_{13}x_3 + \dots + \xi_1$

Aparaadivõrrand

$$y = \mathbf{G}\mathbf{x} + \boldsymbol{\xi} \quad \text{ehk} \quad y_j = \sum_{i=1}^n G_{ji}x_i + \xi_j$$

Aparaadivõrrand on üksikuid kanalisignaale kirjeldavate skalaarsete võrrandite hulk. Nendele võrranditele võib juurde kirjutada apriorseid võrrandeid:

Mõõtmisvõrrandid: $y_1 = g_{11}x_1 + g_{12}x_2 + g_{13}x_3 + \dots + \xi_1$

$$y_2 = g_{21}x_1 + g_{22}x_2 + g_{23}x_3 + \dots + \xi_2$$

.....

Apriorsed võrrandid: $0 = p_{11}x_1 + p_{12}x_2 + p_{13}x_3 + \dots + \xi_{p1}$

$$0 = p_{21}x_1 + p_{22}x_2 + p_{23}x_3 + \dots + \xi_{p2}$$

.....

Tõkestatuse tingimused on $x_i = 0 + \xi$, σ väärtus on apriorne

Esimese astme sileduse tingimused on

$$x_i - x_{i+1} = 0 + \xi, \quad \sigma \text{ väärtus on apriorne}$$

Teise astme sileduse tingimused on

$$x_i - 2x_{i+1} + x_{i+2} = 0 + \xi, \quad \sigma \text{ väärtus on apriorne}$$

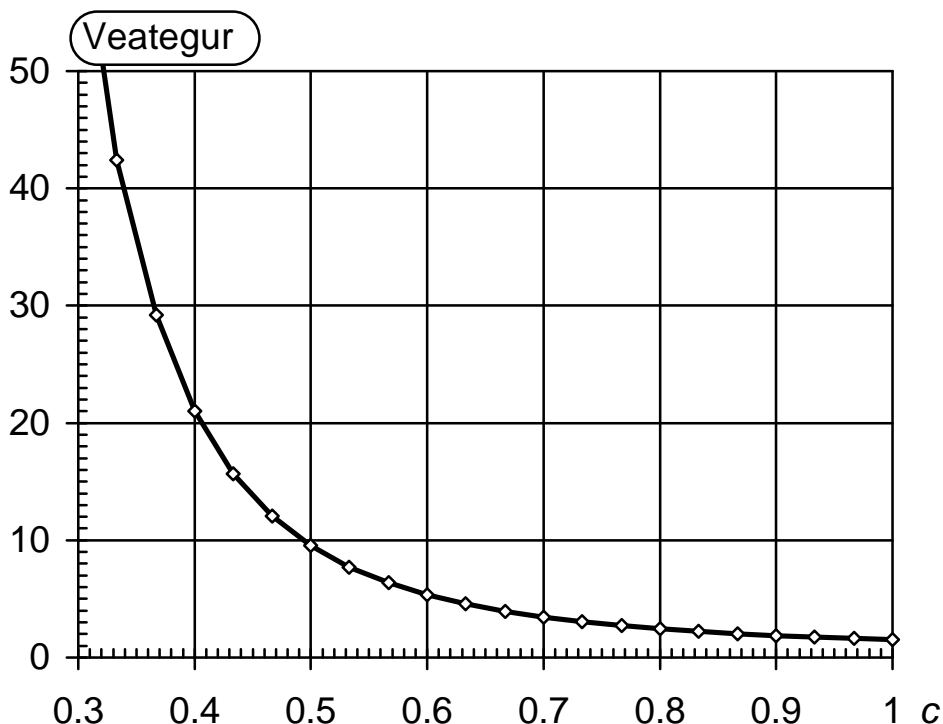
Apriorsetest võrranditest saab apriorse matriksi G_p täpselt samal viisil kui mõõtmisvõrranditest aparaadimatriksi. Kui nüüd edasi jätkata ilma mõõtmisvõrranditeta, saame apriorse informatsioonimatriksi

$$\mathbf{A}_p = \mathbf{G}_p^T \mathbf{D}^{-1} \mathbf{G}_p.$$

Otstarbekam on aga käsitleda arvutamisel kõiki võrrandeid üht ja sama viisi nagu oleks tegemist mõõtmisvõrranditega. Tulemuse poolest on see eelmises punktis käsitletuga ekvivalentne.

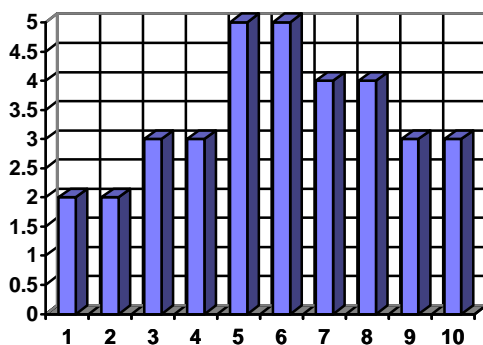
Mõõtmise stabiliseerimine mudeli struktuuri abil.

Meenutame tuttavat graafikut:



Valides diskretiseerimissammuks $c = 0.4$, saame veavõimenduseks 21. Defineerime nüüd aprioorse informatsiooni võrranditega, mis nõuavad spektri naaberpunktide paarikaupa võrdsust ($\sigma = 0$):

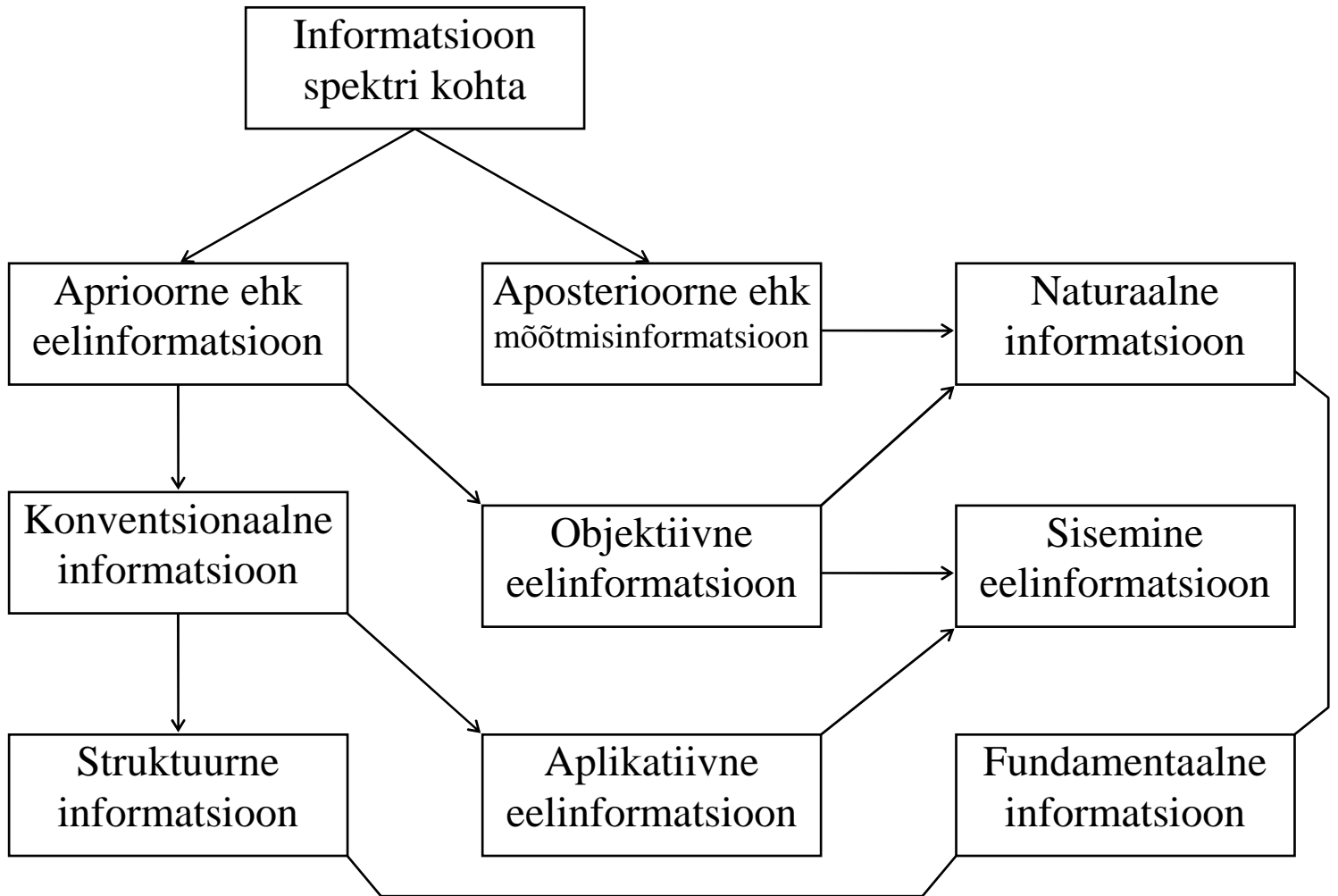
$$x_1 = x_2, \quad x_3 = x_4, \quad x_5 = x_6 \quad \text{jne.}$$



Nõue on samaväärne diskretiseerimissammu kahekordistamisega. Kui $c = 0.8$, saame veavõimenduseks 2.5 mis on 8.5 korda vähem kui algse diskretiseerimissammu puhul. Täpselt sama tulemuse saaksime ka aprioorseid võrrandeid ja/või aprioorset informatsioonimaatriksit kasutades.

Küsimus: Milliseid stabiliseerimismeetodeid oleme käsitlenud ja kui võrdnad sobivad füüsiku jaoks praktiliseks kasutamiseks?

Mõõtmise ja andmeteisenduse informatsiooniallikad



Informatsioon müra kohta?

10. Mõõtmisinformatsiooni hulga hindamine.

Informatsiooni hulga multiplikatiivne mõõt ja aditiivne mõõt.

Teadmise hulka skalaarse suuruse väärtuse kohta mõõdab väärtuse lokaliseerimise vahemiku laius. Vahemik võib olla rangelt või difuusselt piiritletud. Difuusset piiritletust kirjeldab tõenäosusjaotus ja vahemiku laiust jaotuse karakteristik, näiteks standardhälve või selle kordne.

Teadmise hulka ennast ei saa arvuliselt üheselt väljendada, küll aga teadmise hulga muutust. Mõõduks on vahemiku kitsenemise tegur:

$$z = \frac{\Delta x_{alg}}{\Delta x_{lõpp}}$$

Teadmise muutumine tähendab informatsiooni saamist. Tegur z on saadud informatsiooni hulga loomulik mõõt. See mõõt on multiplikatiivne: kui saame algul informatsiooni z_1 ja hiljem z_2 , siis kogu saadud informatsiooni mõõt on

$$z_3 = z_1 \times z_2.$$

Informatsiooni hulk ei kirjelda olemasolevat vaid saadut. Termodünaamika mõttes on informatsiooni hulk protsessifunktsioon.

Informatsiooni hulga tuntuim mõõtühik bitt vastab lokaliseerimise intervalli kahekordsele kitsenemisele.

Mõõtmise korral määravad lokaliseerimise intervalli laiuse algväärtuse eelteadmised mõõdetava suuruse kohta ja lõppväärtuse mõõtmistulemuse vahemikhinnang. Sõltumatute koordinaatide mitmemõõtmelise mõõtmise informatsioonihulgad korrutuvad ja

$$z = \frac{\Delta V_{alg}}{\Delta V_{lõpp}}$$

kus V on mõõdetava vektori lokaliseerimise ruumala.

Harjumuspäraselt soovitakse millegi juurde saamisel juurdesaadu mõõt arv olemasoleva mõõt arvule liita aga mitte korrutada nagu ülalpool. Seepärast eelistatakse informatsiooni hulga aditiivset mõõtu

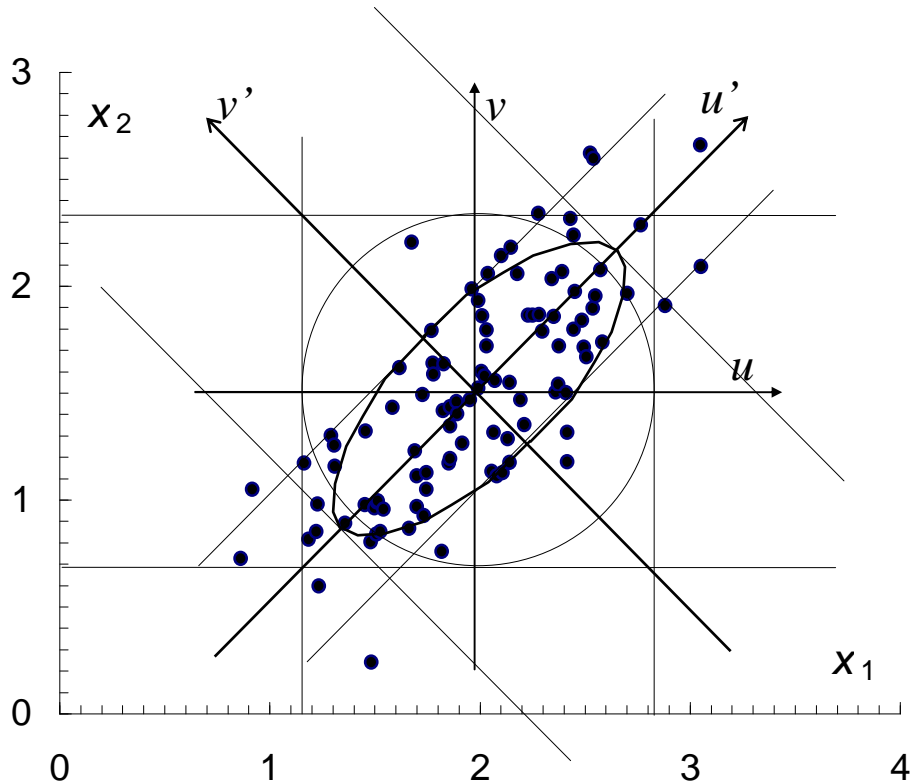
$$I = \log z = \log \frac{\Delta V_{alg}}{\Delta V_{lõpp}}$$

Kui kasutame kahendlogaritmi, siis avaldub I bittides.

Vektormõõtmise informatsioonihulk

Avaldis
$$I = \log z = \log \frac{\Delta V_{alg}}{\Delta V_{lõpp}}$$

kehtib ka sõltuvate koordinaatidega vektori puhul, kuid sel juhul ei ole ΔV enam koordinaatide lokaliseerimise vahemike korrutis



Lokaliseerimise ruumalaks on mõõtmise hajumisellipsoidi ruumala mis sõltub koordinaatide vahelistest korrelatsioonidest. Standardhälvete asemel peaks korrutama ellipsoidi pooltelgi.

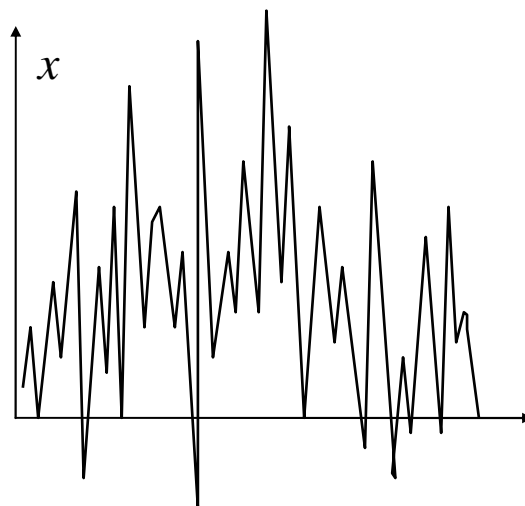
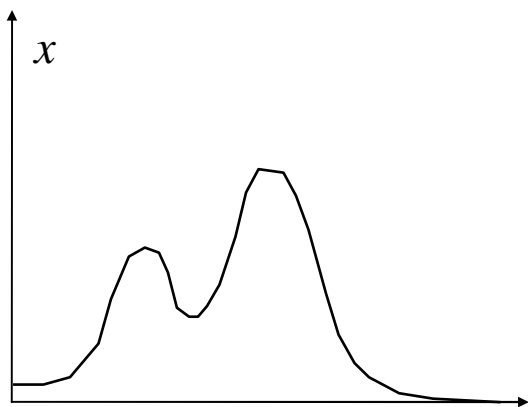
Korreleerimata koordinaatide puhul on C diagonaalne ja standardhälvete korrutis

$$V = \prod_i \sigma_i = \sqrt{\det C}.$$

Korreleeritud koordinaatide puhul saab valida uue koordinaatsüsteemi nii et uued koordinaadid oleks korreleerimata $y' = Ty$, kus T on maatriksi C omavektoritest koostatud ortogonaalpöörde maatriks. Ortogonaalpöörde ei muuda determinanti, $\det C' = \det C$ ning korreleeritud andmete puhul on ikkagi

$$V = \sqrt{\det C}.$$

Formaalne ja kasulik informatsioon

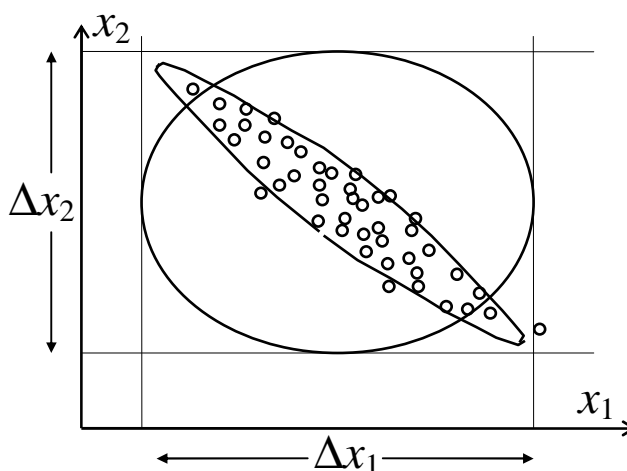


Teine joonis esitab tüüpilist pöördülesande lahendit väikese digitaliseerimissammu puhul. Lahendi vigade kovariatsioonimaatriksis on peadiagonaalil positiivsed suured arvud, naaberkõrvaldiagonaalidel aga negatiivsed suured arvud, mis annavad tunnistust naaberpunktide mõõtmisvigade tugevast negatiivsest korrelatsioonist. Seetõttu ongi võimalik, et naaberpunktide ühendamisel viga kahaneb mitte 1.4, vaid öige mitu korda. Meenutame: $\sigma_{x=y}^2 = \sigma_x^2 + \sigma_y^2 + 2\rho_{xy}\sigma_x\sigma_y$. Seetõttu on kahe naaberpunkti hajumisellipsoid sageli hoopiski väljavenitatum, kui eelmisel joonisel näidatu (vt. eelmine lk.). Formaalne informatsiooni arvutamisel arvestame kitsa ellipsi pindala. Kas vastav informatsioon jõuab spektrigraafiku vaatajani? Vaataja näeb väärtusi ja veahinnanguid aga ei näe veakorrelatsioone. Vastuvõetav ehk kasulik informatsioon on määratud mitte hajumisellipsoidi ruumalaga, vaid üksikkoordinaatide mõõtmisvigadele ehitatud koordinaatteljestiku järgi suunatud telgedega ellipsoidi ruumalaga, mis on võrdeline maatriksi \mathbf{C} diagonaalelementide korrutisega. Informatsioonikadu formaalse tulemuse viimisel selle kasutajani on:

$$\Delta I = \frac{1}{2} \log \frac{\prod_j C_{jj}}{\det \mathbf{C}}$$

Kahe muutuja erijuhul:

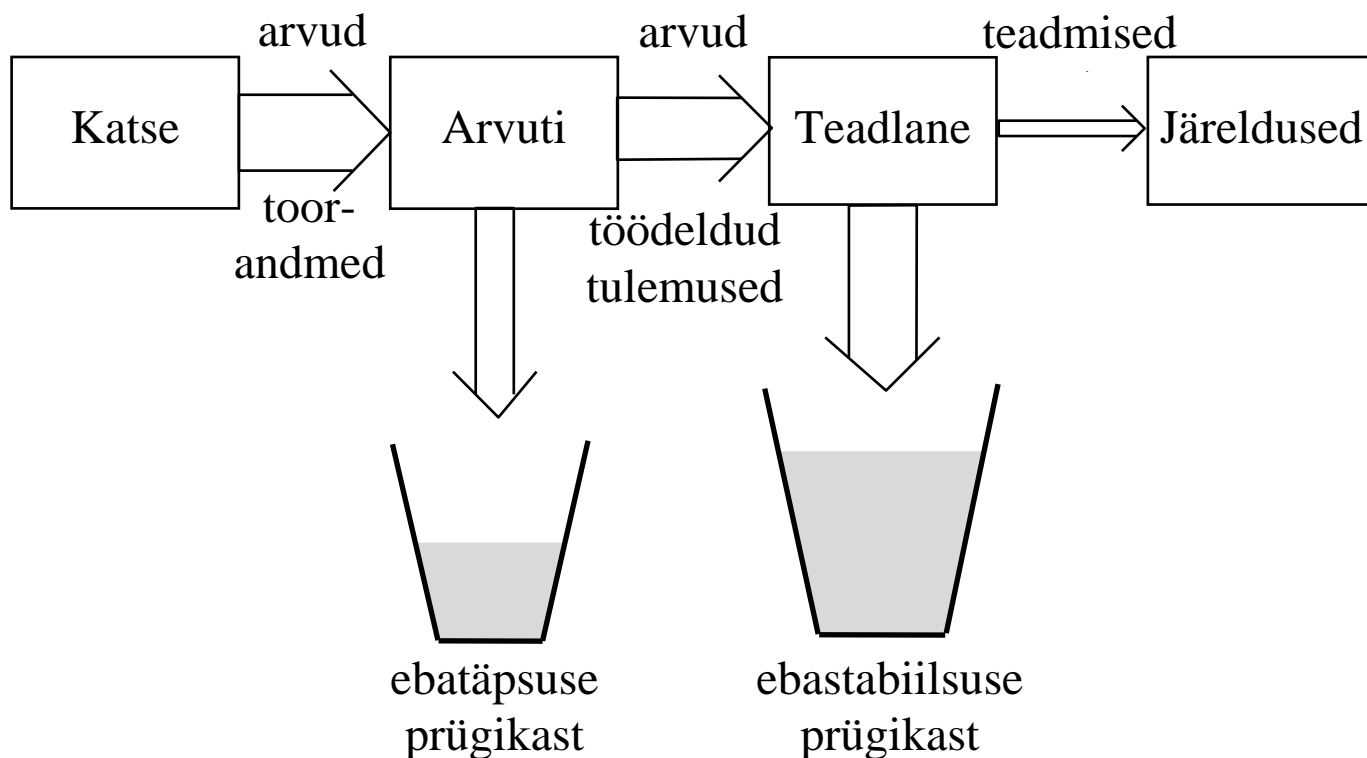
$$\Delta I = \frac{1}{2} \log \frac{1}{1-\rho^2}$$



Informatsiooni teisenemine andmeanalüüsi käigus

Pööratav teisendus ei lisa ega hävita informatsiooni. Optilise spektri kirjeldus lainepikkus-intensiivus-graafiku kujul on informatsiooniliselt võrdväärne selle Fourier kujutisega ehk polükromaatsse valguse interferogrammiga. Kui need kaks kujutist matemaatilise andmetötluse lõppresultaadina osutuvad füüsiku jaoks mittevõrdväärseiks, tähendab see informatsioonikadu tulemuste üleandmisel masinalt inimesele.

Mittepööratav teisendus saab informatsiooni ainult vähendada. Vahetute mõõtmisandmetega edasi antavat formaalse informatsiooni hulka ei ole võimalik ühegi andmetötlusmeetodiga suurendada.



Eesmärk on maksimiseerida lõpp-produkt. Stabiliseerimine vähendab täpsust (mida see tähendab?). Sellega suureneb voog ebatäpsuse prügikasti kuid väheneb voog ebastabiilsuse prügikasti. Ülesandeks on leida optimum.

11. Empiiriliste võrrandite koostamine.

Lähendusvalemid ja empiirilised võrrandid.

Lähendusvalemite üheks tuntud rakenduseks on mittealgebraaliste funktsioonide arvutamine. Suvaline arvuti oskab vahetult teha ainult algebraalisi põhitehteid. Funktsioone alates siinusest ja koosinusest saab arvutada ainult algebraaliste lähendusvalemite abil. Lihtsaim meetod selliste valemite tuletamiseks on funktsiooni astmeritta arendamine ja rea esimeste liikmete kasutamine lähendusvalemina. Arvutusmatemaatika kursustes näidatakse, kuidas lõplike summade kordajaid määrata nii, et saada lõigatud astmereaga võrreldes veidi täpsemaid tulemusi või hoida kokku arvutustööd. Meie kursuses neid probleeme ei käsitleta.

Meie vaatleme probleemi, kus mingi funktsioon on algselt kirjeldatud ainult numbrilise tabeliga ja otsitavaks on funktsiooni ligikaudne analüütiline avaldis.

Algne numbriline tabel võib olla arvutuslikku päritolu, näiteks koostatud mingi diferentsiaalvõrrandi numbrilise lahendamise teel. Niisugust tabelit võib koostada kuitahes täpselt.

Algne numbriline tabel võib olla saadud ka mõõtmiste teel ning varustatud tabeliväärtuse mõõtmisvigade hinnangutega. Sellisel juhul nimetatakse otsitavat funktsiooniavaldist empiiriliseks võrrandiks.

Otsitava funktsiooni matemaatiline struktuur võib olla määratud teoreetiliste kaalutlustega. Tegelikult otsitavateks on siis teoreetilise mudeli parameetrid.

Paraku tuleb sageli ette olukordi, kus otsitava funktsiooni kohta pole pea mingeid eelteadmisi peale mõõtmistulemuste tabeli. Otsitavaks on siis ka funktsiooni matemaatiline struktuur. Funktsiooni matemaatilise struktuuri varieerimiseks on lõpmatu palju võimalusi ja sellise varieerimise kirjeldamiseks ei ole olemas rakendusliku väärtusega teooriat. Funktsiooni struktuuri otsimine on kunst ja meetodiks teoreetilise süsteemita katsetamine.

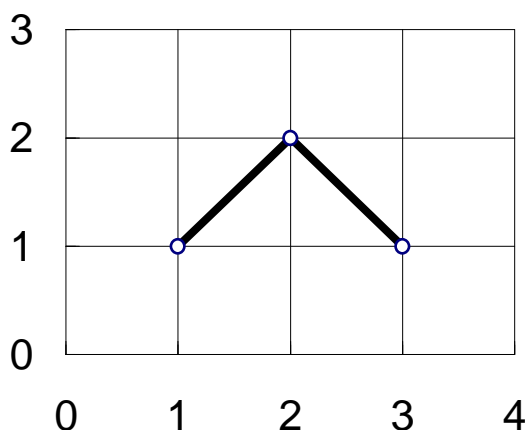
Niiviisi osutub lähendusvalemite koostamine paljuharuliseks õpetuseks, millest me jõuame käsitleda vaid üksikuid valitud küsimusi.

Kahe muutuja lineaarne ülesanne

Vaatleme kaht juhuslikku viga sisaldavat suurust x ja y . Ühe suuruse tinglikuks tõenäosusjaotuseks nimetatakse suuruse jaotust juhul, kui teine suurus pole vaba, vaid fikseeritud kindlal nivool. Kui ühe suuruse tinglik tõenäosusjaotus sõltub teisest suurusest siis on suurused statistiliselt sõltuvad. Üks sõltuvuse kirjeldamise vahendeid on regressioon $\mu_{y/x} = f(x)$. Seda kirjeldavat võrrandit nimetatakse ühe muutuja regressioonmudeliks.

Näide:

x	$\mu_{y/x}$
1	1
2	2
3	1



Lihtsaimal erijuhul on regressioon lineaarne

$$\mu_{y/x} = ax + b, \quad \mu_{x/y} = cx + d.$$

Tõenäosusteoorias näidatakse, et

$$a = \rho_{xy} \sigma_y / \sigma_x, \quad b = \mu_y - a \mu_x.$$

Lineaarse regressioonikonstantide hindamiseks piisab arvutada tavalised ühemõõtmelise analüüsi statistikud ja korrelatsioonikordaja hinnang.

Kudas seda teha, on kirjeldatud juba varem.

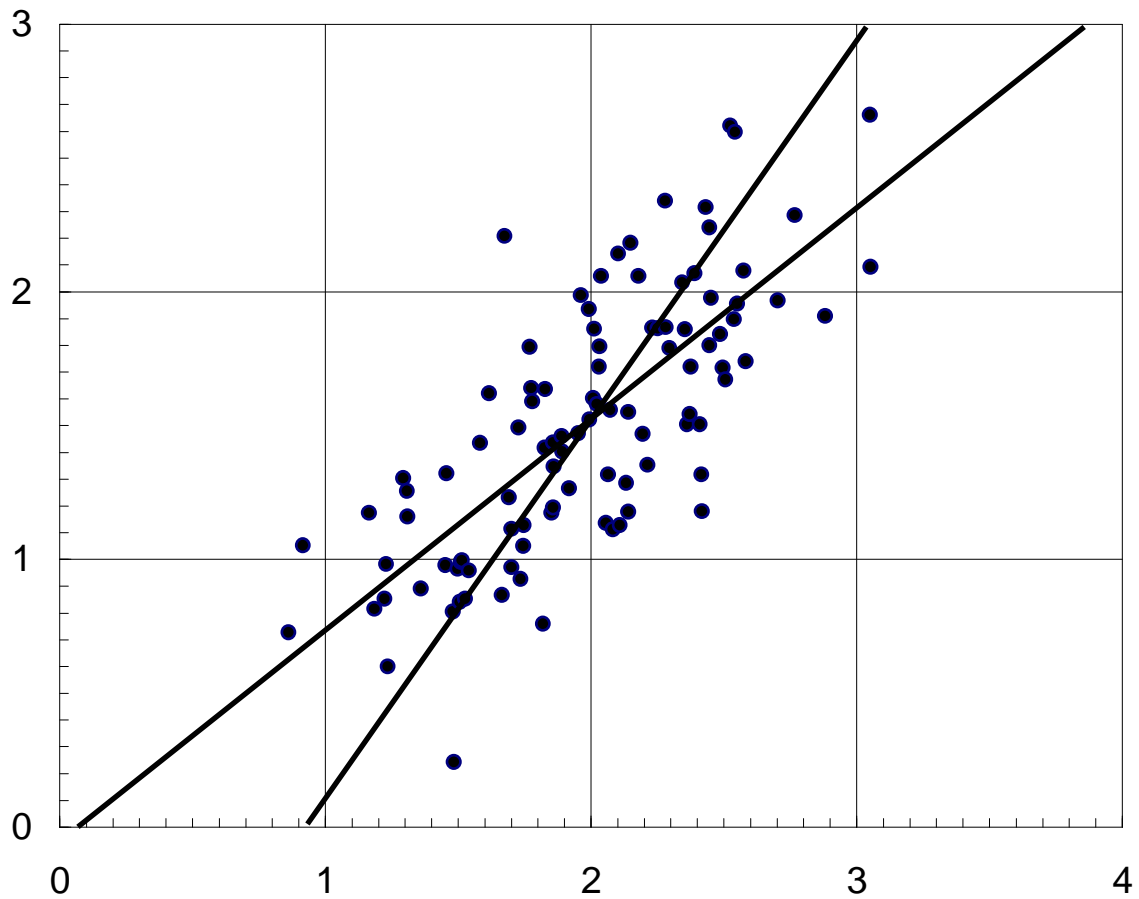
Hälbe $y - (ax + b)$ ruudu keskväärtuse ruutjuurt nimetatakse jääkhälbeks. Jääkhälve on

$$\sigma_{y/x} = \sigma_y \sqrt{1 - \rho_{xy}^2}$$

Hälvete suhe mõõdab lineaarse prognoosi informatsioonihulka

$$I = \log \frac{\sigma_y}{\sigma_{y/x}} = \frac{1}{2} \log \frac{1}{1 - \rho_{xy}^2}$$

Näide:



$$m_x = 1.98, \quad m_y = 1.51, \quad s_x = 0.46, \quad s_y = 0.49, \quad r = 0.742,$$

$$\mu_{y/x} \approx 0.788x - 0.053, \quad \mu_{x/y} \approx 0.698y + 0.927$$

$$\sigma_{y/x} = \sigma_y \sqrt{1 - \rho_{xy}^2} \approx 0.49 \times 0.67 = 0.33$$

Regressioonprognoosi $y \approx 0.788x - 0.053$ $I = 0.58$ bitti.

Mitmene lineaarne regressioonimudel.

Lähteandmete tabel:

j	x_1	x_2	x_n	y_j	$\sigma(y_j)$
1	2.34	14.7	-2.3	18.6	0.9
2	3.12	13.1	-0.2	22.3	1.5
3	0.75	9.9	1.7	31.0	1.3
...

Homogeenne mudel

$$y = \sum_{i=1}^n a_i x_i + \xi$$

Iga üksikmõõtmise jaoks kirjutatuna

$$y_j = \sum a_i x_{ji} + \xi_i = \sum x_{ji} a_i + \xi_i$$

Tähistame

$$\mathbf{X} = \{x_{ji}\} \quad \mathbf{D} = \text{cov}(\xi)$$

Saame

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \xi$$

$$\mathbf{C} := (\mathbf{X}^T \mathbf{D}^{-1} \mathbf{X})^{-1}$$

$$\mathbf{a} := \mathbf{C} \mathbf{X}^T \mathbf{D}^{-1} \mathbf{y}$$

Mittehomogeenses mudelis on lisaks vabaliige

$$y = a_0 + \sum_{i=1}^n a_i x_i + \xi$$

Sellise mudeli saab formaalselt taandada homogeenseks mudeliks valides

$$i = 0 \dots n \quad \text{ja} \quad x_0 \equiv 1$$

Ühe muutuja funktsiooni lähendusvalemi koostamine.

Lihtsaim moodus ebalineaarse funktsiooni kirjeldamiseks on valida mingi baasfunktsioonide komplekt $\{f_i(x)\}$ ja esitada $y \approx y^* = f(x)$ koordinaatfunktsioonide kaalutud summa kujul:

$$y^* = a_1 f_1(x) + a_2 f_2(x) + a_3 f_3(x) + \dots + a_n f_n(x)$$

Näited:

$$n = 2, f_1(x) = x, f_2(x) = 1,$$

$$f_i(x) = x^{i-1},$$

$$n = 5, f_1(x) = 1, f_2(x) = \sin(x), f_3(x) = \cos(x), f_4(x) = \sin(2x), f_5(x) = \cos(2x).$$

Koordinaatfunktsioonide väärtused x fikseeritud tasemetel on arvud:

$$x_{ji} = f_i(x_j) \quad (i \text{ numereerib koordinaatfunktsioone ja } j \text{ katsepunkte})$$

ja prognoosivõrrand saab tuttava kuju:

$$y_j \approx y_j^* = a_1 x_{j1} + a_2 x_{j2} + \dots + a_n x_{jn} = \sum_{i=1}^n a_i x_{ji} = \sum_{i=1}^n x_{ji} a_i, \quad \mathbf{y}^* = \mathbf{Xa}$$

Siinkohal jääb probleemiks: kuidas valida koordinaatfunktsioone?

Mõned lihtsamad ebalineaarsed mudelid taanduvad teistsugusel viisil lineaarmudeliks.

$$\text{Näited: } y = ax^b \rightarrow \ln y = \ln a + b \ln x, \quad y = ab^x \rightarrow \ln y = \ln a + x \ln b,$$

Ebalineaarse ja lineaarmudeliks mittetaandatava parameetrilise mudeli

$$y = f(x|a_1, a_2, \dots, a_m) + \xi$$

parameetrite otsimiseks vähimruutude meetodil tarvis lahendada arvutusmahukas ekstreemumülesanne, mis sõltumatute mõõtmisvigade puhul on

$$g(a_1, a_2, \dots, a_m) = \sum_j \frac{(y_j - f(x_j|a_1, a_2, \dots, a_m))^2}{\xi_j^2} \rightarrow \min$$

Näide: SigmaStat.

Sageli on ebalineaarne mudel pärit teoreetilistest kaalutlustest ja sihitu otsing pole ei tarvilik ega õigustatud. Niisugustel juhtudel räägitakse ka poolempiirilistest mudelitest.

Näide: koordinaatfunktsioonide intuiitiivne otsing.

(K41–43)

Kommentaariid: olgugi, et programm eeldab diagonaalset veamaatriksit ja käsitleb veadispersioonide pöördväärtusi kui kontrollhälvete

$$\delta_j = y_j^* - y_j$$

kaalusid p . Kasutajale esitatav ruutkeskmise hälve on aga arvatud kaalusid arvestamata. Täiendavalt esitatakse absoluutväärtuse poolest maksimaalne hälve, Abbe statistik ja sellest tuletatud ligilähedaselt normaaljaotusega statistik. Abbe statistik arvutatakse eeldusel, et baas sisaldab konstantset koordinaatfunktsiooni ja andmetabeli ridade kaalud on võrdsed.

$$A = \frac{1}{2} \frac{\sum_{j=1}^{m-1} (\delta_{j+1} - \delta_j)^2}{\sum_{j=1}^m \delta_j^2} \qquad N = (A-1) \sqrt{\frac{2m+1}{2-(A-1)^2}}$$

41

K42

K43

Statistilise prognoosi ülesanne.

Prognoositav suurus: *prediktant*, tähistus y

Prognoosi argumendid: *prediktorid*, tähistus x_1, x_2, \dots, x_n

Kasutame varasemat avaldist, mis kirjeldas ühe muutuja lineaarset prognoosi mitme prediktandi järgi:

$$y \approx y^* = a_0 + \sum_{i=1}^n a_i x_i = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

Kordajate a_k hinnangud arvutatakse reeglina vähimruutude meetodi abil.

Vähimruutude meetodi lahend maksimiseerib $\rho(y, y^*)$. ja minimiseerib jääkhälbe

$$\sigma_y^* = \sigma_y \sqrt{1 - R_{y,x}^2},$$

kus R on mitmese korrelatsiooni koefitsient.

NB! Jääkhälvet võib tõlgendada kui prognoosi kontrollviga aga mitte kui prognoosiviga!

Kuidas võtta arvesse prediktorite ebalineaarseid efekte ja prediktorite koosmõjuefekte?

Tuletatud prediktorid $x_{n+1}, x_{n+2}, x_{n+3}, \dots$

Näited: $x_2^2, \sqrt{x_5}, \ln(x_3), x_2 x_7$ jne. Lõpmatult palju võimalusi.

Tuletatud prediktorite lineaarse sõltumatuse nõue.

Põhiprobleem: prediktorite valimine.

Konkureerivad “alt üles” ja “ülalt alla” meetodid.

Näide komplitseeritud mudelist: iooni liikuvuse arvutamine.

[J. Aerosol Sci. **26**, 459–475 (1995)]

$$B = f_1 f_2 \frac{1 + \frac{l}{\delta} \left[a + b \exp\left(-c \frac{\delta}{l}\right) \right]}{6\pi\eta\delta} \quad \delta = r_m + h + r_g(T_\delta)$$

$$f_1 = \sqrt{1 + \frac{m_g}{m_p}} \quad f_2 = \frac{2.25}{(a + b) \left(\Omega_{\infty-4}^{(1,1)*}(T^*) + s(r_m, T_\delta) - 1 \right)}$$

$$\rho = 2.07 \text{ g cm}^{-3} \quad h = 0.115 \text{ nm} \quad r_{cr} = 1.24 \text{ nm.}$$

```

function Mobility
  {velocity/force } (GasMass {amu}, { air nitrogen}
  { (m/s) / fN } Polarizability {nm3}, { 28.96 28.02 }
  VisCon1 {nm}, { 0.00171 0.00174}
  VisCon2 {K}, { 0.3036 0.2996 }
  VisCon3, { 44 40 }
  Pressure {mb}, { 0.8 0.7 }
  Temperature {K},
  ParticleDensity {g cm-3},
  ParticleCharge {e},
  MassDiameter {nm} : real) : real;

function Omega11 (x : real) : real; { $\Omega(1,1)^*(T^*)$  for ( $\infty-4$ ) potential}
  var p, q : real; {and elastic-specular collisions}
  begin
    if x > 1 then Omega11 := 1 + 0.106 / x + 0.263 / exp ((4/3) * ln (x))
    else begin p := sqrt (x); q := sqrt (p); Omega11 := 1.4691 / p - 0.341 / q + 0.181 * x * q + 0.059 end;
  end;

const a = 1.2; b = 0.5; c = 1; {the slip factor coefficients}
  ExtraDistance = 0.115 {nm}; TransitionDiameter = 2.48 {nm};
var GasDiameter, MeanVelocity, Viscosity, FreePath, DipolEffect,
  DeltaTemperature, CheckMark, ParticleMass, CollisionDistance, Kn, Omega, s, x, y : real;
begin
  Viscosity { $\mu$ Pa s} := 0.02713 * sqrt (GasMass * Temperature) /
    sqrt (VisCon1 * (1 + exp (VisCon3 * ln (VisCon2 / Temperature))));
  MeanVelocity {m/s} := 145.5 * sqrt (Temperature / GasMass);
  FreePath {nm} := (166251 * Viscosity * Temperature) / (GasMass * Pressure * MeanVelocity);
  ParticleMass {amu} := 315.3 * ParticleDensity * exp (3 * ln (MassDiameter));
  DeltaTemperature := Temperature;
  repeat
    CheckMark := DeltaTemperature;
    GasDiameter {nm} := VisCon1 * (1 + exp (VisCon3 * ln (VisCon2 / DeltaTemperature)));
    CollisionDistance {nm} := MassDiameter / 2 + ExtraDistance + GasDiameter / 2;
    DipolEffect := 8355 * ParticleCharge * Polarizability / sqrt (sqrt (CollisionDistance));
    DeltaTemperature := Temperature + DipolEffect;
  until abs (CheckMark - DeltaTemperature) < 0.01;
  if ParticleCharge = 0 then Omega := 1 else Omega := Omega11 (Temperature / DipolEffect);
  Kn := FreePath / CollisionDistance;
  if Kn < 0.03 {underflow safe} then y := 0 else y := exp (- c / Kn);
  x := (273.15 / DeltaTemperature) * exp (3 * ln (TransitionDiameter / MassDiameter));
  if x > 30 {overflow safe} then s := 1
  else if x > 0.001 then s := 1 + exp (x) * sqrt (x / (exp (x) - 1)) * (2.25 / (a + b) - 1)
  else {underflow safe} s := 1 + (2.25 / (a + b) - 1);
  Mobility := ((2.25 / (a + b)) / (Omega + s - 1)) * sqrt (1 + GasMass / ParticleMass) *
    (1 + Kn * (a + b * y)) / (6 * PI * Viscosity * CollisionDistance);

```

end;

The empirical data by Kilpatrick and the best fit of the data

Mass (amu)	Sign	Mobility ($\text{cm}^2\text{V}^{-1}\text{s}^{-1}$)			Error (%)
		reduced	measured	calculated	
35.5	-	2.49	4.31	4.41	2.2
46	-	2.36	4.09	4.00	-2.3
60	-	2.06	3.57	3.61	1.3
60	+	2.07	3.59	3.61	0.8
84	+	1.84	3.19	3.19	0.0
90	-	1.83	3.17	3.11	-2.0
114	-	1.69	2.93	2.85	-2.8
114	+	1.73	3.00	2.85	-5.0
127	-	1.52	2.63	2.74	3.8
130	+	1.56	2.70	2.71	0.3
152	-	1.46	2.53	2.56	1.1
152	+	1.44	2.50	2.56	2.5
168	+	1.47	2.55	2.46	-3.4
188	+	1.34	2.32	2.36	1.6
215	+	1.21	2.10	2.24	6.8
222	-	1.27	2.20	2.21	0.5
240	+	1.24	2.15	2.15	0.1
244	+	1.27	2.20	2.13	-3.1
244	+	1.28	2.22	2.13	-3.8
256	+	1.26	2.18	2.09	-4.1
292	-	1.14	1.98	1.99	0.6
292	-	1.13	1.96	1.99	1.5
292	+	1.14	1.98	1.99	0.6
295	-	1.19	2.06	1.98	-4.0
361	-	1.06	1.84	1.83	-0.6
407	-	0.984	1.71	1.74	2.0
430	-	0.976	1.69	1.70	0.6
430	+	0.976	1.69	1.70	0.6
499	-	0.906	1.57	1.60	1.8
499	+	0.906	1.57	1.60	1.8
511	-	0.922	1.60	1.58	-1.0
660	-	0.820	1.42	1.41	-0.6
660	+	0.842	1.46	1.41	-3.2
1061	-	0.607	1.05	1.10	4.5
1612	-	0.497	0.86	0.84	-2.1
2122	-	0.411	0.71	0.71	-1.0

Näide ebalineaarse mudeli otsingust.

[TÜ Toimetised **947**, 136–149 (1992)]

Käsitsi joonistatud leht A (teooria)

Käsitsi joonistatud leht B (mõõtmejaotuse mudelid)

Käsitsi joonistatud leht C (lähendamise kvaliteedi hinnangud)

```

Procedure KLOR (number : integer; {number of distribution points}
  var radius ; {array of radii [1..number], real}
  var dndlnr ; {array of dn/d(lnr) [1..number], real}
  var relerr ; {array of relative errors [1..number], real}
  {Results are four scalar parameters
  of the KL0 model and a measure of
  relative approximation error}
  var      K, {nondimensional}
          L, {nondimensional}
          a, {unit the same as for dn/d(lnr)}
          rx, {unit the same as for radius}
  approxerr : real);

{KL0 model: dn/d(lnr) = r*dn/dr = a / ((r/rx)^K + (rx/r)^L),
approxerr = sqrt (average (sqr (reldev / relerr))), where
reldev = ((approximated dn/d(lnr)) - (given dn/d(lnr))) /
          dn/d(lnr)

Productivity : about 1000 15-point spectra per minute
                when using a 100 MHz 486DX computer}

Const nmax = 100; {technical limit 100-point presentation}
Var      r : array [1..nmax] of real absolute radius;
          fr : array [1..nmax] of real absolute dndlnr;
          re : array [1..nmax] of real absolute relerr;
          lnr, {ln (r)}
          y, {1/fr}
          w {sqr (fr/re)}
          : array [1..nmax] of real;
          ww, {scaling factor}
          p, q, {coefficients of linear submodel}
          level {measure of deviation}
          : real;
          blind : boolean;

procedure test (testK, testL : real);
  var      f, g, wi,
          wff, wfg, wgg, wfy, wgy,
          testp, testq, det, rhs : real;
          ff, gg : array [1..nmax] of real;
          i : integer;

  begin
    rhs := 1E33;
    if abs (K + L) > 0.1 then begin
      wff := 0; wfg := 0; wgg := 0; wfy := 0; wgy := 0;
      for i := 1 to number do begin
        f := exp (testK * lnr [i]); g := exp (-testL * lnr [i]);
        ff [i] := f; gg [i] := g;
        wi := w [i] / ww;
        wff := wff + wi * f * f;
        wfg := wfg + wi * f * g;
        wgg := wgg + wi * g * g;
        wfy := wfy + wi * f * y [i];
        wgy := wgy + wi * g * y [i];
      end;
      det := wff * wgg - wfg * wfg;
      testp := (wfy * wgg - wgy * wfg) / det;
      testq := (wff * wgy - wfg * wfy) / det;
      rhs := 0;
      for i := 1 to number do
        rhs := rhs + w [i] * sqr (testp * ff [i] + testq * gg [i] - y [i]);
      rhs := sqrt (rhs / number);
    end;
    if rhs < level then begin
      blind := false;
      level := rhs;
      K := testK;
      L := testL;
      p := testp;
      q := testq;
    end;
  end;
end;

```

```

Var i : integer;
    h : real;
Begin
h := 0.256;
K := 2.75; L := 0.5; ww := 0;
for i := 1 to number do begin
    lnr [i] := ln (r [i]);
    y [i] := 1 / fr [i];
    w [i] := sqr (fr [i] / re [i]);
    if w [i] > ww then ww := w [i];
end;
level := 2E33; test (K, L);
repeat
    blind := true;
    test (K + h, L);
    if blind then begin
        test (K - h, L);
        if blind then begin
            test (K, L + h);
            if blind then begin
                test (K, L - h);
                if blind then begin
                    test (K + h, L + h);
                    if blind then begin
                        test (K - h, L - h);
                        if blind then begin
                            test (K - h, L + h);
                            if blind then test (K + h, L - h);
                        end;
                    end;
                end;
            end;
        end;
    end;
end;
end;
end;
end;
end;
if blind then h := h / 8;
until h < 0.0004;
rx := exp (ln (q / p) / (K + L));
a := 1 / (p * exp (K * ln (rx)));
approxerr := level;
End;

```