Uz datizraces kursu

Varbūtību teorija un matemātiskā statistika

(atkārtojums – 1.lekcija)

Kārlis Podnieks, LU

Datizraces kursa vieglākai studēšanai ir vēlams atkārtot bakalaura programmas kursa "Varbūtību teorija un matemātiskā statistika" pašus pamatus. Varu ieteikt atkārtošanu sākt ar manu grāmatiņu "Varbūtības".

Labs "atgādinājums" par tālāko (kas nav aprakstīts vidusskolām domātajā grāmatiņā), ir atrodams kursa grāmatas pielikumā (*Appendix: Random Variables*). Bet ļoti labi ilustrētus tekstus par šim lietām var atrast *Wikipedia*.

Ko tad vajadzētu atkārtot? Sāksim ar klasiķa citātu:

Leonard J Savage. Recent Tendencies in the Foundations of Statistics, 1958, *The International Congress of Mathematicians* (was held in Edinburgh from 14 August to 21 August 1958).

1) Etymologically, 'statistics' refers to numerical data about the state.

2) ... [statistics is] **quantitative thinking about uncertainty** as it affects scientific and other investigations.

Varbūtību telpas

Varbūtības un varbūtību teoriju nevajag mistificēt, tā ir tikai viens no modeļu būves līdzekļiem. Un tā radās 17.gs. no azartspēļu analīzes...

Varbūtību telpa (*probability space*): diskrētās (galīgās un sanumurējamās) un nepārtrauktās varbūtību telpas, Kolmogorova aksiomas. Sk. http://en.wikipedia.org/wiki/Probability_space.

Varbūtību telpas kalpo par **modeli**, kas formalizē intuitīvo varbūtības jēdzienu, kas matemātiķiem jau sen bija izveidojies. Šo formalizāciju pirmo reizi ieveda 20.gs. 30-jos gados krievu matemātiķis Andrejs Kolmogorovs.

Ja, prātojot par kādu situāciju, Jūs izmantojat terminu "varbūtība", tad precīzu jēgu Jūsu prātojumi iegūs tikai tad, ja definēsiet atbilstošo varbūtību telpu. Vienkāršākajos gadījumos derēs diskrēta varbūtību telpa, sarežģītākajos vajadzēs ņemt palīgā vispārīgo versiju.

Diskrēta varbūtību telpa ir vienkāršāka. Tā sastāv no (seko t.s. *Kolmogorova aksiomas*):

a) Galīgas vai sanumurējamas kopas E (tās elementus sauc par **elementārajiem notikumiem**).

b) Funkcijas P: E→R, kas katram elementāram notikumam *e* (kopas E elementam), piekārto reālu skaitli *P(e)* (to sauc par notikuma e **varbūtību**).

c) Izpildās divi nosacījumi:

$$P(e) \geq 0 \text{ visiem } e\text{; un } \sum_{e \in E} P(e) = 1 .$$

Kopas E apakškopas, t.i. elementāro notikumu kopas sauc par **notikumiem**. Ja A ir notikums (t.i. $A \subseteq E$), tad tā varbūtību definē šādi:

$$P(A) = \sum_{e \in A} P(e) .$$

Protams, $P(E) = 1$.

Piemērs − spēļu kauliņa mešana.

E = {1, 2, 3, 4, 5, 6}, $P(i)=\frac{1}{6}$ visiem i.

Varam aplūkot tādus notikumus kā "uzkrīt skaitlis, kas dalās ar 3" − tā ir kopa {3, 6}, bet varbūtība: $\frac{1}{6}+\frac{1}{6}=\frac{1}{3}$.

Toties vispārīgajā gadījumā:

Piemērs − šautriņas mešana segmentā [0, 1]. Varbūtība trāpīt konkrētā punktā ir 0. Bet varbūtība trāpīt segmentā $[\frac{1}{3},\frac{1}{2}]$? Ja trāpījumu izkliede ir vienmērīga, tad varbūtība trāpīt segmentā $[\frac{1}{3},\frac{1}{2}]$ būs segmenta garums $\frac{1}{2}-\frac{1}{3}=\frac{1}{6}$.

Tātad vispārīgajā gadījumā notikumu varbūtības neizdodas reducēt uz elementāro notikumu varbūtībām (tās visas iznāk vienādas ar 0), un tāpēc ir vajadzīgs sarežģītāks modelis, kurā varbūtības tiek piekārtotas uzreiz elementāro notikumu kopām. Protams, šis piekārtojums nevar būt patvaļīgs, piemēram, ja kopas A un B nešķeļas, tad varbūtībai $P(A\cup B)$ jābūt vienādai ar $P(A)+P(B)$.

Vispārīgajā gadījumā **varbūtību telpa** ietver (seko t.s. *Kolmogorova aksiomas*) trīs lietas (E, V, P), kur:

a) Kopu E (tās elementus sauc par **elementārajiem notikumiem**), kas var nebūt sanumurējama. [Piemērs: reālo skaitļu segments [0,1].)

b) Kopu V, kas sastāv no kopas E apakškopām. Kopas V elementus (t.i. elementāro notikumu kopas) sauc par **notikumiem**. Kopai V ir jābūt slēgtai pret papildinājuma operāciju (jo katram notikumam A atbilst arī notikums ne-A) un galīga vai sanumurējama skaita kopu apvienojuma operāciju.

[Pārliecinieties, ka tad kopa V ir slēgta arī pret galīga vai sanumurējama skaita kopu šķēluma operāciju.]

Kopas V ar minēto īpašību matemātiķi sauc par "sigma-algebrām".

c) Funkciju P: V→R, kas katram notikumam A (kopas V elementam), piekārto reālu skaitli P(A), to sauc par notikuma A **varbūtību**.

d) Izpildās 3 nosacījumi:

$$P(A) \geq 0 \quad \text{visiem A;}$$
$$P(E) = 1 \;;$$

Ja A = UA_i, un kopas A_i savstarpēji nešķeļas un to skaits ir galīgs vai sanumurējams, tad

$$P(A) = \sum_i P(A_i) \;.$$

Funkciju P ar šīm īpašībām matemātiķi sauc par "sigma-aditīvu mēru" kopā V.

[Kolmogorova nopelns: viņš pamanīja, ka ar tikko minētajām īpašībām pietiek, lai pierādītu (kā teorēmas) **visas** varbūtību īpasības, kas tiek izmantotas, pierādot varbūtību teorijas teorēmas.]

Piemērs − šautriņas mešana segmentā [0, 1].

$$E = [0,1] = \{x \in R \mid 0 \leq x \leq 1\} \;,$$

kopa V satur visus [0, 1] apakšsegmentus [a, b] un tā ir minimālā E apakškopu kopa, kas ir noslēgta pret papildinājuma un galīga vai sanumurējama skaita kopu apvienojuma operācijam,

$$P[a, b] = b - a$$

(t.i. mēs uzskatām, ka iespējas trāpīt visos punktos ir vienādas).

Visu elementāro notikumu varbūtības te ir 0:

$P(\{c\})=0$ jebkuram punktam c.

Gadījuma lielumi

Gadījuma lielums (*random variable*): varbūtību sadalījuma (*distribution*) funkcija, varbūtību blīvuma (*density*) funkcija, šo funkciju grafiku analīze. Sk. http://en.wikipedia.org/wiki/Random_variable, http://en.wikipedia.org/wiki/Probability_density_function.

Formālā definīcija: ja ir dota varbūtību telpa (E, V, P), tad gadījuma lielums X ir funkcija X: E→R, t.i. funkcija, kas katram elementāram notikumam piekārto reālu skaitli.

Piemērs − divu spēļu kauliņa mešana.

E = {(i, j) | i, j = 1, 2, 3, 4, 5, 6}, $P(i,j)=\frac{1}{36}$ visiem pāriem i, j.

Gadījuma lieluma piemērs: X = uzkritušo punktu summa.
X pieņem vērtības no 2 līdz 12.

$$P(X=7)=\frac{6}{36}=\frac{1}{6}$$

[6 kombinācijas: 1+6, 2+5, 3+4, 4+3, 5+2, 6+1].

Gadījuma lieluma X **varbūtību sadalījuma funkcija:**

$$F(x)=P(X\leq x) \ ,$$

tās grafika izskats: līnija, kas "aug" no 0 līdz 1.

$$P(a\leq X\leq b)=F(b)-F(a)+P(X=a) \ .$$

Piemērs: divu kauliņu mešana, punktu summa. [Bilde.]

Ja varbūtību sadalījuma funkcijai eksistē F(x) atvasinājums p(x), tad to sauc par **varbūtību**

blīvuma funkciju. Tad:

$$F(x)=\int_{-\infty}^{x} p(x)\,dx \;.$$ [Bilde.]

Blīvuma funkcijas **grafikā**: laukums zem līknes = 1, laukums starp vertikālām līnijām y=a un y=b – varbūtība

$$P(a\leq X\leq b)=\int_{a}^{b} p(x)\,dx \;.$$

$\int_{-\infty}^{+\infty} p(x)\,dx=1$ – tikai, izpildot šo nosacījumu, p(x) var kalpot par sadalījuma blīvuma funkciju. [Šo intuīciju vajag "kopt".]

Diraka delta-funkcija (nav jālasa)

Ja varbūtību sadalījuma funkcija *ne visur ir diferencējama*, tad kas notiek ar sadalījuma blīvumu?

Piemēram, visai parasts gadījuma lielums X, kas vienmēr pieņem tikai vērtību 0.

Tā sadalījuma funkcija:

ja x<0, tad F(x)=P(X≤x)=0;

ja x≥0, tad F(x)=P(X≤x)=1.

Blīvums būtu atvasinājums, tātad šeit:

δ(x)=F'(x)=0, ja x≠0;

δ(x)=F'(x)=+∞, ja x=0.

Tā arī ir slavenā Diraka delta-funkcija.

Sk. http://en.wikipedia.org/wiki/Dirac_delta_function.

Ko ar to var iesākt? Pirmkārt, uzskatīsim, ka

$$\int_{-\infty}^{+\infty} \delta(x)\,dx=1.$$

Otrkārt , ja varbūtību sadalījuma blīvums ir δ, tad uzskatīsim, ka funkcijas f(x) vidējā vērtība būs:

$$\int_{-\infty}^{+\infty} f(x)\delta(x)\,dx=f(0) \;;$$

$$\int_{-\infty}^{+\infty} f(x)\delta(x-a)\,dx=f(a) \;.$$

Protams, δ(x) nav funkcija šī vārda matemātiskajā nozīmē. Tā ir **vispārinātā**

funkcija:

http://en.wikipedia.org/wiki/Generalized_function .

Vispārinātās funkcijas definē kā “parasto” funkciju robežas:

$$f(x) = \lim_{t \to +\infty} g(x,t) \quad \text{(tas ir fakts);}$$

$$\int_{-\infty}^{+\infty} f(x)\,dx = \lim_{t \to +\infty} \int_{-\infty}^{+\infty} g(x,t)\,dx \quad \text{(šī ir definīcija!);}$$

utt.

Piemērs: δ(x).

Vidējā vērtība

Gadījuma lieluma vidējā vērtība (*expected value, mathematical expectation, mean value*)**, dispersija** (*variance*, t.i. "mainība"). Sk. http://en.wikipedia.org/wiki/Expected_value, http://en.wikipedia.org/wiki/Variance.

Piemērs – spēļu kauliņa mešana. Te lielums K pieņem vērtības 1, 2, 3, 4, 5, 6 ar vienādām varbūtībām $\frac{1}{6}$. Tad (teorētiskajā modelī) vidējā vērtība būs “svērtā summa”:

$$E(K) = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 5\cdot\frac{1}{6} + 6\cdot\frac{1}{6} = \frac{21}{6} = 3{,}5 \ .$$

Ko tas nozīmē praksē? Mēs domājam, ka ja metīsim kauliņu, piemēram, 1000 reizes, tad visbiežāk summā sanāks skaitlis, kas tuvs 350 (sk. tālāk *lielo skaitļu likumu*).

Nianse. "Vidējo vērtību" var iegūt divos veidos:

a) no teorētiska modeļa – tad to pareizāk saukt par *expected value, mathematical expectation,* t.i. **sagaidāmā vērtība, matemātiskā cerība**;

Piemēram, $E(K) = 3{,}5$.

b) no novērotajiem datiem, rēķinot vidējos aritmētiskos – tad tā tiešām ir **vidējā vērtība** (*mean value*).

Piemēram, $\hat{K} \approx \frac{1}{n}\sum_{i=1}^{n} K_i$, ja n eksperimentos ir novērotas lieluma K vērtības K_i.

Nevajag domāt, ka E(K) ir "īstā precīzā" vidējā vērtība, bet $\hat{K}$ – aptuvenā. E(K) ir teorētiskā vērtība modelī, "īstā", varbūt, dabā nemaz neeksistē.

Gadījuma lieluma X vidējo vērtību biežāk apzīmē ar E(X) ("*expected*"), bet ir arī teksti, kur to apzīmē ar M(X) ("mean"). Arī dispersijas apzīmējumi mēdz būt divi – D(X) vai Var(X) ("*variance*").

Teorētiskajā modelī gadījuma lieluma X **sagaidāmo vērtību E(X)** formāli definē kā "svērto summu". Piemēram, diskrētā varbūtību telpā:

$$E(X)=\sum_{e\in E} X(e)\,p(e) .$$

Ja E sastāv no *n* notikumiem e_i, kuru varbūtības ir vienādas ar 1/n, tad:

$$E(X)=\frac{1}{n}\sum_{i=1}^{n} X(e_i) .$$

Šīs definīcijas ideja, protams, ir saistīta ar priekšstatu, ka notikumiem *e* daudzreiz atkārtojoties, vērtību X(e) vidējais aritmētiskais "tieksies" uz E(X). Vēlāk (sk. *lielo skaitļu likumu*) redzēsim, ka šis priekšstats ir teorētiski korekts.

Nepārtrauktā varbūtību telpā E(X) izteiksmē summas vietā nāk integrālis:

$$E(X)=\int_{-\infty}^{+\infty} x\,p(x)\,dx ,$$

kur p(x) ir lieluma X varbūtību blīvuma funkcija.

Teorēma.

$$E(X+Y)=E(X)+E(Y)\ ;$$

$E(aX+b)=aE(X)+b$ jebkuriem skaitļiem a, b.

Bet **ne** vienmēr: $E(X\cdot Y)=E(X)\cdot E(Y)$.

Dispersija

Gadījuma lieluma X **dispersiju** D(X) definē kā

$$E((X-E(X))^2)\ ,$$

jeb **vidējo kvadrātisko novirzi**, un tās uzdevums ir mērīt lieluma X “izkliedi” t.i. "tieksmi" novirzīties no tā vidējās vērtības E(X). Jo lielāka dispersija, jo lielāka varbūtība, ka katrā konkrētā gadījumā X vērtība būs tālu no E(X).

Piemērs − spēļu kauliņa mešana. Kā jau konstatējām, vidējā vērtība E(K)=3,5. Dispersija:

$$D(K)=\frac{1}{6}((1-3{,}5)^2+(2-3{,}5)^2+(3-3{,}5)^2+(4-3{,}5)^2+(5-3{,}5)^2+(6-3{,}5)^2)\ =$$

$$\frac{1}{6}(2{,}5^2+1{,}5^2+0{,}5^2)\cdot 2=\frac{8{,}75}{3}\approx 2{,}92\,.$$

Kāpēc lieluma X "tieksmi" novirzīties no tā sagaidāmās vērtības E(X) nevarētu mērīt ar $E(|X-E(X)|)$, kāpēc jāķeras pie kvadrātiem? Izteiksmei $E(|X-E(X)|)$ nav "skaistu matemātisku īpašību", bet $E((X-E(X))^2)$ tādas ir! Skaistums sākas jau ar šādu teorēmu:

Teorēma.

$$D(X)=E(X^2)-E(X)^2\ ;$$

$$D(aX+b)=a^2D(X)$$

(t.i. skaitlis *b* neietekmē lieluma *aX*+*b* "tieksmi" novirzīties).

Bet **ne** vienmēr: $D(X+Y)=D(X)+D(Y)$.

Kā novirzes mērs, $E((X-E(X))^2)$ nav sliktāks par $E(|X-E(X)|)$, bet tā teorija iznāk skaistāka. Protams, ka ja lielums X ir mērams metros, tad arī E(X) būs metros, bet D(X) būs jau kvadrātmetros.

Un tāpēc līdz ar D(X) izmanto arī $\sqrt{D(X)}$, un to sauc par lieluma X **standartnovirzi**. Tai ir tās pašas mērvienības, kas E(X).

Piemērs − spēļu kauliņa mešana. E(K)=3,5; D(K)≈2,92; standartnovirze: $\sqrt{D(K)} \approx \sqrt{2{,}92} \approx 1{,}71$.

Ja no lieluma X atņem tā sagaidāmo vērtību E(X), tad iegūst lielumu X – E(X) , kura sagaidāmā vērtība ir 0. Šo operāciju sauc par **gadījuma lieluma centrēšanu**.

Ja, tālāk, lielumu X – E(X) izdala ar standartnovirzi $\sqrt{D(X)}$, tad iegūst lielumu $\frac{X-E(X)}{\sqrt{D(x)}}$, kura sagaidāmā vērtība ir 0, bet standartnovirze un dispersija ir 1. Šo operāciju sauc par **gadījuma lieluma centrēšanu un normēšanu**.

Daudzos gadījumos, ja tas neizraisa sajukumu, E(X) apzīmē ar grieķu burtu **μ**, bet standartnovirzi − ar burtu **σ**, tad dispersija būs **σ²**, bet centrētais un normētais gadījuma lielums: $\frac{X-\mu}{\sigma}$.

No **novērotajiem datiem**

dispersiju var aprēķināt, rēķinot vidējos aritmētiskos.

Piemēram, ja *n* eksperimentos ir novērotas lieluma K vērtības K_i, tad vispirms aprēķinām vidējo vērtību $\hat{K} \approx \frac{1}{n}\sum_{i=1}^{n} K_i$, un tad dispersiju:

$$\hat{D} \approx \frac{1}{n-1}\sum_{i=1}^{n}(K_i-\hat{K})^2 .$$

Kāpēc jādala ar $n-1$ (nevis n) – par to vēlāk.

Nevajag domāt, ka D(K) ir “īstā precīzā” vidējā vērtība, bet $\hat{D}$ – aptuvenā. D(K) ir teorētiskā vērtība modelī, “īstā”, varbūt, dabā nemaz neeksistē.

Notikumu un gadījuma lielumu neatkarība

Sk. http://en.wikipedia.org/wiki/Statistical_independence.

Intuitīvi, X un Y neatkarība nozīmē, ka "viens neietekmē otru". Praksē ar tādu priekšstatu bieži vien pietiek, bet teorijā tādu "definīciju" izmantot nav iespējams,.

Tāpēc kā neatkarības definīciju te izmanto (it kā) tās "sekas", t.s. varbūtību reizināšanas likumu.

Notikumu neatkarība

Piemērs – divu spēļu kauliņu mešana. Varbūtība, ka uzkritīs divi sešnieki vai jebkuri divi cipari i un j:

$$P(K_1=i \wedge K_2=j)=\frac{1}{36}; P(K_1=i)\cdot P(K_2=j)=\frac{6}{36}\cdot\frac{6}{36}=\frac{1}{36} \; .$$

Ja

$$P(A \wedge B)=P(A)P(B) \; ,$$

tad notikumus A un B sauc par **neatkarīgiem**.

[Kauliņu mešanai tā ir **teorēma**, bet vispārīgajā gadījumā to izmanto kā neatkarības **definīciju**. Mēdz būt gadījumi, kad $P(A \wedge B)=P(A)P(B)$ konstatē tikai aprēķinu ceļā.

Ar šādu neatkarības definīciju pietiek, lai varētu pierādīt (kā teorēmas) visus ar neatkarību saistītos varbūtību teorijas rezultātus.]

Gadījuma lielumu neatkarība

Ja jebkurām vērtību kopām A un B,

$$P(X \in A \wedge Y \in B)=P(X \in A)P(Y \in B) \; ,$$

tad **gadījuma lielumus X, Y sauc par neatkarīgiem**. Tai skaitā:

$$P(X=a \wedge Y=b)=P(X=a)P(Y=b)$$

Teorēma. Ja X un Y ir neatkarīgi gadījuma lielumi, tad

$$E(XY)=E(X)E(Y)$$ (tā tas ir tikai neatkarīgiem lielumiem!);

$$D(X+Y)=D(X)+D(Y)$$ (tā tas ir tikai neatkarīgiem lielumiem!).

[Tā ir otrā teorēma par dispersijas matemātisko skaistumu.]

No augstam matemātiskām pozīcijām raugoties, varbūtību teorija ir t.s. mēra teorijas (sk. http://en.wikipedia.org/wiki/Measure_theory) speciālgadījums, kur īpaša uzmanība tiek veltīta neatkarības jēdzienam. Šis jēdziens figurē vairumā varbūtību teorijas teorēmu. "Parastajā" mēra teorijā šāda jēdziena nav.

Kovariācija

http://en.wikipedia.org/wiki/Covariance

Gadījuma lielumu X un Y **kovariāciju** ("tieksmi mainīties saskaņoti", atkarību) definē šādi:

$$COV(X,Y)=E((X-E(X))(Y-E(Y))) \ .$$

Korelācijas koeficients (sk. tālāk)

$$\rho(X,Y)=\frac{COV(X,Y)}{\sqrt{D(X)D(Y)}}$$

raksturo lielumu X, Y **lineārās** atkarības pakāpi.

Viegli pārliecinaties, ka

$$COV(X,Y)=E(XY)-E(X)E(Y) \ .$$

Tātad, ja lielumi X, Y ir **neatkarīgi**, tad COV(X, Y)=0. **Bet ne otrādi!**

Ja abu lielumu vidējās vērtības ir 0, tad iznāk vienkārši:

$$COV(X,Y)=E(XY) \ .$$

Tāpat ir ar dispersiju: $D(X)=E((X-E(X))^2)$, bet ja $E(X)=0$, tad $D(X)=E(X^2)$.

No **novērotajiem datiem**

kovariāciju var aprēķināt, rēķinot vidējos aritmētiskos.

Piemēram, ja *n* eksperimentos ir novērotas lielumu K, L vērtības K_i, L_i, tad vispirms aprēķinām vidējās vērtības $\hat{K}\approx\frac{1}{n}\sum_{i=1}^{n} K_i$, $\hat{L}\approx\frac{1}{n}\sum_{i=1}^{n} L_i$ un tad kovariāciju:

$$\hat{COV}(K,L)\approx\frac{1}{n-1}\sum_{i=1}^{n}(K_i-\hat{K})(L_i-\hat{L}) .$$

Kāpēc jādala ar $n-1$ (nevis n) – par to vēlāk.

Kovariāciju matrica

[Ar šo jēdzienu esam jau sastapušies – PCA.]

p gadījuma lielumu $X_1,\ldots,X_p$ **savstarpējās atkarības** kaut kādā mērā raksturo kovariācijas $c_{ij}=COV(X_i,X_j)$. Kopā tās veido *p*x*p* **kovariāciju matricu** $C=(c_{ij})$.

Tā ir **simetriska** matrica: $c_{ij}=c_{ji}$.

Uz diagonāles ir **dispersijas**: $c_{ii}=D(X_i)$.

Ja lielumi X_i ir savstarpēji neatkarīgi, tad C ir **diagonālmatrica** (uz diagonāles – dispersijas, citur – nulles).

Sk. arī http://en.wikipedia.org/wiki/Covariance_matrix

Turpat sk. arī par **korelāciju matricu**.

Lielo skaitļu likums

Sk. http://en.wikipedia.org/wiki/Law_of_large_numbers.

Šis likums teorētiski pamato gadījuma lieluma X sagaidāmās (“teorētiskās vidējās”) vērtības E(X) praktisko nozīmību.

Ja gadījuma lieluma X vērtību "nolasa" *n* neatkarīgos eksperimentos, iegūstot *n* skaitļus $x_1, ..., x_n$, un izrēķina vidējo aritmētisko

$$\frac{x_1+...+x_n}{n},$$

tad "rezultāts būs aptuveni E(X)", un jo lielāks būs *n*, jo tuvāk $\frac{x_1+...+x_n}{n}$ būs E(X). Tāds ir novērojums...

Teorētiskajā modelī šo X vērtību “nolasīšanu” neatkarīgos *n* eksperimentos atspoguļo, ievedot *n* neatkarīgus gadījuma lielumus $X_1, ..., X_n$, kuru varbūtību sadalījumi ir tādi pat kā lielumam X. Tad var jautāt: kāda ir varbūtība, ka vidējais aritmētiskais $\frac{X_1+...+X_n}{n}$ (arī tas ir gadījuma lielums!) atšķirsies no E(X) vairāk par doto mazo skaitli ε:

$$P\left(\left|\frac{X_1+...+X_n}{n}-E(X)\right|>\epsilon\right) \ ?$$

Un, *n* augot, kā šī varbūtība mainīsies? [Kādā **varbūtību telpā** šī varbūtība ir definēta? Lieluma X varbūtību telpas n-tā Dekarta pakāpe plus...]

Teorēma (lielo skaitļu likums). Ja $E(|X|)<\infty$, tad katram mazam ε>0,

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+...+X_n}{n}-E(X)\right|>\epsilon\right)=0 \ .$$

Var pierādīt arī vispārīgāku formulējumu:

Teorēma. Ja neatkarīgiem gadījuma lielumiem

$$X_1, \dots X_n, \dots$$

ir vienāds varbūtību sadalījums ar $E(X_i)=\mu$ un $E(|X_i|)<\infty$, tad katram mazam ε,

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+\dots+X_n}{n}-\mu\right|>\epsilon\right)=0 \ .$$

Piemērs – spēļu kauliņa mešana. Te lielums K pieņem vērtības 1, 2, 3, 4, 5, 6 ar vienādām varbūtībām $\frac{1}{6}$. Tātad E(K) = 3,5. Ja metīsim kauliņu 100 reizes, tad visbiežāk summā sanāks skaitlis, kas tuvs 350 (jeb, vidējais aritmētiskais būs tuvu 3,5). Ja metīsim kauliņu vēl vairāk – 10000 reizes, tad vidējais aritmētiskais parasti būs vēl tuvāk 3,5. Tas ir novērots dabas likums!

Bet cik īsti maza ir varbūtība, ka metot kauliņu 100 reizes, uzkritušo ciparu vidējais aritmētiskais novirzīsies no 3,5 vismaz par 0,1 (t.i. būs zem 3,4 vai virs 3,6)? Un ja metīsim 10000 reizes? Novirzes lielums arī ir gadījuma lielums

$$K'=\frac{K_1+\dots+K_n}{n}-3{,}5 \ ,$$

kam (kā visiem gadījuma lielumiem) ir savs noteikts varbūtību sadalījums. [Kā tas varētu izskatīties? Liekas, blīvuma funkcija varētu būt zvans ar centru 0 – tā prognozē lielo skaitļu likums!]

[Šo intuīciju vajag "kopt"!]

Piemērs. Metīsim kauliņu 100 sērijās par 100 metieniem. Kā tad sadalīsies uzkritušie vidējie aritmētiskie pa intervāliem

... (3,3; 3,4), (3,4; 3,5), (3,5; 3,6) utt.

(histogramma)? Eksperimenti liecina, ka **veidojas zvanam līdzīgs grafiks** ar maksimumu apmēram pie 3,5.

Normālais varbūtību sadalījums un centrālā robežteorēma

Normālais varbūtību sadalījums (*Normal Distribution*, http://en.wikipedia.org/wiki/Normal_distribution) **un centrālā robežteorēma** (*Central Limit Theorem*, http://en.wikipedia.org/wiki/Central_limit_theorem).

Liekas dīvaini, ka par "normālu" sauc (un bieži izmanto kā "default"-o) varbūtību sadalījumu, kura blīvuma funkcija ir

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \text{ (vidējā vērtība 0, dispersija 1).}$$

Kur tāda radusies? Kāpēc tā ir tieši tāda?

Kāpēc jādala ar $\sqrt{2\pi}$? Tāpēc, ka

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi} \cdot$$

Tāpēc ar $\sqrt{2\pi}$ ir jāizdala, lai blīvuma funkcijas integrālis būtu 1.

Bet kā rodas $e^{-\frac{x^2}{2}}$?

Šī funkcija rodas no t.s. **Muavra-Laplasa robežteorēmas.**

Aplūkosim neatkarīgus un vienādi sadalītus gadījuma lielumus

$$X_1, ..., X_n, ... ,$$

kas pieņem vērtības 0, 1, pie tam vērtības 1 varbūtība ir *p* (0<p<1).

Tad:

$$E(X_i) = p, D(X_i) = p(1-p) \;;$$

$$E(X_1+\ldots+X_n)=np\ ;$$
$$D(X_1+\ldots+X_n)=np(1-p)\ ;$$
standartnovirze = $\sqrt{np(1-p)}$.

Robežteorēma. Ja $n\to\infty$, tad centrētās un normētās summas

$$\frac{X_1+\ldots+X_n-np}{\sqrt{np(1-p)}}$$

varbūtību blīvuma funkcija tiecas uz funkciju

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\ .$$

Tas nozīmē, ka šī dīvainā funkcija **izsaka fundamentālu dabas likumu!**

Abraham de Moivre (1667–1754)
Pierre-Simon Laplace (1749–1827)

Piemērs – monētas mešana: $p=\frac{1}{2};\sqrt{np(1-p)}=\frac{\sqrt{n}}{2};$ ja ar C_n apzīmējam *n* metienos uzkritušo ciparu skaitu, tad lieluma $\frac{2}{\sqrt{n}}(C_n-\frac{n}{2})$ varbūtību sadalījuma blīvums tiecas uz $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Un tāpēc varbūtība, ka, piemēram, pie n=10000 izteiksmes $\frac{2}{\sqrt{n}}(C_n-\frac{n}{2})=\frac{1}{50}(C_{10000}-5000)$ absolūtā vērtība nepārsniegs ε=1 (t.i. ka 10000 metienos uzkritušo ciparu skaits būs robežās no 4950 līdz 5050), būs tuvu

$$f(\epsilon)=\frac{1}{\sqrt{2\pi}}\int_{-\epsilon}^{+\epsilon}e^{-\frac{x^2}{2}}dx$$

. Jāsastāda tikai f(ε) tabula (vai arī šīs funkcijas programma ir jāpievieno bibliotēkai)... Sk. tālāk par varbūtību integrāli.

Muavra-Laplasa robežteorēmu var vispārināt lielumiem, kuri pieņem ne tikai vērtības 0, 1. Tādā veidā mēs iegūsim t.s. **centrālo robežteorēmu**:

Centrālā robežteorēma. Ja $X_1, \ldots, X_n, \ldots$ ir jebkuri neatkarīgi gadījuma lielumi, ar vienādu sagaidāmo vērtību μ un dispersiju σ^2, kuriem izpildās t.s. Ļapunova nosacījums, tad, ja $n \to \infty$, tad centrētās un normētās summas

$$\frac{X_1+\ldots+X_n-n\mu}{\sigma\sqrt{n}}$$

varbūtību blīvuma funkcija tiecas uz funkciju

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

Šī teorēma vēl vairāk pastiprina normālā sadalījuma kā fundamentāla dabas likuma reputāciju: ja summējam jebkurus neatkarīgus "faktorus", tad neatkarīgi no faktoru sadalījuma funkcijām, normētā summa "uzvedīsies" kā normāli sadalīts gadījuma lielums!

Normālais sadalījums = “pilnīga haosa” pazīme?
(ja gribam zināt, kas tas ir...)

Varbūtību integrāļa tabulas

Izmantojot normālā sadalījuma funkciju, var izrēķināt atbildi uz jautājumu: kāda ir varbūtība, ka summa $X_1+\ldots+X_n$ **ne**novirzīsies no sagaidāmās vērtības $n\mu$ vairāk par lielumu δ? Ja šo lielumu uzrakstām kā $\delta=\epsilon\sigma\sqrt{n}$ (kur $\sigma\sqrt{n}$ ir summas $X_1+\ldots+X_n$ standartnovirze), tad (lielam n) varbūtība, ka

$$\left|\frac{X_1+\ldots+X_n-n\mu}{\sigma\sqrt{n}}\right|\le\epsilon$$

būs aptuveni

$$f(\epsilon)=\frac{1}{\sqrt{2\pi}}\int_{-\epsilon}^{+\epsilon} e^{-\frac{x^2}{2}}dx \quad .$$

Ja ε=1,960, tad šī izteiksme iznāk 0,95.

Tātad ar varbūtību 95% summa $X_1+\ldots+X_n$ novirzīsies no $n\mu$ ne vairāk par $1{,}960\sigma\sqrt{n}$.

Būtu jāsastāda tikai f(ε) tabula (vai šīs funkcijas programma jāpievieno standartfukciju bibliotēkai)...

Diemžēl, tradīcija ir mazliet savādāka – tabula ir sastādīta citai funkcijai *erf(x)* – t.s. *kļūdu funkcijai*:

$$erf(\epsilon)=\frac{2}{\sqrt{\pi}}\int_{0}^{\epsilon} e^{-x^2}dx \quad .$$

Tad mūs interesējošā varbūtība būs:

$$f(\epsilon)=\frac{1}{\sqrt{2\pi}}\int_{-\epsilon}^{+\epsilon} e^{-\frac{x^2}{2}}dx=\frac{2}{\sqrt{\pi}}\int_{0}^{\epsilon} e^{-\frac{x^2}{2}}dx=\boldsymbol{erf}\left(\frac{\boldsymbol{\epsilon}}{\boldsymbol{\sqrt{2}}}\right)$$

Sk. http://en.wikipedia.org/wiki/Error_function

Piemērs – monētas mešana.

Lielums X_i=1, ja uzkrīt ģerbonis, X_i=0 – citādi.

$$E(X_i)=\frac{1}{2}=\mu\,;D(X_i)=\frac{1}{2}\left(1-\frac{1}{2}\right)=\frac{1}{4}=\sigma^2 \quad ;$$

$$E(X_1+\ldots+X_n)=\frac{n}{2}\,;D(X_1+\ldots+X_n)=\frac{n}{4} \quad .$$

Tātad, ja monētu metīsim n=1000 reizes, tad ar varbūtību 95% summa $X_1+\ldots+X_n$ (t.i. uzkritušo gērboņu skaits) novirzīsies no

$n\mu$ = 500 ne vairāk kā par

$$1{,}960\,\sigma\sqrt{n}=1{,}960\frac{1}{2}\sqrt{1000}\approx 31 \quad .$$

Tātad 1000 metienos ar varbūtību 95% uzkritušo gērboņu skaits būs starp 469 un 531 (t.i. ±3% no 1000).

Attiecīgi, 10000 metienos:

$$1{,}960\,\sigma\sqrt{n}=1{,}960\frac{1}{2}\sqrt{10000}\approx 98 \quad ,$$

t.i. ar varbūtību 95% uzkritušo gērboņu skaits būs starp 4902 un 5098 (t.i. ±1% no 10000).

Vēl daži varianti:

summa $X_1+...+X_n$ novirzīsies no $n\mu$:

ar varbūtību **99%** – ne vairāk par $\mathbf{2{,}576}\,\sigma\sqrt{n}$;
ar varbūtību **99,5**% – ne vairāk par $\mathbf{2{,}807}\,\sigma\sqrt{n}$;
ar varbūtību **99,9%** – ne vairāk par $\mathbf{3{,}291}\,\sigma\sqrt{n}$;
ar varbūtību **99,99999%** – ne vairāk par $\mathbf{5{,}327}\,\sigma\sqrt{n}$.

Raksturīgi, ka varbūtībai strauji tuvojoties 1, koeficients pie $\sigma\sqrt{n}$ aug ļoti lēni.

Ja Jūs interesē citas varbūtības, meklējiet t.s. **Varbūtību integrāļa (jeb kļūdu funkcijas –** ***error function*** $erf(x)$ **) tabulas** (piemēram, http://en.wikipedia.org/wiki/Error_function – sk. lapas beigās, neaizmirstam, ka $x=\frac{\epsilon}{\sqrt{2}}$).

Programmēšanas valodās šo tabulu vietā izmanto iebūvētās funkcijas. Piemēram, MS Excel: ERF(min, max) un ERFC(min, max). Mūs interesē $ERF\left(0,\frac{\epsilon}{\sqrt{2}}\right)$.

Uz datizraces kursu