Teoriniai diskretieji pasiskirstymai

1. Diskretaus teorinio pasiskirstymo supratimas. Bernulio bandymai

Teoriniu vadintinas toks kokio nors dydžio (ar kokio nors požymio reikšmių) pasiskirstymas, kuris nėra „atskleidžiamas“ iš bandymų metu apdorojamos imties duomenų, bet yra *išskaičiuojamas* (sukonstruojamas) pagal žinomus, seniai atskleistus atsitiktinių dydžių dėsningumus. Tad tuo atžvilgiu teoriniai pasiskirstymai priklauso grynai abstraktybių plotmei, bet tikimybiniams bei statistiniams skaičiavimams yra itin reikalingi.

Teoriniai pasiskirstymai, panašiai kaip ir empiriniai, irgi būna dvejopi – diskretieji ir tolydieji – žiūrint kokio tipo atsitiktinius dydžius jie modeliuoja (imituoja).

Diskretusis būtų toks teorinis pasiskirstymas, kuris abstrakčiu lygiu modeliuoja kokio nors diskretaus, dažniausiai – sveikaisiais (o apskritai imant – *baigtiniais*) skaičiais išreiškiamo dydžio (požymio) reikšmių sklaidą. Diskretieji teoriniai pasiskirstymai dažniausiai konstruojami remiantis vadinamaisiais *Bernulio bandymais* (Jakobas **Bernulis** [Bernoulli; 1654–1705] – šveicarų matematikas, tikimybių tyrinėjimo pradininkas), kuriuose svarbu šie momentai:

- Bandymo rezultatas (baigmė) - *binariška*: arba vienokia, arba kitokia, tegiama arba neigima, *x* arba ne-*x*.
- Bandymai *nepriklausomi*, t. y. vieno iš jų reultatas tiesiogiai niekaip nepriklauso nuo kitų bandymų rezultato.
- Teigiamos bandymo baigmės tikimybė ***p*** visuose bandymuose išlieka *pastovi*, nekinta. Jos dydis *žinomas*.
- Neigiamos bandymo baigmės tikimybė irgi išlieka pastovi, ji lygi (***1–p***).

Bernulio bandymų (Bernulio schemos) sąlygas gerai atitinka kapeikos mėtymas, rutuliuko traukimas su grąžinimu iš dvispalvių rutuliukų urnos, berniuko arba mergaitės gimimas ir kiti panašūs realūs ar „dirbtiniai“ atsitiktiniai įvykiai. Tik realių atsitiktinių įvykių atveju gana keblu būna *patikrinti* ir įsitikinti, ar *tikrai* tikimybės visą laiką išlieka pastovios ir ar bandymų rezultatai iš tiesų nepriklauso vieni nuo kitų.

2. Binominis (Bernulio) pasiskirstymas

Tarkime, jog Bernulio bandymas kartojamas tam tikrą (fiksuotą) kiekį kartų: tą kiekį žymėsime *n*. Taip pat žinome, kokia yra tikimybė, jog kiekvienas toks bandymas pasibaigs teigiamai; tos tikimybės dydį pažymėsime *p*. Iš Bernulio bandymų sąlygų aišku, kad kiekvieno bandymo metu ta tikimybė turi išlikti vienoda, nepakitusi. Pavyzdžiui, turime urną su 6 juodais ir 4 baltais rutuliukais. Teigiama bandymo baigme laikysime balto rutuliuko ištraukimą (vadinasi, neigiama – juodo). Tikimybė ištraukti baltą rutuliuką tokiu būdu yra 4/10=0.4, ir jeigu rutuliuką grąžiname atgal į urną, ji tokia pati išlieka ir kitų traukimų metu. Traukimų rezultatai vienas nuo kito irgi nepriklauso. Traukiame rutuliuką, sakysim, dvylika kartų. Tad $p = 0{,}4$; $n = 12$.

Dydžiai *p* ir *n*, kuriuos reikia žinoti ***iš anksto***, yra vadinami šio – binominio arba Bernulio – pasiskirstymo ***parametrais***.

Aišku, kad traukiant rutuliuką 12 kartų, bendras bandymų rezultatas – balto rutuliuko ištraukimų kiekis gali labai įvairuoti, nes jis – atsitiktinis dydis. Gali būti net ir taip, jog balto rutuliuko per 12 traukimų neištrauksime nė karto, gali būti, jog jį ištrauksime 1 kartą, 2 kartus, 3 kartus ir t. t., pagaliau gali atsitikti ir taip, jog visus 12 kartų bus ištrauktas vien tik baltas rutuliukas! Šį mus dominantį atitiktinį dydį – kiek kartų per 12 bandymų ištrauktas baltas rutuliukas – pažymėkime *m*. Iš to, kas sakyta, aikšu, jog *m* teoriškai gali įgauti ***bet kurią*** reikšmę nuo 0 iki 12. Tačiau šansai – nevienodi! Kad per 12 traukimų balto rutuliuko neištrauksime nė karo arba kad jį ištrauksime visus 12 kartų – gana menkai tikėtina.

Binorminio pasiskirstymo tikimybės

Jau seniai yra nustatyta, kaip galima ***teoriškai apskaičiuoti*** kiekvienos iš įmanomų *m* reikšmių tikimybę, sakysim, apskaičiuoti, kokios yra ***teorinės tikimybės***, kad per 12 bandymų balto rutuliuko neištrauksime nė karto, kad jį ištrauksime kaip tik vieną kartą, du kartus ir t. t. iki pat 12 kartų imtinai. Šios tikimybės tiesiogiai priklauso nuo minėtųjų pasiskirstymo parametrų ir yra apskaičiuojamos pagal tokią formulę:

$$P_{(m;n)} = \frac{n!}{m!(n-m)!}\, p^m(1-p)^{n-m}$$

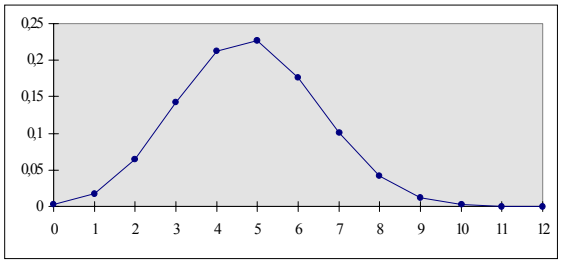
Čia *n* – atliekamų Bernulio bandymų bendras kiekis, *p* – teigiamos pavienio bandymo baigmės tikimybė, *m* – binomiškai pasiskirstęs diskretusis atsitiktinis dydis, galintis įgauti visas reikšmes nuo 0 iki *n*.

Tokiu būdu teoriškai suformuojamas atsitiktinio dydžio *m* (teigiamai pasibaigusio bandymų skaičiaus) *pasiskirstymas*: nustatoma, kokias reikšmes jis gali igauti ir kokia yra kiekvienos iš tų reikšmių tikimybė. Realių bandymų atlikti ne tik kad nereikia, bet jie tiksliai tokių pačių tų tikimybių (tiksliau – santykinių dažnumų!) reikšmių niekuomet ir neduotų, nes būtų jau praktinės, *empirinės* to pasiskirstymo realizacijos, jo „konkretūs atvejai".

Esant anksčiau nurodytiems parametrams (*n*=12, *p*=0,4) dydžio *m* (*teigiamai* pasibaigusių Bernulio bandymų kiekio iš 12 atliktų bandymų) pasiskirstymas būtų toks:

m=	0	1	2	3	4	5	6	7	8	9	10	11	12
P(m|12; 0,4)=	0,0022	0,0174	0,0639	0,1419	0,2128	0,2270	0,1766	0,1009	0,0420	0,0125	0,0025	0,0003	0,0000
F(m)=	0,0022	0,0196	0,0834	0,2253	0,4382	0,6652	0,8418	0,9427	0,9847	0,9972	0,9997	1,0000	1,0000

Pavaizduotas poligonu, tas pats pasiskirstymas atrodytų taip:



1 pav. Binominio pasiskirstymo (*n*=12, *p*=0.40) poligonas

Šiuo atveju sakytume, jog mums rūpintis atsitiktinis dydis *m* turi binominį pasiskirstymą arba yra pasiskirstęs pagal binominį (Bernulio) dėsnį su parametrais *n*=12 ir *p*=0,4. Būtent tie parametrai lemia kiekvienos reikšmės tikėtinumą, jos tikimybės dydį, tad belieka jį tik *apskaičiuoti* pagal nurodytąją formulę.

Dirbant su elektronine skaičiuokle *Excel*, tiesiogiai taikyti minėtosios formulės nereikia, nes čia tam yra skirta speciali Excel funkcija

BINOMDIST(m-reikšmė; n-reikšmė; p-reikšmė; kumuliatyvumas)

Jos parametrai *m-reikšmė*, *n-reikšmė* ir *p-reikšmė* savaime suprantami, o paskutinis, ketvirtasis, parametras *kumuliatyvumas* yra *loginio* tipo: kai jo reikšmė nurodoma FALSE, visa ši funkcija apskaičiuoja *m* atitinkančią tikimybę, o kai TRUE - tą pačią *m* atitinkančią pasiskirstymo funkcijos [$F(x)$] reikšmę.

Pagrindinės charekteristikos

Dar reikia atsiminti, kad *dydžio, pasiskirsčiusio pagal binominį dėsnį su parametrais* ***n*** *ir* ***p*** *pagrindinės charakteristikos yra tokios*:

$m_{vid.} = np$
$s^2 = np(1-p)$,
vadinasi $s = \sqrt{s^2} = \text{sqrt}(np(1-p))$

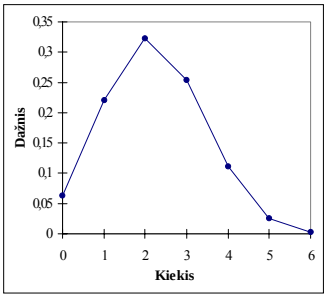
3. „Lingvistinis" binominio pasiskirstymo pritaikymo pavyzdys

Iš palyginti didelio žodžių kiekio (13869 žodžių) yra nustatyta, kad *dviskiemenių* žodžių santykinis dažnumas lietuviškuose tekstuose yra 37,01% (arba 0,3701; žr. *Kalbotyra*, t. 41(1), p. 39). Vadinasi, galima gana patikimai spėti, kad maždaug tokia yra ir tikimybė, kad bet kuris atsitiktinai iš teksto paimtas žodis būsiąs dviskiemenis. Taisyklių, griežtai apsprendžiančių, kad kokio nors ilgio (vadinasi, ir dviskiemenių!) žodžių pavartojimas tekste būtų reglamentuojamas kitų, prieš juos ar po jų einančių žodžių, – irgi nėra, tad galima daryti prielaidą, kad pagrindinės Bernulio bandymų sąlygos (tikimybės stabilumas ir įvykio baigmių nepriklausomumas) šiuo atžvilgiu išlaikomos. Todėl binominiu pasiskirstymu galėtume *teoriškai modeliuoti* labiausiai tikėtiną dviskiemenių žodžių *kiekių pasiskirstymą* kokio nors fiksuoto ilgio teksto atkarpose, sakysim, sakiniuose, sudarytuose iš 6 žodžių (ar bet kokio kito tipo teksto fragmentuose, sudarytuose iš bet kokio kitokio *fiksuoto* žodžių kiekio).

Tarkime, mums rūpi tik sakiniai, turintys po 6 žodžius. Tuomet mus dominančio teorinio dviskiemenių žodžių kiekio pasiskirstymo *parametra*būtų: *n=6* ir *p=0,3701*, o pats pasiskirstymas būtų toks, koks pavaizduotas lentelėje ir 2 pav.

Teorinis dviskiemenių žodžių kiekio pasiskirstymas sakiniuose, sudarytuose iš 6 žodžių
(lentelė)

kiekis (m)=	0	1	2	3	4	5	6
P(m|6; 0,3701)=	0,0625	0,2202	0,3235	0,2534	0,1117	0,0262	0,0026
F(m)=	0,0625	0,2827	0,6061	0,8595	0,9712	0,9974	1,0000



2 pav. Teorinis dviskiemenių žodžių kiekio pasiskirstymas sakiniuose, sudarytuose iš 6 žodžių (poligonas)

Pagrindinės šio pasiskirstymo charakteristikos tada būtų:

$m_{\text{vid.}} = np = 6 \text{ x } 0{,}3701 = \mathbf{2{,}2206}$
$s = \text{sqrt}(np(1–p)) = \text{sqrt}\,(2{,}2206 \text{ x } 0{,}6299) = \mathbf{1{,}1827}$

Vadinasi, tarsi gauname pakankamai solidų atramos tašką: remdamiesi tam tikromis išankstinėmis prielaidomis apskaičiuojame, kokios yra tikimybės, kad sakinyje, sudarytame iš 6 žodžių, dviskiemenio žodžio nebus nė vieno, kad bus tik vienas, bus jų du ir t. t.; kartu apskaičiuojame, koks yra dviskiemenių žodžių kiekio *vidurkis* šitokio ilgio sakiniuose ir koks – standartinis (vidutinis kvadratinis) nuokrypis. Su tokiu teoriniu pasiskirstymo modeliu jau būtų racionalu lyginti ir konkrečius empirinius duomenis. Tarkime, jog lingvistas sužiūri visus šešiažodžius kokio nors stambaus tiriamojo teksto sakinius ir nustato, keliuose iš jų dviskiemenių žodžių apskritai nėra, keliuose – yra po vieną, po du, po tris ir t.t. dviskiemenius žodžius, o po to mėgina atsakyti į klausimą: ar *empirinis* dviskiemenių žodžių kiekio pasiskirstymas tokiuose sakiniuose *atitinka* analogiško „teorinio“ atsitiktinio dydžio (kurį apskaičiuotasis teorinis pasiskirstymas ir „reprezentuoja“) pasiskirstymą tokiomis pačiomis sąlygomis, ar ne. Lyginimo „instrumentas“ yra specialūs *statistiniai kriterijai* (apie juos – vėliau), o vienas iš *lyginamųjų objektų*, pasakytum, (pa)lyginimo „predikatas“ (tai, ***su kuo*** empirinis pasiskirstymas lyginamas) ir yra aptartuoju būdu randamas teorinis pasiskirstymas.

Atkreipkime dėmesį!

Pasinaudojant „tikimybių algebra“ iš teorinio paiskirstymo duomenų tiesiog *sumavimo būdu* lengvai galima apskaičiuoti bet kurių mus dominančių bandymo ***baigčių*** (mūsų atveju – dviskiemenių žodžių kiekių šešiažodžiuose sakiniuose) ***derinių*** tikimybes. Pvz., tikimybė, kad šešiažodžiame sakinyje dviskiemenių žodžių bus:

* *ne daugiau kaip 3* – yra lygi 0,0625+0,2202+0,3255+0,2534 = **0,8596**
* *nuo 2 iki 4* (*imtinai*) – yra lygi 0,3235+0,2534+0,1117 = **0,6886**
* *ne daugiau kaip 1 arba 4 ir daugiau* – yra lygi 0,0625+0,2202+0,1117+ +0,0262+0,0026=**0,4232**

Panašiai apskaičiuotume ir visų kitų derinių teorines tikimybes.

Bet tais atvejais, kai rūpi rasti *ištisinį m* reikšmių *intervalą* atitinkančių tikimybių sumą, galima pasinaudoti ir *pasiskirstymo funkcija*. Jei to intervalo apatinę ribą pažymėsime *a*, o viršutinę *v*, tai tikimybė, jog binomiškai pasiskirstęs dydis *m* įgis bet kurią reikšmę nuo *a* iki *v* imtinai, yra:

$$P_{(a<=m<=v)} = F(v)–\text{F}(a–1)$$

Sakysim, norėdami iš pateiktojo pavyzdžio nustatyti, kokia yra tikimybė, jog šešiažodžiuose sakiniuose dviskiemenių žodžių rasime nuo 3 iki 5 imtinai, galėtume pasinaudoti kaip tik šia „paprastesne“ formule:

$$P_{(3<=m<=5)} = F(5)–\text{F}(3–1) = F(5)–F(2) = 0{,}9974–0{,}6061=0{,}3913$$

Savaime aišku, jog tą pačią reikšmę gautume ir susumavę atitinkamas tikimybes:

$$P_{(3<=m<=5)} = P(3)+P(4)+P(5) = 0{,}2534+0{,}1117+0{,}0262 = 0{,}3913$$

4. Binominio pasiskirstymo apibendrinimas dideliam bandymų kiekiui (Muavro–Laplaso teorema)

Kadangi tikimybių apskaičiavimo binominio pasiskirstymo atvejui formulėje figūruoja dydžių *n* ir *m*, o taip pat ir jų skirtumo *faktorialai* [*n*!, *m*! ir (*n*-*m*)!], tai sunkumų atsiranda tada, kai šie dydžiai tampa pakankamai dideli (net dalis kalkuliatorių neleidžia apskaičiuoti skaičių, didesnių už 63, faktorialo!). O praktiniame darbe neretai tenka modeliuoti situacijas ir su dideliais, kur kas didesniais nei 63 bandymų kiekiais (kitaip tariant, *n* būna palyginti didelis skaičius, šimtai ar net tūkstančiai, o *m* – igyja reikšmes iš intervalo [*0*, *n*]). Binominių tikimybių apskaičiavimo uždavinį tokiems atvejams yra išsprendę, rodos, nepriklausomai vienas nuo kito, Abrahamas **Muavras** (Moivre, 1667–1754) ir Pjeras Simonas **Laplasas** (Laplace, 1749–1827). Muavras rado būdą apskaičiuoti visoms *m* reikšmių tikimybėms, kai *p*= 0,5, o Laplasas – kai *p* reikšmė

yra bet kokia, todėl tas būdas tikimybių teorijoje ir matematinėje statistikoje dažniausiai ir siejamas su abiejų šių matematikų vardais.

Vėl – lingvistinis pavyzdys

Sakysim, yra žinoma, jog tam tikro žanro prozos tekstuose *ilgų* (ilgesnių negu 17 žodžių) sakinių dažnumas (vadinasi, gana patikimai galima spėti, jog – ir tikimybė *p*) yra **0,208** (plg. *Kalbotyra*, t. 43(1), p.23). Rūpi apskaičiuoti, kokios būtų teorinės tikimybės, kad palyginti dideliuose tekstuose (tarkime, kuriuose iš viso yra po 850 sakinių) „ilgų“ sakinių nebus nė vieno, bus vienas, du, trys ir t. t., pagaliau – kad jų bus 848, 849 ir galiausiai – visi 850 sakinių bus vien tiktai „ilgi“ (t.y. turės ne mažiau kaip po 18 žodžių). Galima daryti šiaipjau pakankamai grubią prielaidą, kad visuose tuose tekstų fragmentuose, turinčiuose po 850 sakinių, „ilgų“ sakinių tikimybės nesikeičia, visur išlieka vienodos, nors „tikro“ mokslinio tyrimo interesais vis dėlto reikėtų tą prielaidą kaip nors patikrinti ar bent jau pagrįsti. Paprasčiau yra su antrąja Bernulio bandymų sąlyga: gana akivaizdu, kad vieno sakinio ilgis tiesiogiai niekaip neįtakoja kitų – nei prieš jį, nei po jo einančių – sakinių ilgio. Tačiau skaičiuodami be kompiurerio, anksčiau pateikta binominio pasiskirstymo formule praktiškai negalėtume šiuo atveju pasinaudoti, nes skaičiavimai pagal ją pasidarytų neišpasakytai griozdiški.

Muavro ir Laplaso yra įrodyta, jog kai *n* pakankamai didelis, *m* reikšmes atitinkančios tikimybės yra:

$$P_{(m,\,n)} \approx \frac{1}{\sqrt{np(1-p)}} \cdot \frac{1}{\sqrt{2\pi}}\, e^{-\frac{\left(\frac{m-np}{\sqrt{np(1-p)}}\right)^2}{2}}$$

Ši formulė „baisi“ tik iš pirmo žvilgsnio: iš tikrųjų ji remiasi jau žinomais kintamaisiais *n*, *p* ir *m* bei pora konstantų. Viena iš jų yra visiems žinoma konstanta $\pi \approx 3{,}14159...$, o kita iš jų, konstanta *e*, vadinama *Neperio skaičiumi* ir yra apytikriai lygi **2, 71828...** (ši konstanta matematikoje irgi dažna, ji yra *natūrinių logaritmų* pagrindas). Jeigu dėl paprastumo pažymėtume, pvz, jog ***m–np / sqrt(np(1–p))= x***, tai ta pati formulė supaprastėtų ir atrodytų taip:

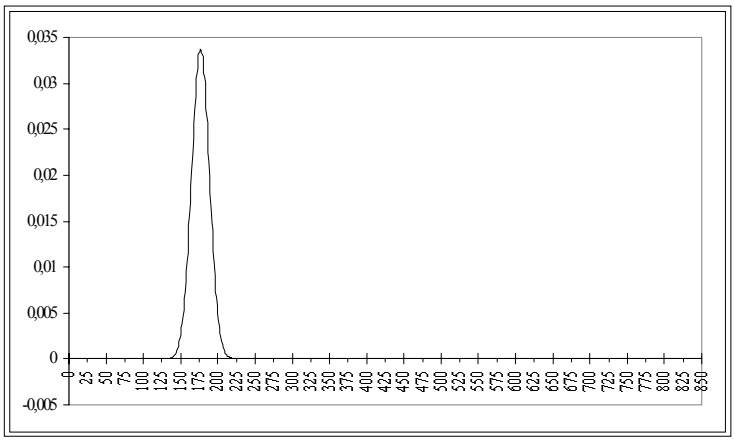
$$P_{(m,\,n)} \approx \frac{1}{\sqrt{np(1-p)}} \cdot \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

Antrasis dauginamasis šioje išraiškoje yra tik dydžio *x* (kuris priklauso nuo tų pačių *n*, *p* ir *m*) *funkcija*, ir tos funkcijos reikšmes matematinės statistikos vadovėliai dažnai pateikia specialiomis lentelėmis, tad belieka jas ten tik susirasti, o su kompiuteriu – nėra jokių problemų ir paskaičiuoti jas „tiesiogiai“. Jeigu tos funkcijos reikšmę pažymėtume *f(x)*, tai „baisioji“ formulė taptų ir visai „paprasta“:

$$P_{(m,\,n)} \approx f(x) \,/\, \mathrm{sqrt}(np(1-p))$$

Jei šias formules pritaikytume anksčiau pateiktam pavyzdžiui – „ilgų“ sakinių kiekių tikimybėms tekste iš 850 sakinių – tai ištisinė, neintervaluota pasiskirstymo lentelė užimtų net keliolika puslapių, bet „ilgų“ sakinių *pavienių* kiekių tikimybės būtų palyginti labai mažos. Tai ir natūralu: kadangi „ilgų“ sakinių kiekis tekste iš 850 sakinių bent jau teoriškai gali įgauti *bet kurią* reikšmę nuo 0 iki 850, tai *pilnąją* elementariųjų įvykių erdvę sudaro 851 įvykis, ir visų jų tikimybių suma lygi vienetui (vienoks koks nors „ilgų“ sakinių kiekis būtinai bus rastas). Kitaip tariant, tą vienetą dabar tenka „išdalinti“ net 851 elementariajam įvykiui, ir išdalinti ne lygiomis dalimis, o tokiomis proporcijomis, kokios apskaičiuojamos pagal pateiktąsias formules. Kadangi pasiskirstymo lentelė didelė, čia pateikiamas tiktai jo grafikas (poligonas). Net ir iš jo matyti, kad bent kiek žymiau besiskiriančias nuo nulio reikšmes tikimybė $P_{(m,\,n)}$ pradeda įgauti tik tada, kai *m* priartėja prie 140, ir vėl nustoja – kai *m* viršija 220. Didžiausią reikšmę (0,033709) ši tikimybė įgauna, kai *m*=177. Pačioje pasiskirstymo lentelės pradžioje, kai m=0, ši tikimybė yra neapsakomai maža, ir kompiuteris ją užrašo vadinamuoju „inžineriniu“ formatu: 1,13E-50. „Normaliai“ gi užrašytas šis skaičius atrodytų taip:

0,00000000000000000000000000000000000000000000000000000113.



3 pav. „Ilgų“ sakinių pasiskirstymas tekstų fragmentuose po 850 sakinių (poligonas)

O kai *m* pasiekia 621, net ir šiuolaikinis kompiuteris gaunamą tikimybės reikšmę sutapatina su vadinamuoju mašininiu nuliu, kitaip sakant – nebeskiria nuo nulio („normaliu“ formatu rašant tokius skaičius po kablelio tektų rašyti gerokai daugiau kaip 300 nulių...). Atkreiptinas dėmesys į du specifinius šio pavyzdžio niuansus.

Pirma, didėjant bandymų kiekiui (mūsų atveju – tekstų apimčiai), o tuo pačiu – ir ilgėjant galimų *m* reikšmių intervalui diskretusis pasiskirstymas vis labiau darosi panašus į tolydųjį: žingsnelių (grafike – tiesiųjų laužtinės atkarpėlių) daugėja, ir bendra poligono forma glotnėja, tolydėja. Kuo jų daugiau, tuo daugiau ir tolydumo, tuo mažiau toksai diskretusis pasiskirstymas besisikiria nuo tolydžiojo.

Antra, kuo daugiau bandymų ir teoriškai galimų *m* reikšmių, tuo į daugiau dalių, tegu ir nelygių, tenka „išdalinti“ būtinojo įvykio tikimybės reikšmę – vienetą. Vadinasi, tuo pačiu atitinkamai mažėja atskiroms *m* reikšmėms tenkančių tikimybių $P_{(m,n)}$ dydis (vertė), o kai bandymų kiekis (ir – galinė *m* reikšmių intervalo riba) artėja į begalybę, tų $P_{(m,n)}$ dydis artėja į 0. Tai – iš pirmo žvilgsnio keistas, bet protu gana lengvai suvokiamas dalykas, su kuriuo aktualiai susidursime aptardami „tikruosius“ tolydžiuosius pasiskirstymus.

Pasigilinus net ir į čia pateiktą pavyzdį nesunku suprasti, jog praktiškai retai kuomet reikia ar tikslinga būna teoriškai modeliuoti visą (pilną) pasiskirstymą, ypač – tuomet, kai bandymų skaičius didelis ir – tuo pačiu – atskirų tikimybių $P_{(m,n)}$ reikšmės mažos. Bene dažniau prisieina apskaičiuoti „sudėtines“ tikimybes, atitinkančias situaciją, kai *m* papuola į vienokį ar kitokį dėl kokių nors priežasčių „įtartiną“ reikšmių *intervalą*. Grįžtant prie pavyzdžio, nesunku būtų, sakysim, įsivaizduoti, jog dėl kokių nors priežasčių parūpsta, kokia yra teorinė tikimybė, kad „ilgų“ sakinių kiekis 850 sakinių tekste bus **nuo 100 iki 150**. Visi kiti galimi „ilgų“ sakinių kiekiai šiuo atveju tegu nerūpi. Savaime suprantama, jog praktiniams reikalams tokias „sudėtines“ tikimybes nėra keblu su kompiuteriu apskaičiuoti ir *tiesioginio sumavimo* būdu, tačiau daug logiškiau būtų, sakysim, dirbant su *Excel*'iu, pasinaudoti kaip tik pasiskirstymo funkcija: ją pakaktų apskaičiuoti tik dviems *m* reikšmėms: 150 ir 99. Tad:

$$P_{(100<=m<=150)} = F(150)–F(100-1) = F(150)–F(99) = 0{,}0120–1{,}55E\text{-}12 \approx 0,\ 0120$$