

Univerzita Karlova v Praze  
Matematicko-fyzikální fakulta

**BAKALÁŘSKÁ PRÁCE**



Ondřej Šedivý

**Aplikace geometrické statistiky na měření charakteristik rovinných objektů**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Ivan Saxl, DrSc.

Studijní program: Obecná matematika

2006

Moje vřelé a upřímné poděkování patří RNDr. Ivanu Saxlovi, DrSc. za ochotu a trpělivost při vedení mé bakalářské práce.

Prohlašuji, že jsem svou bakalářskou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 24. 5. 2006

Ondřej Šedivý

# Obsah

<b>1. Pravděpodobnost a statistika.....</b>	<b>5</b>
1. 1.    Základní pojmy a principy.....	5
1. 2.    Klasická versus geometrická statistika.....	5
<b>2. Geometrické objekty.....</b>	<b>7</b>
2. 1.    Zanedbávání rozměrů .....	7
2. 2.    Geometrický výběr .....	8
<b>3. Aplikace geometrické statistiky.....</b>	<b>10</b>
3. 1.    Bodové systémy .....	10
3. 2.    Dvourozměrné částice v rovině.....	13
3. 3.    Buffonovy úlohy .....	15
<b>4. Měření charakteristik rovinných objektů pomocí geometrické statistiky....</b>	<b>21</b>
4. 1.    Měření obsahu rovinného obrazce .....	21
4. 2.    Měření délky křivky .....	23
<b>5. Zkoumání trojrozměrných objektů .....</b>	<b>25</b>
<b>6. Dodatek .....</b>	<b>26</b>
<b>Literatura.....</b>	<b>27</b>

**Název práce:** Aplikace geometrické statistiky na měření charakteristik rovinných objektů

**Autor:** Ondřej Šedivý

**Katedra:** Katedra pravděpodobnosti a matematické statistiky

**Vedoucí bakalářské práce:** RNDr. Ivan Saxl, DrSc.

**e-mail vedoucího:** saxl@math.cas.cz

**Abstrakt:**

Cílem práce je podat stručný a co nejpřístupněji napsaný přehled odhadů charakteristik rovinných objektů z výběrů pořízených geometrickými prostředky. První dvě kapitoly jsou obecnějšího rázu. Pojednávají o rozdílech mezi klasickým a geometrickým výběrem z populace a naznačují některé základní principy dále využívané. Následují ukázky použití geometrické statistiky k odhadům základních charakteristik pro rovinné soubory bodů, křivek a ploch. Krátká závěrečná stať je věnována způsobu zkoumání trojrozměrných objektů. V textu jsou také podrobně zpracovány historické Buffonovy úlohy o jehle a čtverci.

**Klíčová slova:** geometrická statistika, klasický a geometrický výběr, Buffonovy úlohy

**Title:** Application of geometrical statistics for measuring characteristics of planar objects

**Author:** Ondřej Šedivý

**Department:** Department of Probability and Mathematical Statistics

**Supervisor:** RNDr. Ivan Saxl, DrSc.

**Supervisor's e-mail address:** saxl@math.cas.cz

**Abstract:**

The aim of this work is to offer a short and clearly written review of estimations of characteristics of planar objects obtained by geometrical sampling. The first two chapters are more general. They are devoted to the differences between the classical and geometrical sampling of populations and suggest some basic principles used in what follows. In the next two chapters, the methods of estimating geometrical characteristics of planar populations of points, curves and figures are covered. A short closing chapter sketches the approaches to the 3D populations of geometrical objects. The historical Buffon needle and square problems are also explained in detail.

**Keywords:** geometrical statistics, classical and geometrical sampling, Buffon problems

# 1. Pravděpodobnost a statistika

## 1. 1. Základní pojmy a principy

Se statistikou se všichni setkáváme v běžném životě velice často. Kdybychom chtěli vymezit, co je jejím obecným cílem, dalo by se říci, že je to nashromáždění kvantitativních informací o určitém jevu, jejich analýza a následné vyhodnocení umožňující popsání zkoumaného jevu. Jak takový postup vypadá v praxi?

Jedním typem jsou situace, kdy máme veškeré informace o celém jevu, ty zpracujeme a vyhodnotíme. Například máme veškeré podrobnosti o jednotlivých utkáních v hokejové lize a na jejich základě zjistíme nejlepšího střelce, nejtrestanějšího hráče apod. Mne však budou spíše zajímat situace, ve kterých máme pouze omezené informace, z nichž pouze odhadujeme charakteristiky celého jevu.

Sběr těchto informací (dat) se většinou uskutečňuje tím způsobem, že se ze zkoumaného souboru vybere jistá jeho část, běžně označovaná jako *výběrový soubor*, a na ní se zjistí vlastnost, která je předmětem našeho zájmu. Získaný výsledek se pak vhodným způsobem vztáhne na celý soubor. Čím větší výběrový soubor máme, tím přesnější představu o celku teoreticky dostáváme. Tak například při předvolebním průzkumu položíme náhodné skupině lidí otázku, jakou budou volit politickou stranu. Vyhodnocením ankety následně získáme výsledek, který budeme považovat za odhad rozložení voličů v celé populaci.

V souvislosti s názvem této kapitoly se naskytá otázka, jaký je vztah mezi statistikou a pravděpodobností? Pravdou je, že tyto dvě disciplíny jsou si v matematice velmi blízké. Mnohdy se jedná pouze o interpretaci téhož jevu dvěma způsoby. U předvolebního průzkumu jsme získali určitou statistiku, v médiích se výsledek zobrazuje nejčastěji pomocí známého „koláče“. Nyní si můžeme položit následující otázku: jaká je na základě získaných dat pravděpodobnost, že náhodně oslovená osoba volí jednu konkrétní politickou stranu? Odpověď zní, že jejím odhadem na základě našeho výběrového souboru je právě procentuelní zastoupení voličů této strany v našem průzkumu.

## 1. 2. Klasická versus geometrická statistika

Pojmy pravděpodobnost a statistika jsou intuitivně srozumitelné. Nyní však docházíme na důležité rozcestí. Jak je psáno výše, v obecném schématu se z nějakého malého vzorku odhaduje vlastnost celého souboru. Dosud jsme však uvažovali jen případy, kdy zkoumaný soubor má nějaký konkrétní (konečný) počet prvků, z nich pak jednoduše některé vybereme. Z více než deseti milionů obyvatel ČR se třeba tisíci lidem položí otázka, kterou politickou stranu budou volit. Způsobů, jak náhodně vybrat tisíc lidí, je sice mnoho, ale určitě je to konečný počet<sup>1</sup>. A právě takovými konečnými výběry z konečné populace se zabývá *klasická statistika*.

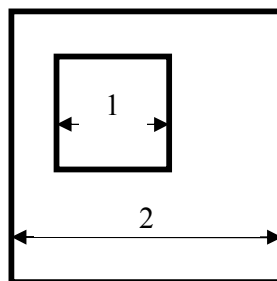
---

<sup>1</sup> V reálné situaci je to  $\binom{n}{1000}$ , kde  $n$  je počet obyvatel ČR, kterým má smysl takovou otázku položit.

Představme si nyní následující situaci: hospodyně peče v troubě koláč a chce zjistit, jestli je již všude dostatečně propečený. Zkusí do něj proto na několika místech píchnout špejlí. Její počínání lze opět chápat jako výběr malého statistického vzorku, z něž se pak odhadne charakteristika celého objektu. Kuchařka však v této situaci nemá vyčíslitelný počet možností, jak to udělat, ona si může vybrat naprosto libovolné body a těch je nekonečně (dokonce nespočetně) mnoho. Ve skutečnosti svůj výběr zvolí šikovně, aby zjistila propečenost koláče uprostřed i na jeho okraji, v přední i zadní části trouby. A samozřejmě čím více píchnutí provede, tím jistější si může být ve svém úsudku o propečenosti celého koláče.

V předchozím odstavci jsem popsal realizaci tzv. *geometrického výběru*, s jehož obecným pojetím čtenáře seznámím v kapitole nazvané Geometrický výběr (2.2.). *Geometrická statistika* se zabývá zkoumáním objektů v prostoru právě realizací nějakého geometrického výběru.

Podobné je to s pravděpodobností. Jestliže chceme spočítat, jaká je šance, že nám při hodu obyčejnou kostkou padne sudé číslo, vydělíme počet příznivých možností (tj. 3) počtem všech možností (tj. 6) a získáme poměr  $\frac{1}{2}$ . V takovém případě se jedná o *klasickou pravděpodobnost*. Jiná situace nastane, budeme-li se ptát, jaká je šance, že náhodně zvolený bod ve čtverci o straně 2 bude i bodem čtverce o straně 1, který je ve větším čtverci celý obsažen:



Obr. 1 – Ilustrace geometrické pravděpodobnosti

Tady již nemůžeme uvažovat počet příznivých možností a počet všech možností. Zvolit bod ve čtverci lze totiž nekonečně mnoha způsoby. Na tomto místě přichází ke slovu *geometrická pravděpodobnost*. Správný výsledek dostaneme jako poměr obsahů malého a velkého čtverce, neboli podíl obsahu oblasti znázorňující příznivé možnosti ku obsahu celé oblasti:

$$p = \frac{1^2}{2^2} = \frac{1}{4} .$$

Na tomto místě je užitečné si uvědomit, že svět kolem nás je doslova přeplněn náhodnými jevy, jejichž realizacemi jsou různé geometrické objekty, které nás obklopují. Příkladem je každý plod jabloně nebo květ tulipánu. V krajině se střídají pohoří s nížinami, řeky si najdou postupem času svoje řečiště, také vývoj světadílů do dnešní podoby byl velmi spletitý. Stejně tak můžeme přejít do mikrosvěta a podívat se na tvar buněk, mikroorganismů, atd.

Zkrátka svět se v podstatě skládá ze samých geometrických objektů, které spolu interagují a vytvářejí veškeré krásy přírody. Atomy se shlukují do molekul,

molekuly zase do větších celků a nakonec i my sami jsme výsledek řetězce složitých a dlouhodobých náhodných jevů.

Geometrickou statistiku lze tedy chápat i jako nástroj, kterým usilujeme o popsání a pochopení okolního světa. Je to prostředek, který umožňuje studovat geometrické objekty, jejich tvar, velikost, vnitřní strukturu a další vlastnosti.

## 2. Geometrické objekty

### 2. 1. Zanedbávání rozměrů

Veškeré objekty kolem nás jsou trojrozměrné. V některých situacích však můžeme, nebo dokonce musíme jeden, nebo i více rozměrů při jejich zkoumání zanedbat a tím si zjednodušit situaci. Co ve skutečnosti zanedbávání rozměrů obnáší? Začneme od samé podstaty věci.

Každý objekt zaujímá v prostoru nějaký objem, rozměry objektu se měří ve třech směrech, obvykle na sebe kolmých. Kdyby byl objekt například dvourozměrný, znamenalo by to, že jeden z rozměrů objektu je nulový. To v praxi není možné, ale při jisté idealizaci, která je v matematice běžná, můžeme některé objekty nebo jejich části za dvourozměrné považovat.

Vezměme si třeba list papíru. Jeho tloušťka je oproti zbylým dvěma rozměrům velmi malá, přesto může v některých situacích hrát rozhodující roli – např. budeme-li zjišťovat, jestli se štos určitého počtu papírů vejde do skříňky dané výšky. Na druhou stranu existuje mnoho situací, ve kterých je tloušťka papíru zanedbatelná. Nakreslíme-li na papír obrázek, pak je pro nás tloušťka papíru při prohlížení tohoto výtvaru naprosto bezvýznamná. Ve skutečnosti je to tak, že využíváme pouze vrchní stranu papíru, v geometrickém smyslu vrchní stěnu kváдру.

Při kreslení se tedy dá při troše idealizace říci, že se hrot tužky pohybuje ve dvourozměrném prostoru a vytváří v něm nový objekt – obrázek. Kolik má rozměrů? I tady záleží na tom, jaký obrázek vytváříme a co přesně nás zajímá. Budeme-li kreslit (resp. malovat) krajinku, vytvoříme na papíře jakýsi celek a bude se jednat o dvourozměrný objekt. Avšak jestliže nakreslíme nějakou rovnou nebo klikatou čáru a bude nás zajímat pouze její délka, nebude hrát šířka čáry vůbec žádnou roli, tento rozměr zanedbáme a vzniklý objekt budeme považovat za jednorozměrný – jediný rozměr čáry bude právě její délka. Nakonec nakreslíme-li pouze jednotlivé body a bude nás zajímat pouze jejich rozmístění na papíře, dá se říci, že jsou to bezrozměrné („nularozměrné“) útvary. Nezajímá nás žádný jejich rozměr, nýbrž pouze jejich počet a poloha.

Podívejme se nyní na věc z obecného hlediska. Při zkoumání objektu je nutné si nejprve rozmyslet, v jakém prostoru ho budeme uvažovat (v trojrozměrném prostoru, v rovině, nebo např. na přímce), a dále jaké rozměry budeme u objektu zanedbávat. Rozhodovací proces má přitom brát v úvahu kromě povahy situace zejména to, jaké vlastnosti (rozměry) objektu zkoumáme.

Uvedu ještě jeden příklad ukazující, jak je případné zanedbávání rozměrů závislé na tom, co přesně chceme u daného jevu zjistit. Představme si, že chceme vyrobit mapu nějaké oblasti znázorňující silniční síť. Nyní si můžeme klást různé požadavky, jimž bude odpovídat uvažované zanedbávání rozměrů u řek a silnic:

- i) má-li být z mapy čitelné, jak se lze po silnicích dostat z jednoho místa na druhé, přičemž všechny silnice jsou pro nás průjezdné, pak nezáleží na šířce čar znázorňujících silnice a můžeme je považovat za jednorozměrné,
- ii) budeme-li rozlišovat různé druhy silnic podle šířky, pak vstoupí do hry druhý rozměr,
- iii) kdybychom čistě teoreticky navíc chtěli mít informaci i o výšce použitého silničního materiálu, museli bychom vyrobit trojrozměrný model nebo třetí rozměr nahradit například různými barevnými odstíny.

## 2. 2. Geometrický výběr

Jak jsem již dříve naznačil, informace o zkoumaném objektu se získávají pomocí tzv. geometrického výběru. Ten spočívá v tom, že se vhodně zvolí nějaké podprostory prostoru, ve kterém objekt původně zkoumáme, a sice většinou podprostory nižší dimenze (tj. body nebo přímky v rovině, roviny v trojrozměrném prostoru apod.), načež se vyšetřovaná vlastnost zjišťuje pouze na jejich průnicích s daným objektem. Opět se zde tedy naplňuje schéma, kdy se vybere statistický vzorek, resp. výběrový soubor, a z něj se usuzuje na vlastnosti celého zkoumaného souboru, v tomto případě geometrického objektu v prostoru.

Není-li předchozí odstavec pro čtenáře dostatečně srozumitelný, snáze si udělá názornou představu o podstatě geometrického výběru z některých jeho konkrétních realizací:

O jedné takové již byla řeč u modelové situace týkající se hospodyně, která zjišťuje propečenost koláče a za tím účelem do něj píchne špejlí na několika náhodně vybraných místech. Pokud nás zajímá pouze poloha bodů, kde tak učiní, přičemž koláč (resp. jeho horní stěnu) považujeme za rovinný útvar (jeden rozměr – výšku – zanedbáváme), lze její počínání nazvat jako geometrický výběr na dvourozměrném objektu prováděný pomocí bodů<sup>2</sup>.

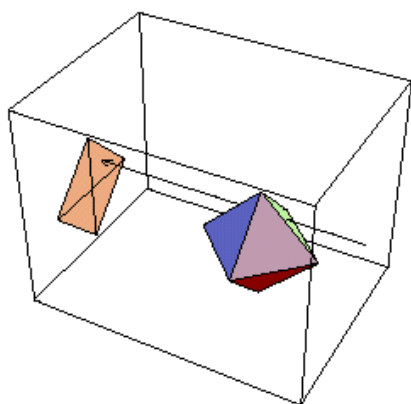
Pro ilustraci geometrického výběru rovinou může posloužit tento příklad: zajímá-li nás rozložení plísňových oblastí v sýru, naskytá se možnost ho rozříznout a podívat se, jak to vypadá na provedeném řezu. Stejně můžeme zjišťovat rozložení dutinek v chlebu nebo děr v ementálu apod.

Zmíním se dále ještě o jedné hojně používané metodě geometrického výběru a tou je *projekce*. Na obrázku 2a (převzato z [7]) je těleso, na které je zaměřen paprsek. Kdybychom ozářili těleso celým svazkem rovnoběžných paprsků, na zadní stěně by byl patrný stín, jako je to naznačeno na obrázku. V řeči matematiky: provedli bychom projekci tělesa pomocí svazku rovnoběžných přímek na rovinu k nim kolmou. Na obrázku 2b je zachycena projekce svazkem rovnoběžných přímek v rovině.

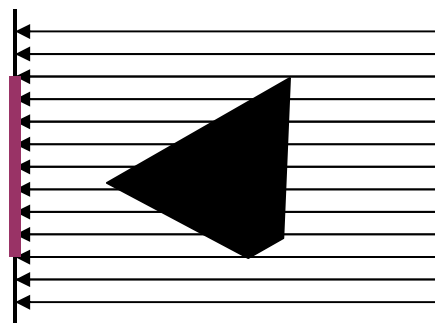
---

<sup>2</sup> Ve skutečnosti je tento proces složitější, protože propečenost koláče se určuje na základě způsobu pronikání špejle do těsta, případně zjišťováním, jestli na špejli po vyjmutí zůstaly zbytky těsta. Při vši důslednosti se tedy jedná o zkoumání koláče jakožto trojrozměrného objektu povrchem špejle, tedy dvourozměrným válcovým pláštěm.





Obr. 2a – Projekce v prostoru



Obr. 2b – Projekce v rovině

Měřením délky úsečky, která je výsledkem projekce na obr. 2b, při různých polohách promítané částice bychom třeba mohli odhadnout střední šířku částice<sup>3</sup>. Význam projekce je tedy zjednodušeně řečeno ten, že zobrazí zkoumané těleso nebo jeho část do prostoru nižší dimenze, kde se pak snáze vyšetřují některé vlastnosti objektu.

Z obecného hlediska jsou pravidla, jak provádět geometrický výběr, dost složitá a obtížně definovatelná. Zkoumané objekty totiž mohou být značně nesourodé a těžko popsatelné a existuje přitom nekonečné množství možností, jak geometrický výběr pořídit. Proto je také třeba dávat velký pozor jak na volbu výběru, tak na správnou interpretaci získaných výsledků.

Připomeňme ještě jednu důležitou vlastnost. Při výběru z většího množství objektů jsou přednostně (tj. s větší pravděpodobností) vybírány větší objekty. V práci se ovšem zabýváme především jednotlivými objekty a bodovými soubory, a proto tuto vlastnost, významnou především u zkoumání objektů v trojrozměrném prostoru, nebudeme dále rozebírat. Svou roli však sehraje zejména v kap. 3.2.

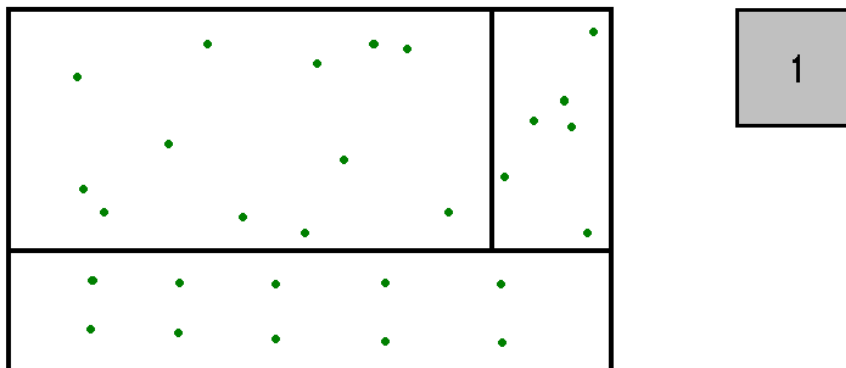
Od této chvíle se budeme omezovat pouze na rovinu, tedy dvourozměrný prostor. Předmětem našeho zkoumání budou situace, které je možné názorně nakreslit tužkou na papír. Konkrétně to budou v podstatě tři druhy objektů: jednotlivé body, dále čáry, tj. přímky, křivky apod., a nakonec nějaké oblasti v rovině, tj. rovinné geometrické obrazce. Naším cílem bude jejich základní popis pomocí vhodně zavedených veličin.

<sup>3</sup> Střední šířka obrazce v rovině je střední hodnota délky jeho ortogonální projekce na přímky všech orientací.

### 3. Aplikace geometrické statistiky

#### 3. 1. Bodové systémy

Začneme nejjednodušším případem - souborem bodů. Pokud budeme sledovat osázení stromy v zahrádkářské kolonii<sup>4</sup>, můžeme situaci zakreslit následovně:



Obr. 3 – Intenzita výsadby v zahrádkářské kolonii, šedý čtverec znázorňuje jednotku obsahu

Celá kolonie je rozdělena na tři různě velké parcely. Na každé z nich můžeme vyčíslit hustotu osázení, neboli průměrný počet stromů na jednotkové oblasti (jednotka obsahu je na obrázku vpravo, jako tomu bude i na mnoha dalších obrázcích). Tato veličina se nazývá *intenzita*, značí se  $\lambda$  a vypočítá se tak, že počet stromů (obecně bodů) vydělíme obsahem dané oblasti. Označíme-li  $\lambda_1$  intenzitu na levé horní,  $\lambda_2$  na pravé horní a  $\lambda_3$  na spodní parcele, pak podle obecného vzorce

$$\lambda_i = \frac{n_i}{a_i}, \quad (1)$$

kde  $n_i$  je počet stromů na  $i$ -té oblasti a  $a_i$  její obsah, dostaneme pro  $i = 1, 2, 3$  tyto hodnoty:

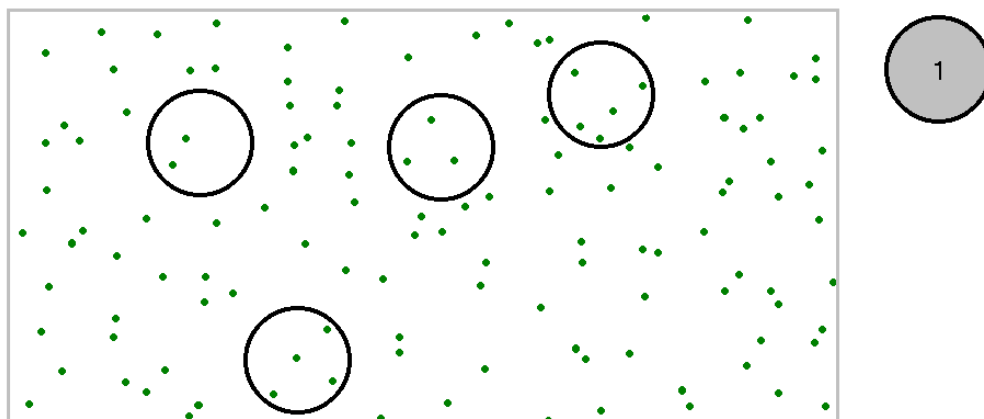
$$\lambda_1 = \frac{12}{8} = 1,5, \quad \lambda_2 = \frac{6}{2} = 3, \quad \lambda_3 = \frac{10}{5} = 2.$$

V tomto případě tedy vyšla intenzita největší na oblasti, která má nejmenší obsah. Dále si všimněme, že ve spodní oblasti jsou stromy vysázeny celkem pravidelně, to ovšem nemá na intenzitu v této parcele pražádný vliv. I kdyby bylo

<sup>4</sup> Nejedná se o úplně náhodný proces, stromy například nemohou růst bezprostředně vedle sebe, navíc mají nezanedbatelnou tloušťku. Pro naše účely je však tento příklad dostatečně názorný.

všech 10 stromů soustředěno do malého háječku v pravém dolním rohu kolonie, dostali bychom stejný výsledek.

Pokud budeme zjišťovat intenzitu výsadby na nějakém rozsáhlém lesním komplexu, jehož obdélníkovou část zachycuje obr. 4, bylo by obtížné počítat všechny stromy, ale dá se učinit odhad pomocí intenzit v nějakých menších podoblastech.



Obr. 4 – Odhad intenzity výsadby v lesním komplexu

Vypočítáme intenzitu pouze na čtyřech kruhových oblastech, bude to součet stromů v těchto kruzích dělený součtem obsahů kruhů:

$$\lambda = \frac{\sum_i n_i}{\sum_i a_i} . \quad (2)$$

V našem případě nabývá index  $i$  hodnot 1, 2, 3, 4 a pro každý je  $a_i = 1$ .

Pokud ze vzorečku (1) vyjádříme  $n_i$  a dosadíme do (2), můžeme po úpravě dospět ještě k jinému vyjádření  $\lambda$  :

$$\lambda = \frac{\sum_i a_i \lambda_i}{\sum_i a_i} = \frac{a_1 \lambda_1 + a_2 \lambda_2 + \dots}{\sum_i a_i} = \frac{a_1}{\sum_i a_i} \lambda_1 + \frac{a_2}{\sum_i a_i} \lambda_2 + \dots$$

a označíme-li  $c_i = \frac{a_i}{\sum_i a_i}$  tzv. plošnou váhu  $i$ -té oblasti, dostaneme

$$\lambda = \sum_i c_i \lambda_i . \quad (3)$$

Dále dosadíme  $c_1 = c_2 = c_3 = c_4 = \frac{1}{4}$ ,  $\lambda_1 = 2$ ,  $\lambda_2 = 3$ ,  $\lambda_3 = 5$ ,  $\lambda_4 = 4$  :

$$\lambda = \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 3 + \frac{1}{4} \cdot 5 + \frac{1}{4} \cdot 4 = 3,5 .$$

Stejnou hodnotu bychom pochopitelně dostali dosazením počtů stromů  $n_1 = 2$ ,  $n_2 = 3$ ,  $n_3 = 5$ ,  $n_4 = 4$  a obsahů oblastí  $a_1 = a_2 = a_3 = a_4 = 1$  do vzorečku (2).

Pokud bychom chtěli pro srovnání spočítat intenzitu v celém obdélníku na *obr. 4*, získali bychom vydělením celkového počtu stromů 128 obsahem obdélníka 40,74 hodnotu cca 3,14, což teoreticky poskytuje přesnější odhad skutečné intenzity v celém lese. Zda a jak je vychýlený záleží na volbě obdélníkové oblasti.

Vzorce (2), (3) jsou samozřejmě aplikovatelné i na první případ – zahrádkářskou kolonii. Aplikací (2) na kolonii na *obr. 3* bychom dostali celkovou intenzitu v kolonii:

$$\lambda = \frac{\sum_i n_i}{\sum_i a_i} = \frac{12 + 6 + 10}{8 + 2 + 5} = \frac{28}{15} \doteq 1,87$$

a stejnou hodnotu bychom získali dosazením příslušných  $c_i$  a  $\lambda_i$  do (3).

Ale zatímco u zahrádkářské kolonie poskytuje  $\lambda$  přesnou hodnotu intenzity v celé zkoumané oblasti, protože známe celkový počet stromů a příslušné rozměry, u lesního komplexu nám dává pouze odhad celkové intenzity ze znalosti jen několika podoblastí (resp. je to přesná hodnota jen na sjednocení těchto podoblastí).

U takového odhadu můžeme zjistit ještě tzv. *směrodatnou odchylku*, která udává jakousi míru variability intenzity na jednotlivých zkoumaných podoblastech, značí se  $\sigma$ . Je-li intenzita na každé z nich blízká hodnotě, kterou odhad poskytuje, pak je směrodatná odchylka malá, zatímco liší-li se hodnoty intenzity na různých oblastech výrazněji, odpovídá tomu i vyšší hodnota směrodatné odchylky. Dosadme již známé hodnoty intenzit do obecného vzorce:

$$\sigma = \sqrt{\frac{\sum_i (\lambda_i - \lambda)^2}{n - 1}} = \sqrt{\frac{(2 - 3,5)^2 + (3 - 3,5)^2 + (5 - 3,5)^2 + (4 - 3,5)^2}{3}} \doteq 1,29$$

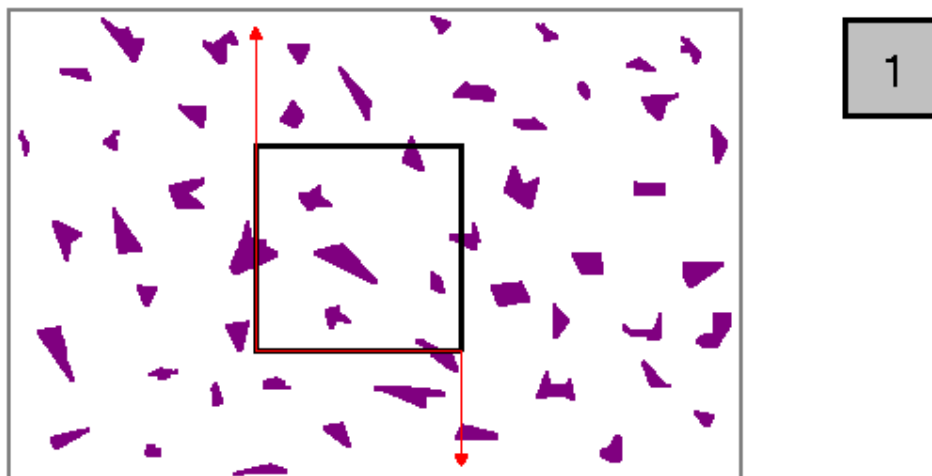
Na závěr této kapitoly převedu celou záležitost do řeči geometrické statistiky: úkolem bylo zkoumat intenzitu bodového systému v omezené rovinné oblasti. Zatímco v první úloze jsme měli k dispozici celý statistický soubor a ze zjištěných charakteristik (počty bodů a obsahy oblastí) jsme mohli přesně určit požadovanou intenzitu, ve druhé situaci jsme nejprve provedli geometrický výběr jako průnik čtyř kruhových oblastí se zkoumaným systémem, na něm jsme zjistili potřebné charakteristiky a výsledek posloužil jako odhad intenzity celého zkoumaného souboru, navíc doplněný o směrodatnou odchylku.

Je třeba podotknout, že přesnost tohoto odhadu velmi závisí na tom, které oblasti vybereme. Pokud zahrneme pouze ty uprostřed lesa, kde zrovna bude hustota zalesnění výrazně větší oproti okrajovým oblastem, bude odhad nadhodnocený,

obrácený výsledek můžeme naopak dostat při výběru pouze okrajových oblastí, nebo pokud se například trefíme na rozsáhlou mýtinu. Již tady je patrné, s jakými obtížemi se setkáváme při realizaci geometrického výběru. Dále je i zde vhodné připomenout platnost obecného statistického principu, a sice že odhad je tím přesnější, čím větší výběrový soubor máme. Výběrovým souborem je v tomto případě souhrnný obsah oblastí, na kterých zjistíme přesnou intenzitu systému.

### 3. 2. Dvourozměrné částice v rovině

V předchozí kapitole jsme se zabývali pouze soubory bodů v rovině, ale nyní přijdou ke slovu objekty složitější, alespoň tedy co do počtu rozměrů. Jako jsme odhadovali intenzitu bodů pomocí několika výřezů, budeme nyní chtít odhadnout intenzitu částic rozmístěných na nějaké velké ploše, přičemž budeme mít k dispozici pouze malé pozorovací okénko.



Obr. 5 – Částice v rovině, čtvercové pozorovací okénko a červená vylučovací čára, která je ve vertikálních částech neomezená

Zatímco u souboru bodů se mlčky předpokládalo, že žádný z bodů neleží na okraji pozorovacího okénka<sup>5</sup>, v tomto modelu již k tomu dojít může. Protnutí částice okrajem okénka budeme nazývat *okrajovým jevem* a jak se ukáže, jedná se o klíčový problém celého našeho úkolu. Otázkou totiž je, jestli se protnuté částice mají považovat za součást okénka, anebo nikoliv.

Kdybychom započítávali všechny protnuté částice, byl by celkový odhad nadhodnocený, při nezahrnutí žádné z nich naopak podhodnocený. Navíc nestejně rozměry částic způsobují, že každá z nich může mít jinou pravděpodobnost, že u ní nastane okrajový jev, a proto ani započítání poloviny protnutých částic neposkytuje

<sup>5</sup> V matematickém modelu je pravděpodobnost takové koincidence rovna nule. V praxi by se pak mohla do vnitřku okénka započítat polovina takových sporných bodů.

dostatečně „spravedlivý“ odhad. Ukážeme si dva způsoby, jak se s okrajovými jevy vypořádat:

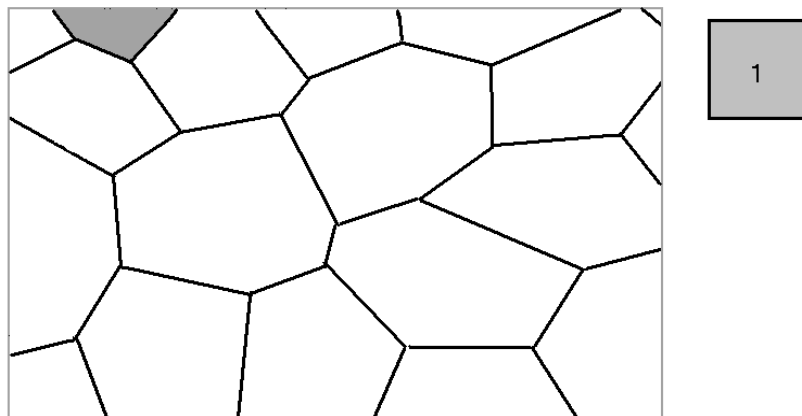
- i) Z protnutých částic budeme do pozorovacího okénka započítávat pouze ty, které nejsou protnuty červeně vyznačenou neomezenou *vylučovací čarou*<sup>6</sup>. Pokud si představíme, že je celá rovina pokrytá takovými čtverečky s příslušnými vylučovacími čarami, zajišťuje nám toto pravidlo, že každá částice bude započítána v právě jednom okénku. Jelikož je obsah okénka 4 jednotky a tímto postupem napočítáme 6 částic, dostáváme podle známého vzorce odhad intenzity  $\lambda = \frac{6}{4} = 1,5$ .

- ii) Každé částici se přiřadí řekněme čtyři *vztažné body* - nejsevernější, nejjihnější, nejvýchodnější a nejzápadnější (resp.  $\max(y)$ ,  $\min(y)$ ,  $\min(x)$ ,  $\max(x)$ ). Potom stačí vydělit počet všech vztažných bodů v pozorovacím okénku čtyřmi a dostaneme odhad průměrného počtu částic na ploše o velikosti našeho pozorovacího okénka. 23 vztažných bodů dává 5,75 částic a odhad intenzity je  $\lambda = \frac{5,75}{4} \doteq 1,44$ .

Získali jsme odhad intenzity dvěma různými metodami. Ta druhá v pořadí je obecně přesnější, navíc je možné ji vylepšit určováním většího počtu vztažných bodů (podle nějakých složitějších pravidel). Nevýhodou prvního postupu je také nepoužitelnost v případě, že pozorovací okénko bude například kruhové.

Pokud nás pro srovnání zajímá přesná intenzita částic v celém obdélníku na obr. 5, pak 53 částic na ploše asi 36,06 jednotek dává hodnotu přibližně 1,47. Přesnost odhadů je zde lepší, než jak tomu bylo u souboru bodů na obr. 4. Již pouhý pohled napovídá, že částice jsou zde rozmístěny pravidelněji a netvoří shluky ani výraznější prázdná místa. Při větším počtu pozorování by se tak dala očekávat i menší směrodatná odchylka.

Analogií předchozího jevu je určování intenzity buněk mozaiky:



Obr. 6 – Určování intenzity výskytu buněk

<sup>6</sup> Tzv. Gundersenovo výběrové pravidlo (viz [7]).

Také tady by se daly obě uvedené metody použít. Je však potřeba dávat pozor na to, že kromě pozorovacího okénka je často nutné vidět ještě jeho dostatečné okolí. Považujeme-li zde celý obrázek za pozorovací okénko, není například jasné, zda tmavě vyznačená buňka protíná pomyslnou vylučovací čáru (jako je nakreslená na *obr. 5*) či nikoliv. Ze stejného důvodu by selhala i druhá metoda, určování vztažných bodů by mohlo být ještě nejasnější. Všechny buňky (resp. částice), které alespoň z části zasahují okénko, musíme pokud možno vidět celé.

Existují však také odhady, které kalkulují pouze s buňkami viditelnými uvnitř okénka bez nutnosti znát zmiňované okolí. Nejpoužívanějším je

$$\lambda = \frac{n_u + 0,5 \cdot n_p - 1}{a}, \quad (4)$$

kde  $n_u$  je počet buněk, které leží celé uvnitř okénka,  $n_p$  počet buněk zasahujících okénko pouze zčásti a  $a$  obsah okénka. Aplikace tohoto vzorečku je jednoduchá, ale nepřihlíží příliš k charakteru konkrétní situace. Pro *obr. 6* dostáváme jeho použitím při obsahu okénka  $a = 22,4$ :

$$\lambda = \frac{3 + 0,5 \cdot 16 - 1}{22,4} \doteq 0,45 .$$

Pro porovnání<sup>7</sup>:

Metodou i) vychází hodnota  $\lambda = \frac{10}{22,4} \doteq 0,45$ , protíná-li tmavá buňka vyluč. čáru,

$$\lambda = \frac{11}{22,4} \doteq 0,49 \text{ v opačném případě.}$$

Použitím metody ii) se čtyřmi vztažnými body pak dostáváme po napočítání 38

vztažných bodů hodnotu  $\lambda = \frac{9,5}{22,4} \doteq 0,42$ .

### 3. 3. Buffonovy úlohy

Po souborech bodů a dvourozměrných částic nyní přijdou na řadu systémy přímků a brzy nám tento oddíl nabídne velmi zajímavé výsledky. Naší snahou opět bude propracovat se k nějakému odhadu intenzity. Cestu nám při tom zpříjemní dvě hry, které nesou jméno francouzského učenice G. Buffona (1707-1788).

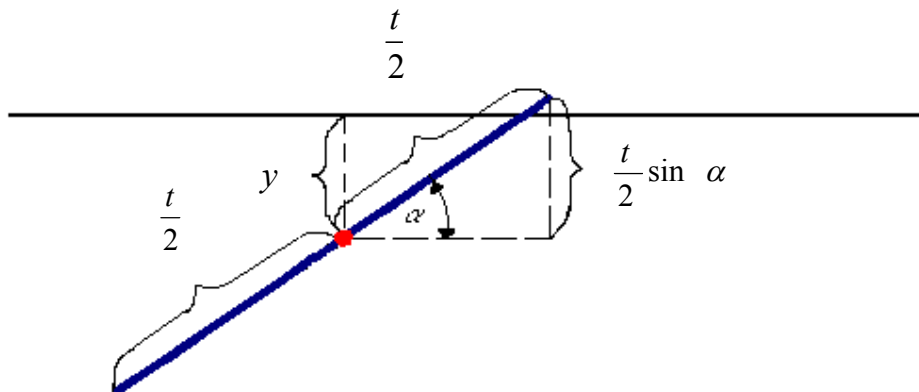
#### • Buffonova úloha o jehle

Představme si, že jehlu délky  $t$  házíme na podlahu, na které je síť rovnoběžných přímků vzdálených od sebe o délku  $d > t$  (např. rýhy mezi prkny v dřevěné podlaze). Úkolem je určit pravděpodobnost, že jehla protne některou z

<sup>7</sup> Pro použití metod i), ii) je tady nutný předpoklad konvexity všech buněk mozaiky.

přímek. Čtenář méně zblhlý v matematice může následující řešení převzaté z [11] přeskočit, zajímavý bude zejména výsledek (4).

Každou polohu jehly můžeme popsat dvěma souřadnicemi - vzdáleností  $y$  jejího středu od nejbližší přímky a úhlem  $\alpha$ , který svírá jehla s daným systémem rovnoběžek<sup>8</sup>:

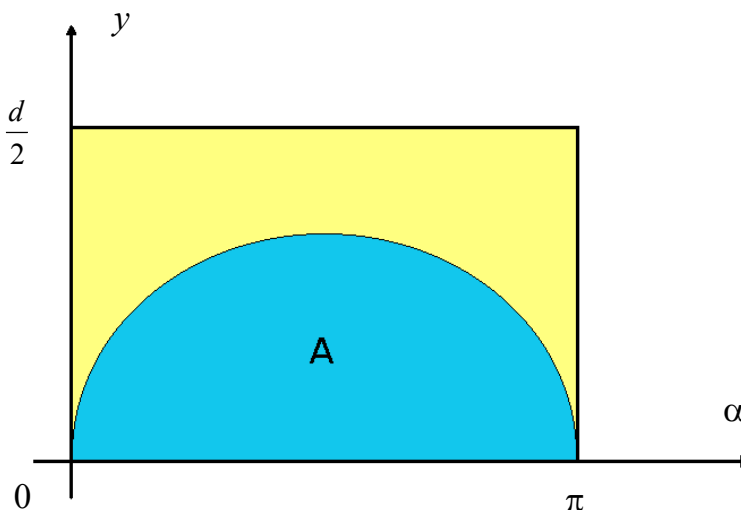


Obr. 7 – Vzájemná poloha jehly a přímky

Hodnoty  $y$ ,  $\alpha$  jsou omezeny nerovnostmi  $0 \leq y \leq \frac{d}{2}$ ,  $0 \leq \alpha < \pi$ .

Jehla protne nejbližší přímku, jestliže  $\frac{t}{2} \sin \alpha \geq y$  (jako je tomu i na obr. 7).

V následujícím grafu je na vodorovnou osu vynesena úhel  $\alpha$  a na svislou osu vzdálenost  $y$ . Modře vyznačená oblast A znázorňuje ty hodnoty  $y$  a  $\alpha$ , při kterých jehla protne nejbližší přímku, celý obdélník pak zahrnuje všechny přípustné hodnoty  $y$  a  $\alpha$ :



Obr. 8 – Graf znázorňující polohy jehly

<sup>8</sup> Uvažuje se zde úhel  $\alpha \in [0; \pi)$ . Díky vlastnostem funkce sinus je rozměr  $\frac{t}{2} \sin \alpha$  (viz obr. 7)

platný i pro  $\alpha \geq \frac{\pi}{2}$ , i když to na obrázku přímo není vidět.



Obsah oblasti A („příznivé případy“) se vypočítá:

$$|A| = \int_0^{\pi} \frac{t}{2} \sin \alpha \, d\alpha = \left[ -\frac{t}{2} \cos \alpha \right]_0^{\pi} = \frac{t}{2} + \frac{t}{2} = t.$$

Protože obsah celého obdélníka („všechny případy“) je  $\pi \frac{d}{2}$ , dostáváme hledanou pravděpodobnost

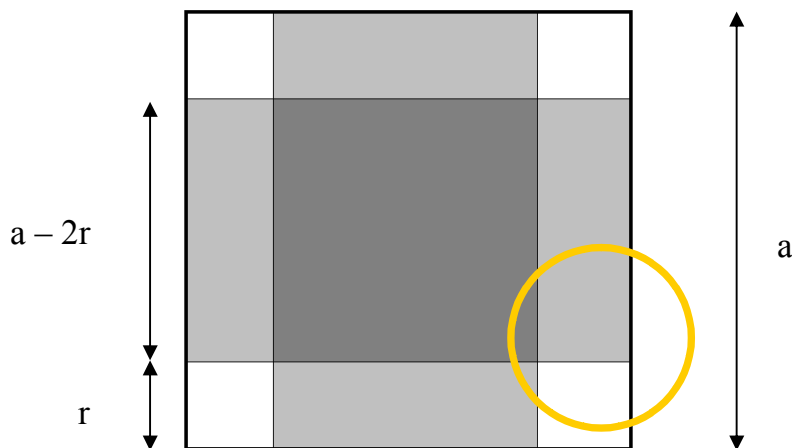
$$P = \frac{2 \cdot t}{\pi \cdot d} \quad (5)$$

Tento výsledek je zajímavý tím, že se v něm vyskytuje číslo  $\pi$ . V dodatku se lze dočíst, jak se toho dá snadno využít k aproximaci hodnoty  $\pi$ .

#### • Buffonův čtverec

Obdobou předchozí úlohy je hra, jíž si údajně krátila dlouhou chvíli již renesanční šlechta. Jedná se v tomto případě o házení mince na podlahu se čtvercovou sítí (např. dlaždičky). Cílem je zasáhnout některou z čar. My opět určíme pravděpodobnost, že k zásahu dojde, ovšem při náhodném, tedy neměřeném hodu.

Poloměr mince označíme  $r$ , stranu dlaždičky  $a$ . Stačí se zaměřit na jednu dlaždičku a podívat se, kde leží střed mince. Čtverec si rozdělíme následovně<sup>9</sup>:



Obr. 9 – Buffonův čtverec, jedna z možných poloh mince

<sup>9</sup> Předpokládá se, že  $a > 2r$ . Pak lze toto rozdělení provést.

Nyní je snadné si rozmyslet, že mince nezasáhne žádný okraj, jestliže její střed bude ležet v tmavě vyznačené oblasti o obsahu  $S_0 = (a - 2r)^2$ . V případě zásahu může dojít ke dvěma možnostem:

- i) Mince protne pouze jednu spáru, a to ve dvou průsečících (obr. 9). Stane se tak v případě, že střed mince bude ležet v jedné ze světle šedých oblastí, součet jejich obsahů je  $S_1 = 4r(a - 2r)$ .
- ii) Mince protne dvě spáry – jednu horizontální a jednu vertikální. Vytvoří se tak čtyři průsečíky (některé mohou být již mimo obr. 9). Střed mince by pak musel ležet v jednom z bílých čtverečků o úhrnném obsahu  $S_2 = 4r^2$ .

Pokud se ptáme, zda mince zasáhla některou z čar, je nám jedno, o který z obou předchozích případů se jedná. Pravděpodobnost tohoto jevu určíme podle schématu, se kterým se již čtenář setkal ve druhé kapitole u *obr. 1* (a také u řešení předchozí úlohy). Tentokrát je obsah oblastí zahrnující příznivé výsledky roven  $S_1 + S_2$ , všechny možné polohy středu mince znázorňuje celý čtverec o obsahu  $S = a^2$ . Proto dostáváme

$$P = \frac{S_1 + S_2}{S} = \frac{4r(a - 2r) + 4r^2}{a^2} = \frac{4r(a - r)}{a^2}. \quad (6)$$

Ekvivalentně se dá  $P$  počítat jako  $P = \frac{S - S_0}{S}$  se stejným výsledkem.

Dále je možné vyjádřit pravděpodobnosti, že dojde k jednotlivým případům i), ii):

$$P_1 = \frac{S_1}{S} = \frac{4r(a - 2r)}{a^2}, \quad (7)$$

$$P_2 = \frac{S_2}{S} = \frac{4r^2}{a^2}. \quad (8)$$

A pro úplnost: pravděpodobnost, že mince žádnou čáru nezasáhne, je

$$P_0 = \frac{S_0}{S} = \frac{(a - 2r)^2}{a^2}. \quad (9)$$

Dalo by se na tomto místě namítnout, že v rozboru nejsou zahrnuty takové varianty, kdy se mince nějaké strany nebo rohu dlaždičky pouze dotkne, čili okraj mince je přesně na spáře. Došlo by k tomu v případě, že by její střed ležel na rozhraní některých různobarevných oblastí na *obr. 9*. V reálné situaci by se snad mohlo stát, že bychom minci hodili tak, že by měla se spárou jeden dotykový bod, na vině by však byly zejména nedostatečné rozlišovací schopnosti a fakt, že spára ve skutečnosti nemá zanedbatelnou šířku. Tyto situace jsou ovšem v matematickém modelu nezajímavé z toho důvodu, že pravděpodobnost jejich výskytu je nulová. Zde je stručné a zjednodušené vysvětlení:

Pokud budeme pravděpodobnost počítat stejným způsobem jako u ostatních případů, zjistíme nejprve obsah oblasti, kde musí ležet střed mince, aby k žádanému jevu došlo. Zde je však zásadní fakt, že se nyní jedná pouze o úsečky – jednorozměrné útvary, jejichž obsah je nulový (úsečku si lze představit jako obdélník, jehož jedna strana má délku nula). Pravděpodobnost by se pak vyčíslila jako podíl nuly a obsahu celého čtverce, tedy opět nula.

### • *Určování intenzity*

Podívejme se nyní na obě předchozí úlohy z jiného hlediska a začněme Buffonovou úlohou o jehle. Nebudeme tentokrát házet jednu, ale více jehel, a budeme zjišťovat, kolik vytvoří na systému rovnoběžek průsečíků.

V předpokladech úlohy bylo, že délka jehly je menší než vzdálenost mezi sousedními přímkami, proto každá jehla protne nejvýše jednu přímkou. Počet indukovaných průsečíků  $N$  je tedy roven počtu jehel, které protnou některou z rovnoběžek. My tento počet nemůžeme samozřejmě určit přesně, jedná se o náhodný pokus, jehož výsledek nelze s jistotou předpovědět. Ovšem můžeme získat nějaký odhad výsledku, a to pomocí již vypočítané pravděpodobnosti  $P$ , že jedna hozená jehla protne některou z přímek (vzorec (5)). Vynásobíme-li tuto pravděpodobnost počtem hozených jehel, označme ho  $n$ , získáme právě hledaný odhad počtu průsečíků:

$$N = P \cdot n = \frac{2 \cdot t \cdot n}{\pi \cdot d} . \quad (10)$$

Připomeňme, že  $t$  je délka jehly,  $d$  vzdálenost mezi sousedními rovnoběžkami.

Nyní se již dostáváme ke slibované intenzitě, otázkou je, co má zde tento pojem vyjadřovat. Na podlaze máme v tuto chvíli naházeno  $n$  jehel o celkové délce  $l = n \cdot t$ . Zavedeme nejprve pojem *intenzita průsečíků* jako podíl počtu průsečíků jehel s přímkami a celkové délky jehel. Tato veličina vyjadřuje, kolik průsečíků připadá na každou jednotku celkové délky naházených jehel, vezměme kupříkladu jeden metr<sup>10</sup> (vztahuje se zde k jednotce délky, ne plochy!). Ve skutečnosti však máme k dispozici jenom odhad počtu průsečíků, proto následující vzoreček bude pouze odhadem intenzity:

$$\lambda_p = \frac{N}{l} . \quad (11)$$

Po dosazení za  $N$  ze vzorce (10) a za  $l$  jako  $l = n \cdot t$  dostaneme:

$$\lambda_p = \frac{2 \cdot t \cdot n}{\pi \cdot d \cdot n \cdot t} = \frac{2}{\pi \cdot d} . \quad (12)$$

<sup>10</sup> Místo metru lze za jednotku zvolit libovolnou jinou délkovou míru.

Házíme-li jehly na podlahu se vzdáleností spár  $d = 7$  cm, můžeme říci, že na jeden metr délky jehel připadá podle tohoto odhadu

$$\lambda_p = \frac{2}{\pi \cdot 0,07} \doteq 9,09 \quad \text{průsečíků.}$$

Stojí za povšimnutí, že ve vzorečku (12) nevystupuje délka jehly  $t$  ani počet jehel  $n$ . Na intenzitu má vliv pouze vzdálenost  $d$  mezi spárami – čím je menší, tím větší očekáváme intenzitu průsečíků.

Dá se říci, že síť rovnoběžek je tím hustší, čím je jejich vzájemná vzdálenost menší. Pro charakterizaci oné hustoty má smysl zavést veličinu *délková intenzita čar*, označme ji  $\lambda_c$ , která bude vyjadřovat průměrnou délku úseček vymezených přímkami na ploše 1 metr. Protože vzdálenost sousedních přímek je v našem případě rovna  $d$ , je délková intenzita čar

$$\lambda_c = \frac{1}{d} . \quad (13)$$

Porovnáním s (12) dostaneme vztah<sup>11</sup>

$$\lambda_p = \frac{2}{\pi} \cdot \lambda_c . \quad (14)$$

Tento výsledek se dá interpretovat tak, že mezi intenzitou průsečíků indukovaných jehlami na systému rovnoběžek a délkovou intenzitou (hustotou) těchto přímek je přímá úměra. To odpovídá představě, že čím jsou rozestupy mezi přímkami menší, tím větší je šance, že jehlou některou z nich zasáhneme, což dává při dané celkové délce naházených jehel větší počet průsečíků.

Zaměřme se nyní na druhou Buffonovu úlohu. Zde bude určování intenzity nepatrně složitější, postup však bude zcela analogický. Mějme naházeno  $n$  mincí o poloměru  $r$  a zajímejme se o počet průsečíků jejich obvodů se čtvercovou sítí na podlaze.

Vzorce (7) a (8) udávají pravděpodobnost, že jedna hozená mince takto vytvoří 2, respektive 4 průsečíky. Na celkový počet mincí  $n$  tedy připadá průsečíků:

$$N = 2nP_1 + 4nP_2 = \frac{8nr(a-2r)}{a^2} + \frac{16nr^2}{a^2} = \frac{8nr(a-2r+2r)}{a^2} = \frac{8nr}{a} . \quad (15)$$

Opět zavedeme veličinu intenzita průsečíků. V tomto případě bude vyjadřovat, kolik průsečíků připadá na jeden metr obvodů mincí. Označíme-li ve shodě s rozbořením první situace celkovou délku obvodů naházených mincí  $l$ , pak definujeme intenzitu průsečíků stejným vztahem:

$$\lambda_p = \frac{N}{l} . \quad (16)$$

<sup>11</sup> Vztah je platný díky tomu, že jsme implicitně zahrnuli předpoklad, že orientace jehel jsou izotropní, neboli každá orientace jehly je stejně pravděpodobná.

Délku  $l$  můžeme snadno vyjádřit ze znalosti poloměru mince, obvod totiž vypočítáme jako  $2\pi r$ , což násobeno počtem mincí dává  $l = 2\pi rn$ . Do výše uvedené intenzity lze tedy dosadit:

$$\lambda_P = \frac{8nr}{2\pi rn} = \frac{4}{\pi a}. \quad (17)$$

I zde popíšeme hustotu čtvercové sítě veličinou délková intenzita čar, která bude vyjadřovat délku čar na ploše  $1 \text{ m}^2$ , budeme přitom chtít dospět k vyjádření pomocí známé strany čtverce  $a$ . Každý čtverec má obvod  $4a$ , jelikož je však každá strana na rozhraní dvou čtverců, je nutné ji započítat pouze u jednoho z nich, na každý čtverec tak připadá délka spár rovná polovině jeho obvodu, tj.  $2a$ . Kýženou délkovou intenzitu získáme vydělením této hodnoty obsahem jednoho čtverce:

$$\lambda_C = \frac{2a}{a^2} = \frac{2}{a}. \quad (18)$$

Porovnáním posledních dvou vzorců dojdeme znovu ke vztahu

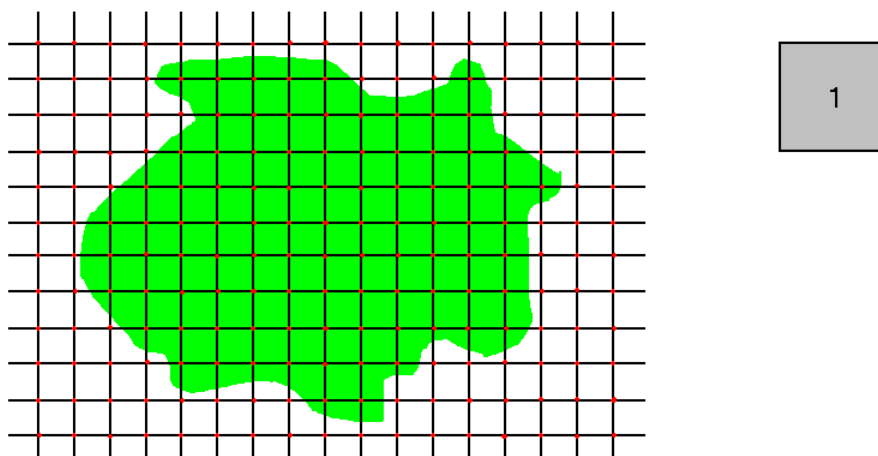
$$\lambda_P = \frac{2}{\pi} \cdot \lambda_C.$$

Různých variací na téma Buffonových úloh by se dalo samozřejmě vymýšlet mnoho a mnoho. Co se týče tvaru házených předmětů a struktury zasahovaného čárového systému, neklade si fantazie žádné meze. Již srovnání předchozích dvou ukázek ovšem napovídá, že některé vlastnosti mají obecnější charakter. O jejich praktickém využití bude pojednávat následující kapitola. Více podrobností o obou uvedených úlohách se lze dočíst např. v [3] a [12].

## 4. Měření charakteristik rovinných objektů pomocí geometrické statistiky

### 4. 1. Měření obsahu rovinného obrazce

Představme si, že jsme postaveni před následující praktický úkol: odhadnout obsah nějakého rovinného obrazce. O žádný problém se samozřejmě nejedná v případě, že jde například o trojúhelník, lichoběžník nebo kružnici, pak stačí změřit vhodné rozměry a podle jednoduchého vzorce obsah dopočítat. V obecném případě to však možné není a je potřeba učinit nějaké přiblížení. Ukážeme si, jak s výhodou využít metodu využívající nejzákladnější poznatky geometrické statistiky.



Obr. 10 – Měření obsahu obrazce pomocí mřížkového testovacího systému

Obr. 10 zachycuje rovinný obrazec a na něj přiloženou mřížku, jejíž rozměry můžeme snadno určit a využít ji jako tzv. *testovací systém*. Průsečíky čar mřížky, které jsou na obrázku zvýrazněny červenými tečkami, budeme nazývat *mřížkové body*. Porovnáním jejich rozestupů s jednotkovým čtvercem se dá ověřit, že na jeden bod připadá plocha  $1/9$ . Nyní stačí zjistit počet mřížkových bodů ležících uvnitř obrazce<sup>12</sup>, což je 92, a provést triviální úvahu: Na každý bod připadá plocha  $1/9$  a uvnitř našeho obrazce je 92 bodů. Budeme-li tuto závislost uvažovat jako přímou úměrnost, můžeme psát

$$S \doteq \frac{1}{9} \cdot 92 \doteq 10,22$$

a máme tak hledaný odhad obsahu  $S$ .

Uvedená metoda se dá precizněji interpretovat pomocí pojmu intenzita bodového systému, zavedeného v kapitole 3.1. Tato veličina je definována jako podíl počtu bodů ku obsahu oblasti, na které se vyskytují. Zapsáno vzorcem:

$$\lambda = \frac{n}{S},$$

kde  $n$  je počet bodů testovacího systému uvnitř obrazce o obsahu  $S$  a  $\lambda$  je intenzita systému.

Tentokrát však pro nás není neznámou intenzita  $\lambda$ , kterou máme přesně určenou z rozměrů testovací mřížky (v našem případě je rovna 9), ale chceme vyjádřit obsah  $S$ . Takže klíčový vzorec této kapitoly, který dává obecný návod na odhadnutí obsahu obrazce pomocí libovolného pravidelného testovacího systému o známé intenzitě, je

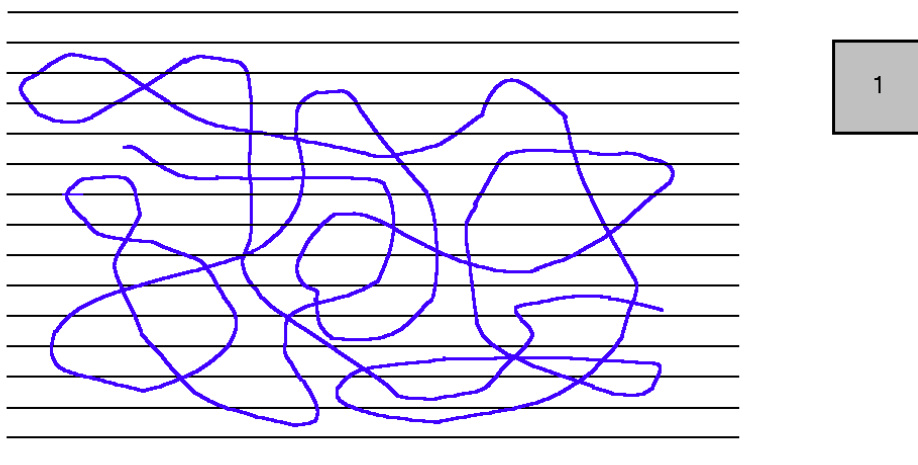
$$S = \frac{n}{\lambda}. \quad (19)$$

<sup>12</sup> Vyskytují-li se v reálné situaci body, u nichž nelze rozlišit, zda leží uvnitř, nebo vně obrazce, můžeme do vnitřku zahrnout polovinu z nich.

Nutno říci, že uvedená metoda při základním použití, které bylo právě popsáno, nemusí být příliš přesná. Ve skutečnosti se ovšem obvykle neaplikuje pouze jednou, ale realizuje se série několika měření, přičemž výsledný odhad se získá zprůměrováním zjištěných hodnot. Praktické provedení právě popsaného spočívá v tom, že máme testovací systém nakreslený na průhledné fólii, kterou opakovaně přikládáme v různých polohách na měřený obrazec a pokaždé zjistíme počet průsečíků ležících uvnitř. Čím vícekrát takto učiníme, tím přesnější odhad teoreticky dostaneme.

#### 4. 2. Měření délky křivky

V tomto oddíle bude naším úkolem nějakým způsobem odhadnout délku křivky v rovině. V první Buffonově úloze v kapitole 3.3. jsme házeli jehly na systém rovnoběžek a zjistili jsme vztah mezi intenzitou průsečíků jehel s přímkami a délkovou intenzitou čar systému (vzorec (14)). Libovolnou křivku si ovšem můžeme představit jako řadu za sebou naházených drobných jehliček, jejichž délka se limitně blíží k nule. To je zhruba důvod, proč lze znalosti nabyté v kapitole 3.3. využít k řešení našeho problému.



Obr. 11 – Měření délky křivky pomocí souboru rovnoběžek

Délkovou intenzitu rovnoběžek známe, v našem případě je to  $\lambda_C = 3$ . Nyní stačí zjistit počet průsečíků křivky s rovnoběžkami, což je  $N = 97$ , a máme k dispozici vše potřebné. Dosadíme-li totiž do vzorce (14) za  $\lambda_p$  výraz z rovnice (11), dostaneme již výsledný odhad pro délku křivky:

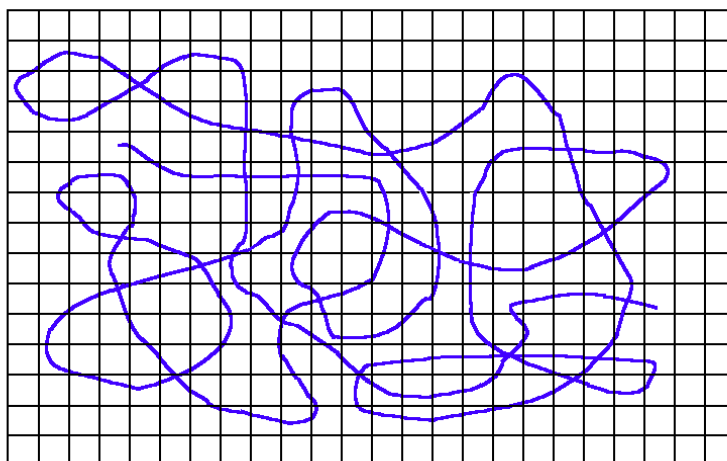
$$\frac{N}{l} = \lambda_p = \frac{2}{\pi} \cdot \lambda_C \quad \Rightarrow \quad l = \frac{\pi \cdot N}{2\lambda_C} . \quad (20)$$

Takže konkrétně po dosazení:  $l \doteq \frac{\pi \cdot 97}{2 \cdot 3} \doteq 50,79$  ,

přičemž jednotka délky je dána stranou čtverce vpravo na *obr. 11*. Později uvidíme, proč je tento odhad dosti nepřesný.

I tady se dá měření opakovat při různých polohách testovacího systému, abychom zprůměrováním naměřených hodnot dostali co nejpřesnější výsledek. Další vylepšení skýtá například zmenšení rozestupů mezi rovnoběžkami. Tady je ale nutné vznést požadavek, aby křivka byla *izotropní*, tj. aby orientace jejích tečen byla rovnoměrně náhodná. Typickým příkladem ukazujícím, proč v opačném případě nemusí metoda fungovat, je obyčejná úsečka. Pokud totiž testovací systém přiložíme tak nešťastně, že jeho čáry budou s touto úsečkou rovnoběžné, pak se nevytvoří žádný průsečík. Nepomůže přitom ani libovolné zjemňování sítě rovnoběžek (tj. zmenšování jejich rozestupů). Naopak při kolmosti přímek na úsečku se vytvoří mnohem větší počet průsečíků, než se teoreticky očekává.

Také je možné používat jiné testovací systémy, obecně to musí být pravidelné čárové systémy, jejichž intenzitu dokážeme určit. Jako vhodný kandidát se může jevit mřížka stejně jako v předchozí kapitole, tentokrát ovšem nikoliv pouze její mřížkové body, ale celé přímky.



*Obr. 12 – Měření délky křivky pomocí mřížky*

Předešlá úvaha naznačuje, že znalost délkové intenzity čárového systému, která je dvojnásobná oproti předchozímu modelu, nám spolu s napočítáním průsečíků křivky se systémem<sup>13</sup> (zde je to  $N = 239$ ) postačuje k určení odhadu délky křivky. Stačí opět dosadit do univerzálního vzorce (20):

$$l \doteq \frac{\pi \cdot 239}{2 \cdot 6} \doteq 62,57 .$$

Tato hodnota je poměrně výrazně odlišná od prvního odhadu, ve kterém byl použit nevhodný testovací systém. Křivka je totiž přednostně horizontálně

<sup>13</sup> V případě, že křivka prochází mřížkovým bodem, je nutné počítat průsečík s horizontální i vertikální čarou.



orientována a protíná tak více vertikálních než horizontálních čar. Narážíme zde na typický problém geometrické statistiky. Máme nekonečné množství možností, jaký testovací systém zvolit a jakým způsobem ho na objekt položit, ale snažíme se to udělat způsobem, který se zdá být nejlepší. Obecně je potřeba provádět co nejvíce měření s co nejjemnějšími modely a z naměřených hodnot odvodit dostatečně přesný výsledek.

Dlužno podotknout, že oba použité modely byly velice hrubé v porovnání s těmi, které se používají v praxi. Značně odlišné odhady by proto neměly svádět k unáhleným závěrům o nepoužitelnosti geometrické statistiky ke zjišťování charakteristik rovinných objektů při záruce dostatečné přesnosti. Právě naopak, je to v podstatě jediný nástroj k tomuto účelu využitelný a pouze rozvíjení metod geometrické statistiky a schopnost správně vyhodnotit získané výsledky může vést ke kvalitnímu popisu reálných geometrických objektů.

Rád bych ještě zmínil jednu souvislost. V kapitole 3.3. je kromě úlohy o Jehle pojednáno ještě o druhé Buffonově úloze, ve které se hází mince právě na čtvercovou síť, neboli mřížku. Nezávislou metodou tam byl odvozen výše používaný vztah mezi intenzitou průsečíků a délkovou intenzitou čar mřížky. Jak Jehly v první úloze, tak mince ve druhé jsou realizacemi obecného jevu – průniku dvou čárových systémů, z nichž jeden je pravidelný (anizotropní) a druhý náhodný (izotropní). V praxi využijeme jeden z nich jako testovací systém a ze vzájemné interakce odhadneme potřebné charakteristiky druhého.

## 5. Zkoumání trojrozměrných objektů

Tato práce pojednává pouze o charakteristikách rovinných objektů, vztah ke zkoumání trojrozměrných těles je však zcela bezprostřední. Jak bylo popsáno v kapitole 2.2., základem geometrické statistiky je provádění geometrických výběrů, tj. projekcí či průniků, a jejich zkoumání obvykle v prostoru nižší dimenze. Speciálně u trojrozměrných objektů se nejčastěji provádějí řezy rovinou, ty se prozkoumají a z nich se pak usuzuje na vlastnosti celého tělesa.

Například jestliže jsou nějaké 3D částice rovnoměrně náhodně rozmístěny v prostoru a jejich objemový podíl (střední hodnota jejich celkového objemu v jednotce objemu obklopujícího prostoru) je  $V_V$ , pak střední hodnota plošného podílu  $A_A$  jejich řezů libovolnou rovinou je odhadem veličiny  $V_V$ . Stejně tak lze ze stop 3D ploch a křivek v rovině řezu odhadnout jejich 3D intenzity  $S_V$ ,  $L_V$  (střední hodnoty plošného obsahu či délek v jednotce objemu). Podrobnosti lze najít např. v [8], [9].

Vědní disciplína zabývající se odhady vlastností geometrických těles na základě výběrů pořízených geometrickými prostředky se nazývá *stereologie* ([1], [10]). Její použití v nejrůznějších oborech lidské činnosti je velice široké. Uplatňuje se zejména při zkoumání struktury materiálů v metalografii, krystalografii a dalších materiálově-technických vědách, jiným oborem úzce spjatým se stereologií je biomatematika (pěkné pojednání je v [4]), kde jsou naopak předmětem zkoumání živé organismy, vyšetřují se prostorové vlastnosti orgánů, buněčné struktury a podobně. Metodami geometrické statistiky se detekuje výskyt nákazy infekčních

nemocí, své použití najde i v krajinné ekologii, demografii nebo také astronomii. A tím výčet zdaleka není vyčerpán.

I přes širokou škálu využití se však zdá geometrická statistika býti ve školních osnovách trochu opomíjená a navzdory její užitečnosti a významnosti se jedná o pojem pro spoustu lidí neznámý. Zvláště v době rychle se rozvíjejících technologií je to přitom obor velmi dynamický. Konstruování stále dokonalejších nástrojů a zpracovávání získaných výsledků na výkonných počítačích umožňuje neustálé zkvalitňování odhadů prostorových vlastností těles a na jejich základě rozvoj mnoha vědních oborů, které těchto výsledků využívají.

## 6. Dodatek

V Buffonově úloze o jehle (str. 15) jsme dospěli k zajímavému vzorečku. Při hození jehly délky  $t$  na podlahu se systémem rovnoběžných přímek, jejichž rozestupy jsou  $d$ , je pravděpodobnost zásahu nějaké přímky rovna

$$P = \frac{2 \cdot t}{\pi \cdot d} .$$

Speciálně, budeme-li uvažovat vzdálenost mezi sousedními přímkami dvojnásobnou oproti délce jehly ( $d = 2t$ ), vychází

$$P = \frac{1}{\pi} .$$

To poskytuje praktický návod, jak jednoduchou pravděpodobnostní metodou aproximovat hodnotu  $\pi$  (odlehčenou formou je to popsáno v [5]). Dejme tomu, že  $p$ -krát hodíme jehlu na podlahu se systémem rovnoběžek, vzdálených od sebe o dvojnásobek délky jehly. Počet pokusů, při kterých jehla protne nějakou přímku, označíme  $s$ . Potom pravděpodobnost  $P$  můžeme vyjádřit dvěma způsoby. Za prvé jakožto přesnou hodnotu, což je popsáno předchozím vzorcem, a za druhé coby experimentálně získaný odhad tak, že klasicky vydělíme počet příznivých jevů počtem všech jevů:

$$\frac{1}{\pi} \doteq \frac{s}{p} \quad \Rightarrow \quad \pi \doteq \frac{p}{s}$$

Nesmíme samozřejmě opominout, že je to metoda založená na pravděpodobnosti. Jestliže pokus provedeme tisíckrát, nemáme žádnou záruku, že se k číslu  $\pi$  přiblížíme, ale je to nanejvýš pravděpodobné.

Na internetu existuje mnoho simulátorů, které uvádějí tuto metodu v praxi. Kdokoliv tak může nechat svůj počítač virtuálně provést takové množství hodů, na které by ve skutečnosti bylo jen ztěžší pomyslet. Zájemce odkazují např. na [2], [6].

## Literatura

- [1] Baddeley A., Vedel Jensen E. B.: *Stereology for Statisticians*. Chapman & Hall/CRC, Boca Raton 2005.
- [2] Hurben M. J.: *Buffon's needle*. URL ke dni 24. 5. 2006:  
<http://www.angelfire.com/wa/hurben/buff.html>.
- [3] Ilucová L.: *Buffonova úloha o štvorci*. In: Zborník konferencie Matematika v škole dnes a zajtra. PedF Katolíckej univerzity Ružomberok. 12.-14. 9. 2005, 101-106.
- [4] Mandarim-de-Lacerda C. A.: *Stereological Tools in Biomedical Research*. Rio de Janeiro 2003. URL ke dni 24. 5. 2006:  
<http://www.scielo.br/pdf/aabc/v75n4/a06v75n4.pdf>.
- [5] Poskitt K.: *Murderous Maths: Do you Feel Lucky?* Scholastic Ltd., London 2001.
- [6] Reese G.: *Java Applet for Buffon's needle*. URL ke dni 24. 5. 2006:  
<http://www.mste.uiuc.edu/reese/buffon/buffjava.html>.
- [7] Saxl I., Beneš V.: *Geometrický výběr*. In: J. Antoch, J. Štěpán (eds.): *Výuka statistiky v České Republice I*. Matfyzpress, Praha 2002, 79-89.
- [8] Saxl I.: *Geometrická pravděpodobnost*. In: J. Antoch, D. Hlubinka, I. Saxl (edit.): *Sborník Didaktického semináře pro profesory středních škol*. Matfyzpress, Praha 2005, 47-72.
- [9] Saxl I., Beneš V.: *Stereologie: interpretace výběrů pořázených geometrickými prostředky*. In: J. Antoch, G. Dohnal, J. Štěpán: *Výuka statistiky v České Republice II*. Sborník semináře *STAKAN*, Bystřice p. Hostýnem, 23. - 25. 5. 2003. Praha 2005, 107-116.
- [10] Stoyan D., Kendall W.S., Mecke J.: *Stochastic Geometry and its Applications*. J. Wiley & Sons, New York 1995.
- [11] Šmajstrla V.: *Řešené příklady*. URL ke dni 24. 5. 2006:  
<http://homen.vsb.cz/~s1m30/cdpast1/soubory/ZDROJE/resprik2/respr2.htm>
- [12] The University of Alabama in Huntsville, Department of Mathematical Sciences: *Buffon's problems*. In: *Virtual Laboratories in Probability and Statistics*.  
URL ke dni 24. 5. 2006: <http://www.math.uah.edu/stat/buffon/Bufferon.xhtml>