Statistické vyhodnocování experimentálních dat

Mgr. Martin Čada, Ph.D.

- Ústav fyziky a biofyziky, PřF JU
- E-mail: mcada@prf.jcu.cz
- Tel.: 266052418
- Organizace výuky, zkouška, zápočet
- Přednášky a případné informace k výuce budou k dispozici na webu:

www.fzu.cz/~cada

Časový plán kurzu:

1/10 – přednáška + přednáška
15/10 – cvičení + přednáška
29/10 – přednáška + cvičení
12/11 – přednáška + přednáška
26/11 – cvičení + přednáška
10/12 – přednáška + cvičení
17/12 – přednáška + cvičení (zápočet)

Sylabus

- Pravděpodobnost a její definice
- Náhodná proměnná a její charakteristiky
- Distribuční a pravděpodobnostní funkce
- Sdružené distribuce a korelace
- Zákon velkých čísel a centrální limitní věta
- Statistické modely a soubory
- Bootstrap
- Odhady a maximální věrohodnostní odhad
- Metoda nejmenších čtverců a intervaly spolehlivosti
- Testování hypotéz

Literatura

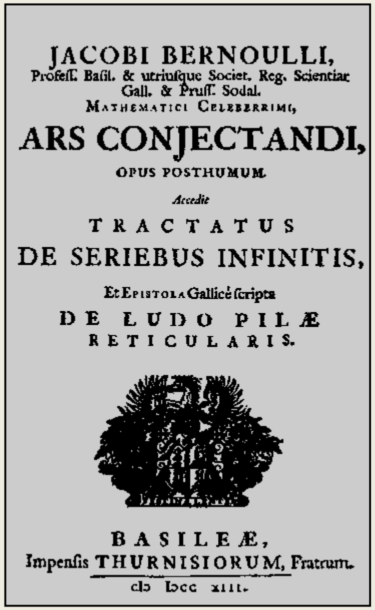
- M. Meloun, J. Militký, „Statistická analýza experimentálních dat“, Academia, Praha 2004.
- J. Štěpán, „Teorie pravděpodobnosti“, Academia, Praha 1987.
- J. Anděl, „Statistické metody“, MatFyzPress, Praha 1998.
- M. Kopecký, „Úvod do počtu pravděpodobnosti a matematické statistiky“, Ped. Fakulta, UPOL 2001.
- K. Zvára, J. Štěpán, „Pravděpodobnost a matematická statistika“, MatFyzPress, Praha 2006.
- A. Plocki, P. Tlustý, „Pravděpodobnost a statistika“, Prometheus, Praha 2007.
- R.J. Barlow, „Statistics. A guide to the use of statistical methods in the physical sciences“, John Wiley & Sons, Chichester 1989.
- G. Cowan, “Statistical Data Analysis”, Oxford Science Publications, Clarendon Press, Oxford 1998.
- J. Pitman, „Probability“, Springer, New York 1993.
- D.C. Montgomery, G.C. Runger, „ Applied Statistics and Probability for Engineers“, John Wiley & Sons, New York 2003.
- F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, L.E. Meester, „A Modern Introduction to Probability and Statistics “, Springer, London 2005.

Historický exkurz

- Hazardní hry provází lidstvo od nepaměti.
- 15. stol. italští matematici řeší partikulární problémy hazardních her (Luca Pacioli, Nicolo Tartaglia, G. Galilei):
  - šance, že padne určitý počet ok na kostce,
  - po kolika hodech bude šance na hození šesti ok větší jak 50%,
  - jak dělit sázku mezi hráče atp.
- Základy teorie pravděpodobnosti položeny v dopisech matematiků: Blaise Pascal a Pierre de Fermat z roku 1654 týkající se hazardních her – nedefinují pojem „pravděpodobnost“.
- Christian Huygens (1657) vydává první spis o počtu pravděpodobnosti při hazardních hrách. Vychází z Pascala a de Fermata. Matematicky formuluje první základní myšlenky počtu pravděpodobnosti.

Historický exkurz

- Jacob Bernoulli ve své práci „Ars Conjectandi“ pokládá základy dnešního mat. pojetí pravděpodobnosti – posmrtně r. 1713.
- Reverend Thomas Bayes publikoval svou zásadní práci týkající se podmíněné pravděpodobnosti – tzv. Bayesův teorém roku (1763).
- Pierre-Simon de Laplace vydává roku 1812 práci „Théorie Analytique des Probabilités“, která je první zásadní analytickou prací v teorii pravděpodobnosti bez vazby na analýzu hazardních her. Definuje $P(A) = m/n$.
- Andrej N. Kolmogorov roku 1933 v knize „Foundations of the Theory of Probability“ pokládá základy moderní matematické teorie pravděpodobnosti založené na axiomech.



Historický exkurz

- 1834 - byla založena první odborná společnost pro statistiku v Londýně.
- 1853 – první statistická konference (věnována biologii a zemědělství).
- K. Pearson a R.A. Fisher na přelomu 19. a 20. století zakládají moderní statistiku na základě statistické analýzy dat.
- J. Neyman a E.S. Pearson pokládají základy statistickému testování hypotéz roku 1933.

Historický exkurz

- Rok 2013 byl vyhlášen mezinárodním rokem statistiky.
- Připomínali jsme si 300 let od vydání Bernoulliho knihy Ars Conjectandi.
- http://www.worldofstatistics.org/
- Zásadní rozvoj statistiky ve 20. století.

Motivace

- Náhodnost versus determinovanost.
- Statistika poskytuje analytické metody pro odhalení variability získaných dat ve fyzikálních, chemických, biologických, medicínských, sociálních, ekonomických… atp. oborech lidské činnosti.
- Na základě velkého množství experimentálních dat mi statistika může poskytnout návod, jak nalézt správné řešení problému.

Motivace - biometrie

- Biometrie – identifikace osob (duhovka). Umožňuje jednoznačnou identifikaci osob podle snímku duhovky.
  - Transformace obrázku duhovky na 2048 bit kód (John Daugman). Kód je velmi citlivý na velikost duhovky a zornice.
  - V různých časech se může stejná duhovka lišit barvou a tvarem.
  - Tolerance detekce stejného člověka: **34% bitů se může lišit!!!**
  - Můžeme se na tuto identifikaci spolehnout, i když třetina kódu se může lišit? **Ano**, díky skutečnosti, že duhovky lidí jsou velmi odlišné.
  - Poměr 0 a 1 v kódu je přibližně 1:1 podle velkého množství experimentálně získaných kódů. Tedy hodnota bitu je náhodná, ale jejich vzájemná korelace je kruciální.
  - Experimentálně ukázáno, že 266 bitů může být považováno za nekorelované.
  - Extrémně malá pravděpodobnost, že duhovky dvou různých lidí se budou lišit v méně jak 34% bitů ($1:10^{80}$).

Motivace - biometrie

- Charakteristické znaky duhovky se vytváří v prenatálním stádiu vývoje náhodně → dvojčata mají různé kódy.
- Hammingova vzdálenost: poměr rozdílných znaků na stejných pozicích v řetězci vůči stejným.

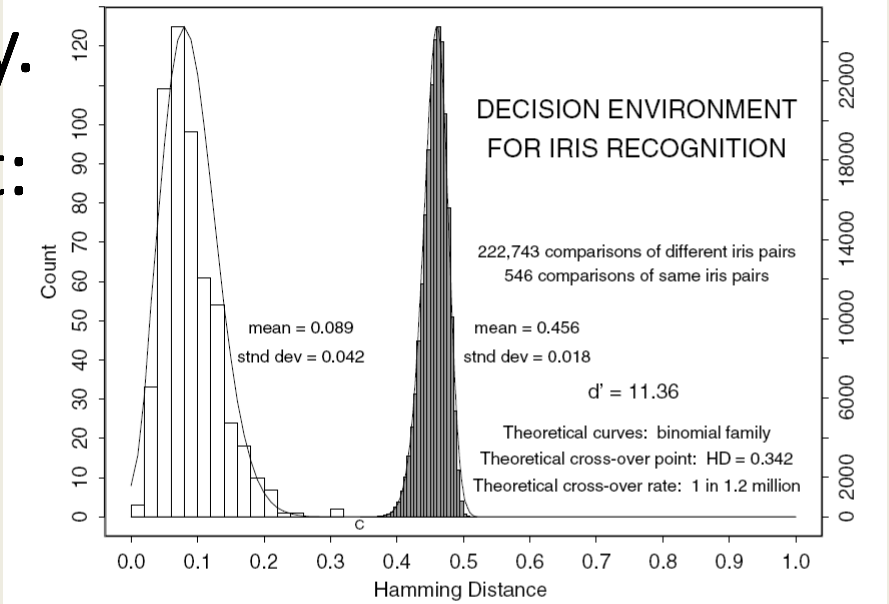


Fig. 1.1. Comparison of same and different iris pairs.

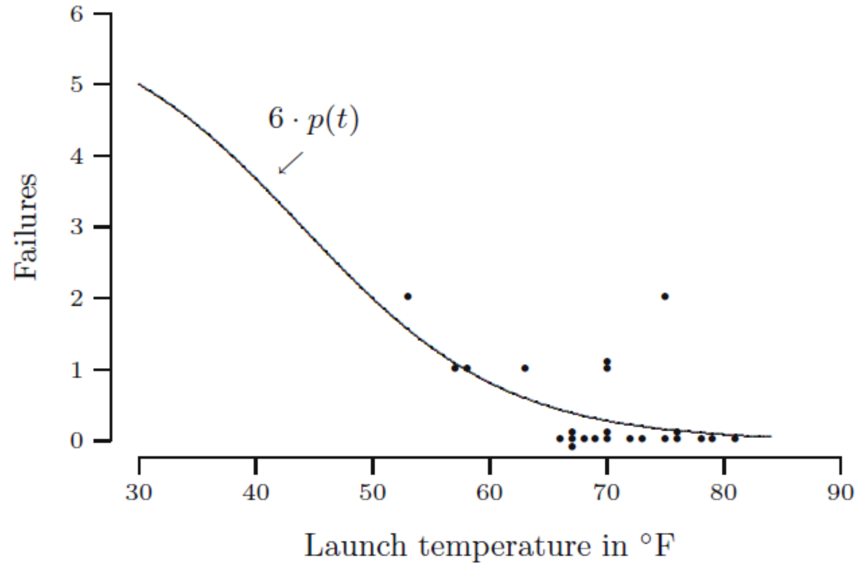
Source: J.Daugman. *Second IMA Conference on Image Processing: Mathematical Methods, Algorithms and Applications*, 2000. © Ellis Horwood Publishing Limited.

Motivace – zkáza raketoplánu Challenger

- 28/1/1986 minutu po startu explodoval.
- Příčina: výbuch hlavní nádrže díky plamenům vycházejících z boku pomocných raketových motorů.
- Ty jsou složené ze segmentů -> těsnění je o-kroužek a tmel -> má požadované vlastnosti při teplotách vyšších jak cca. 4 °C.
- V den startu byla teplota vzduchu -0,5°C (31 °F)!!!
- Challenger byla 24. mise. Byla známa data o poškozených o-kroužcích z předchozích letů (raketový motor je vyloven z oceánu).
- Každý raketový motor má 3 o-kroužky -> 6 dat za start.

Motivace – zkáza raketoplánu Challenger

- Efekt teploty z naměřených dat není zřejmý.



Source: based on data from Volume VI of the Report of the Presidential Commission on the space shuttle Challenger accident, Washington, DC, 1986.

Fig. 1.3. Space shuttle failure data of pre-*Challenger* missions and fitted model of expected number of failures per mission function.

Motivace – zkáza raketoplánu Challenger

- Pravděpodobnost $p(t)$, že při dané teplotě t selže jeden o-kroužek lze namodelovat tzv. binomickým rozdělením pravděpodobnosti.

$$p(t) = \frac{\mathrm{e}^{a+b\cdot t}}{1+\mathrm{e}^{a+b\cdot t}}$$

- Parametry a a b zvolíme tak, abychom dobře fitovali exp. data.
- Pak se dá spočítat p(-0,5°C) = 0,8178.
- Očekávaný počet poškozených o-kroužku je $6 \cdot p(t)$.

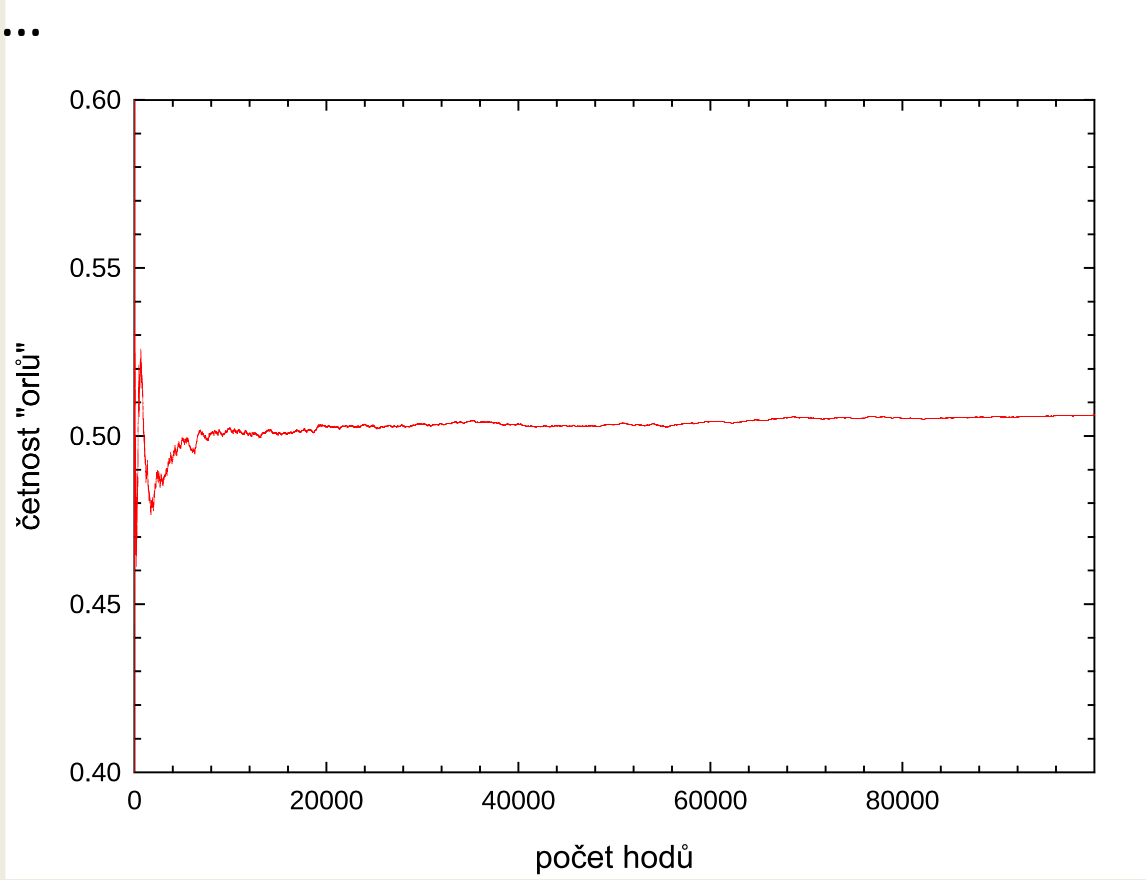
Pravděpodobnost × Statistika

- „Nematematické“ definice: pravděpodobnost je pohled, jak měřit (kvantifikovat) naší nejistotu pro jevy, které nemohu popsat nebo předvídat.
- „Pravděpodobnost“: teorie → data
  - Mám dobře definovaný problém a spočítám všechny možné výstupy (jejich četnosti) mého experimentu.
- „Statistika“: data → teorie
  - Mám sadu dat z experimentu a chci najít zákon/teorii/pravidla, kterým se data řídí
  - Zjistím parametry: chyba, odchylka
  - Testuji hypotézy

Pravděpodobnost

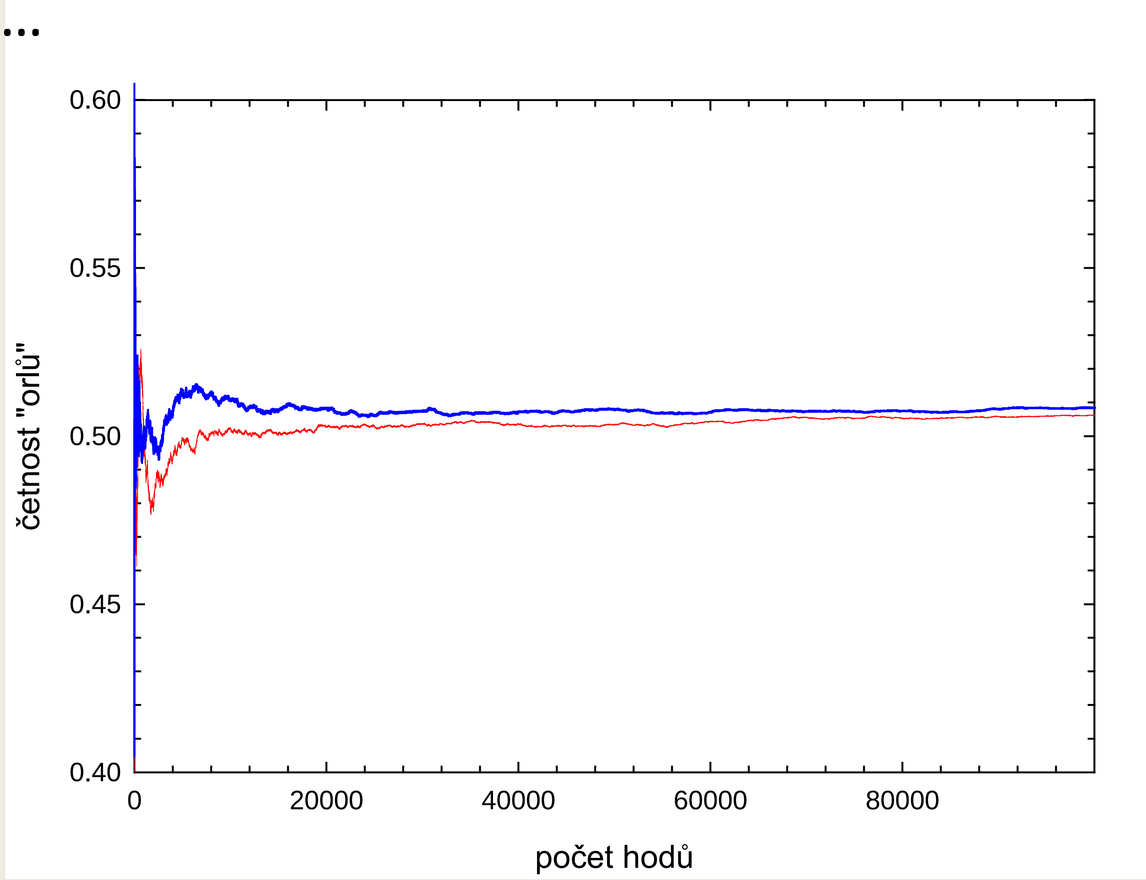
Pravděpodobnost

- Náhodný experiment (**pokus**) – házení mincí, kostkou, střílení na terč,...



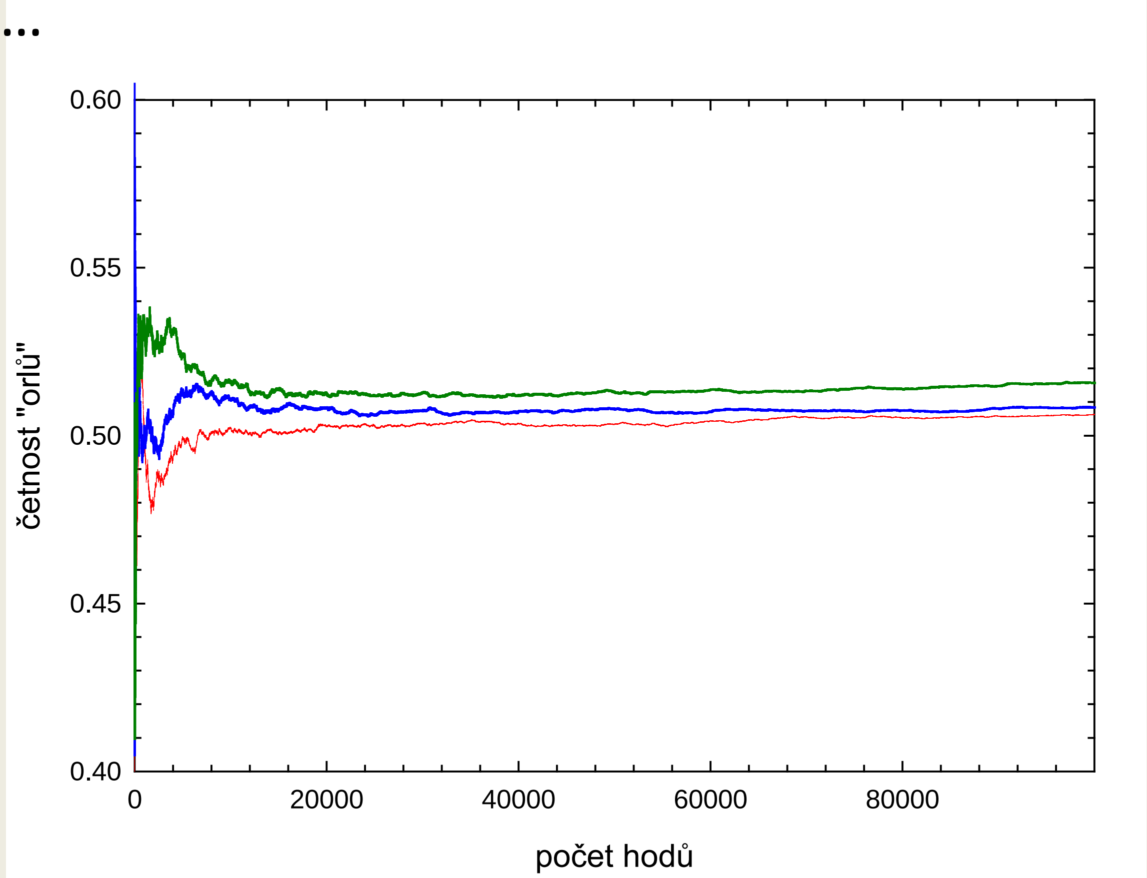
Pravděpodobnost

- Náhodný experiment (**pokus**) – házení mincí, kostkou, střílení na terč,…



Pravděpodobnost

- Náhodný experiment (**pokus**) – házení mincí, kostkou, střílení na terč,…



Pravděpodobnost

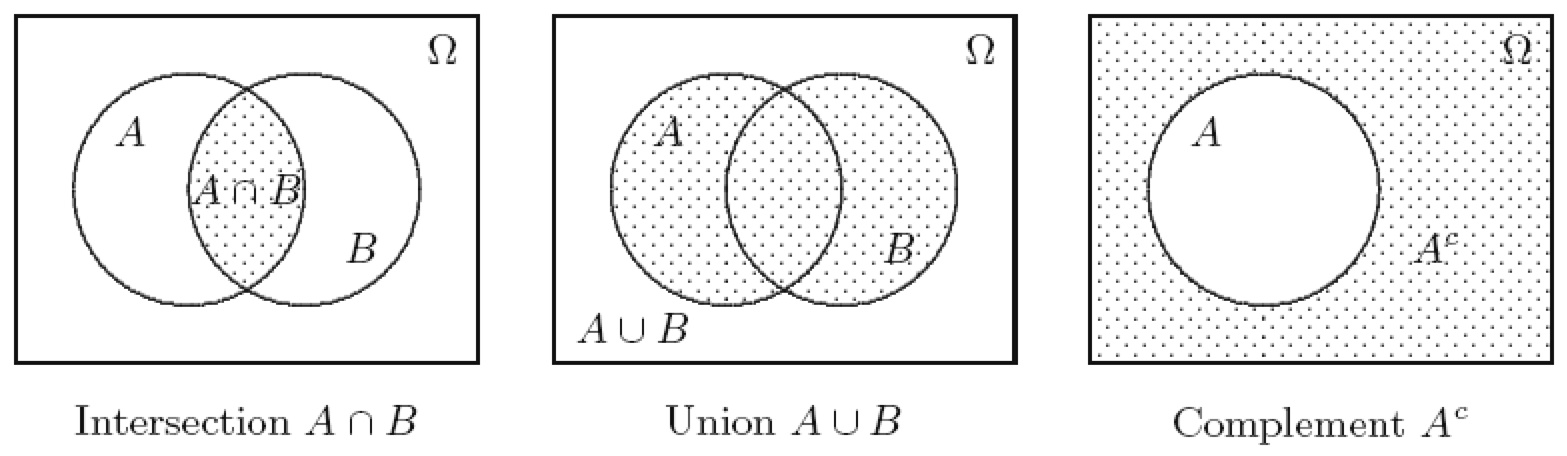
- Pozorujeme **jevy** – výsledky opakovaných měření nebo pozorování.
- Dělíme je na jevy:
  - **Jisté**: určitě nastane
  - **Nemožné**: určitě nenastane
  - **Náhodné**: nastane nebo nenastane
- Náhodný jev je důsledek působení mnoha náhodných příčin.
- Zkoumáme náhodné jevy, které mohou být pozorovány mnohokrát ne jen jednou.

Interpretace pravděpodobnosti

- Existuje několik různých definicí pravděpodobnosti. Zásadně se liší ve způsobu interpretace výsledků.
- Interpretace:
  - matematická (axiomatická, množiny)
  - empirická (relativní četnost náhodného jevu pro $n \to \infty$)
  - subjektivní (Bayesiánská)
  - objektivní (četnost m/n)

Pravděpodobnost - množiny

- Konečný prostor (množina) Z všech možných *elementárních jevů* (výsledků experimentu).
  - $Z = \{A_1, A_2, A_3, \ldots, A_n,\}$
- Náhodný jev je podmnožinou A množiny Z.
- Jistý jev – celá množina Z.
- Nemožný jev – prázdná množina Ø.



Pravděpodobnost - množiny

- Dva jevy A_i a A_j se vzájemně vylučují, pokud nenastanou ve stejném okamžiku při náhodném experimentu. (Př.: náhodně vybrané číslo nemůže být zároveň sudé i liché)
- $A_i \cap A_j = \emptyset$
- Pokud průnik každé dvojice A_i a A_j je prázdná množina, pak všechny A_i tvoří úplný systém, jejichž sjednocením je množina Z.
- Jev, který nelze dále rozložit (není sjednocením jiných jevů) se nazývá elementární jev. Sjednocením všech elementárních jevů je množina Z.
- DeMorganův zákon:

$$(A \cup B)^C = (A^C \cap B^C) \text{ a } (A \cap B)^C = (A^C \cup B^C)$$

Pravděpodobnost - axiomatická

- Axiomatická definice: Pro každý jev A existuje pravděpodobnost (číslo) P(A), pro nějž platí:
  - $P(A) \in [0, 1]$
  - $P(A_1 \cup A_2 \cup A_3, \dots, \cup A_n) = P(A_1) + P(A_2) + P(A_3) + \dots P(A_n)$, pokud se jevy A_i vzájemně **vylučují**
  - pravděpodobnost jistého jevu je rovna jedné, tj. P(Z) = 1
- Z axiomatické definice plynou následující vlastnosti:
  - $A_1 \in Z$, $A_2 \in Z$, $A_1 \in A_2$, pak $P(A_1) \leq P(A_2)$
  - $P(\bar{A}) = 1 - P(A)$
  - $P(\emptyset) = 0$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, pro případ, že jevy A a B se vzájemně **nevylučují**

Pravděpodobnost - axiomatická

- Problém opakujících se experimentů
- Hod mincí - panna × orel mohu opakovat *n*-krát.
- Každému hodu odpovídá $Z_i = \{p, o\}$
- $Z = Z_1 \times Z_2 \times Z_3 \ldots \times Z_n = \{p, o\} \times \{p, o\} \times \{p, o\} \ldots \times \{p, o\}$.
- Mohu vybrat *n* náhodných jevů s pravděpodobností P_n. Pak pravděpodobnost, že *n* náhodných jevů nastane současně je:

  $$P((A_1, A_2, A_3, \ldots, A_n)) = P_1 \cdot P_2 \cdot P_3 \cdot \ldots \cdot P_n$$

Pravděpodobnost - axiomatická

- Množina všech možných jevů Z má nekonečně mnoho prvků.
- $Z = \{A_1, A_2, A_3, \ldots\}$.
- Platí stejná definice jako pro případ konečné množiny Z jen máme nekonečně mnoho prvků.
- Např. při výpočtu pravděpodobnosti nějaké jevu, který nastane třeba při nekonečně mnoha opakováních hodem mincí, je třeba použít limity pro $n \to \infty$.

Pravděpodobnost - objektivní

- Množina Z bude obsahovat konečný počet *n* elementárních jevů, které jsou stejně možné.
- Jev A se skládá z *m* elementárních jevů. Pak pravděpodobnost jevu A je:

$$P(A) = \frac{m}{n}$$

- Je to poměr příznivých jevů, ke všem možným.
- Je definována „objektivně“ na základě experimentu (třeba házení kostkou).
- Neplatí pro spojitou proměnnou, např. měření náhodného úhlu.
- Geometrická definice je grafickou obdobou (šipky v terči).

Pravděpodobnost - empirická

- Vykonám experiment *n* krát naprosto stejnými pokusy.
- Náhodný jev A nastane *m* krát.
- Pravděpodobnost jevu A je dána:

$$P(A) = \lim_{n \to \infty} \frac{m}{n}$$

- Nazývá se to také jako statistická pravděpodobnost.
- Zajímá mě **četnost** s jakou nastane jev A celkem *m* krát, pokud pokus opakuji nezávisle *n* krát.
- Počet opakování *n* musí být velké číslo a pak je pravděpodobnost rovna objektivní definici pravděpodobnosti.
- Realita je komplikovaná: kolik *n* musím vykonat? V realitě striktní matematická limita neexistuje.

Pravděpodobnost - subjektivní

- Pravděpodobnost jako míra (stupeň) věrohodnosti (spolehlivosti, víry) nějakého tvrzení.
- Pravděpodobnost, že hypotéza A platí je rovna míře naší víry ve správnost hypotézy A.
- P(A) je tedy míra našich informací/znalostí I o problému. Tedy P(A) je podmíněno znalostí I – Bayesův teorém.
- P(A)=$(\mathrm{A}|\mathrm{I})$.
- V reálném životě lidi uvažují podle subjektivní pravděpodobnosti
  – „Zítra bude velmi pravděpodobně pršet!“ Budeme tvrzení „silněji“ věřit, pokud už prší celý týden a navíc předpověď počasí to očekává.
- Je nutné znát informace před tím, než vyslovíme pravděpodobnost.

Pravděpodobnost - interpretace

- Různé lidské činnosti vyžadují různé interpretace pravděpodobnosti.
- Částicový fyzik v CERN se bude spoléhat na empirickou interpretaci a pravděpodobnost určí z četnosti měřených dat.
- Obchodník pracující s konečným počtem položek bude spoléhat na objektivní pravděpodobnost, když bude chtít znát pravděpodobnost výskytu reklamací.

Podmíněná pravděpodobnost

- Náhodný jev A nastane za splnění určitých podmínek – nastal již nějaký náhodný jev B.
- Podmíněná pravděpodobnost jevu A vůči jevu B.
- Značíme $P(A|B)$.
- Házení dvěma kostkami. Jev B: sudé číslo na druhé kostce; jev A: součet čísel na obou kostkách je 5. Ř.: P(B)=18/36 -> P(A|B) = 2/18 = 1/9. P(A∩B) = 2/36 = 1/18.

In the previous chapter we encountered the events L, "born in a long month," and R, "born in a month with the letter r." Their probabilities are easy to compute: since $L = \{\text{Jan, Mar, May, Jul, Aug, Oct, Dec}\}$ and $R = \{\text{Jan, Feb, Mar, Apr, Sep, Oct, Nov, Dec}\}$, one finds

$$P(L) = \frac{7}{12} \quad \text{and} \quad P(R) = \frac{8}{12}.$$

Now suppose that it is *known* about the person we meet in the street that he was born in a "long month," and we wonder whether he was born in a "month with the letter r." The information given excludes five outcomes of our sample space: it cannot be February, April, June, September, or November. Seven possible outcomes are left, of which only four—those in $R \cap L = \{\text{Jan, Mar, Oct, Dec}\}$—are favorable, so we reassess the probability as 4/7. We call this the *conditional probability of R given L*, and we write:

$$P(R \,|\, L) = \frac{4}{7}.$$

This is not the same as $P(R \cap L)$, which is 1/3. Also note that $P(R \,|\, L)$ is the proportion that $P(R \cap L)$ is of $P(L)$.

Podmíněná pravděpodobnost

- Jak spočítat podmíněnou pravděpodobnost:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Pravděpodobnost jevu B musí být nenulová.
- Hledáme tedy takovou část pravděpodobnosti jevu A, kdy zároveň nastává jev B.
- Platí: $P(\bar{A}|B) = 1 - P(A|B)$. Pravidlo pravděpodobnosti komplementárních jevů platí i pro podmíněnou pravděpodobnost.

Podmíněná pravděpodobnost

- Pravidlo násobení pravděpodobností

  $$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(B)$$

- Mám rovnici pro průnik jevů, které se vzájemně nevylučují. V tomto případě nemohu použít násobení pravděpodobností P(A) a P(B).
- Jak spočítat pravděpodobnost jevů, které **nenastávajících současně**, tedy se vzájemně vylučují? *Srovnej s nezávislými jevy.*
- Př.: jaká je pravděpodobnost, že i = {1, 2, 3, 4, … n} náhodně vybraných lidí bude mít narozeniny v rozdílných dnech?
- Pro i = 1 je $P(B_1) = 1$, je to jev jistý
- Pro i = 2 je $P(B_2) = 1 - (1/365)$
- Pro i = 3 je pravděpodobnost jevu B_3 průnikem jevu B_2 a jevu A_3 „3. osoba má narozeniny v jiný den než dvě předchozí“. $P(B_3) = P(A_3 \cap B_2) = P(A_3|B_2) \cdot P(B_2) = (1 - (2/365)) \cdot (1 - (1/365))$
- Pro i = n je $B_n = A_n \cap B_{n-1}$ a $P(B_n) = P(A_n|B_{n-1}) \cdot P(B_{n-1}) = (1-(n-1/365)) \cdot P(B_{n-1})$

Podmíněná pravděpodobnost

- Např. pro již 23 náhodně vybraných lidí je pravděpodobnost, že data narození budou v různých dnech ≈ 50%.

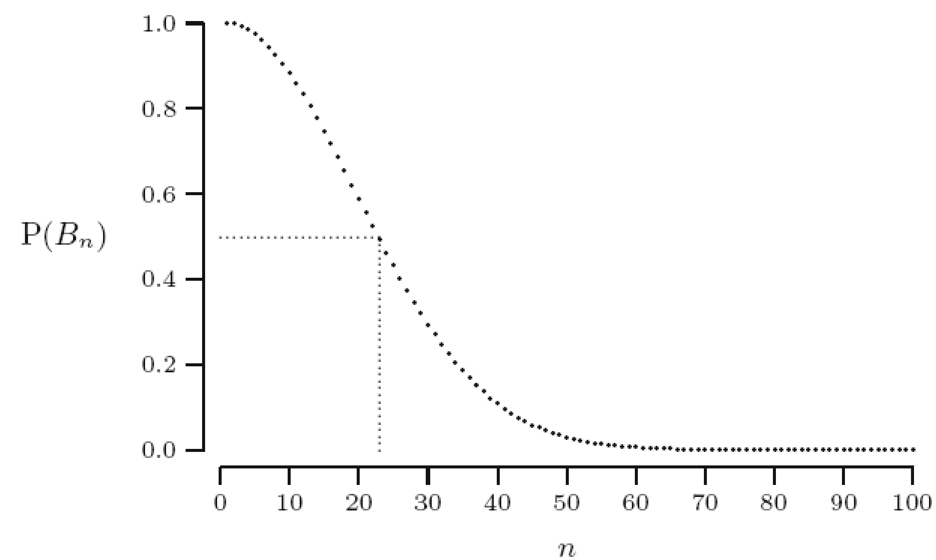


Fig. 3.1. The probability $P(B_n)$ of no coincident birthdays for $n = 1, \ldots, 100$.

Zákon celkové pravděpodobnosti

- Nechť C_1, C_2, C_3,… , C_n jsou náhodné jevy, které nejsou slučitelné a platí $C_1 \cup C_2 \cup C_3 \cup \ldots \cup C_n = Z$. Pak pravděpodobnost libovolného náhodného jevu A může být vyjádřena následovně:

$$P(A) = P(A|C_1)P(C_1) + P(A|C_2)P(C_2) + P(A|C_3)P(C_3) + \ldots + P(A|C_n)P(C_n)$$

- Použili jsme pravidla pro násobení jevů, které se vzájemně nevylučují

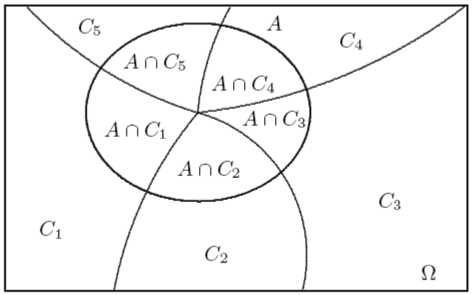


Fig. 3.2. The law of total probability (illustration for $m = 5$).

Bayesovo pravidlo

- Nechť $C_1, C_2, C_3, \ldots , C_n$ jsou náhodné jevy, které nejsou slučitelné a platí $C_1 \cup C_2 \cup C_3 \cup \ldots \cup C_n = Z$. Pak pro podmíněnou pravděpodobnost C_i za předpokladu, že nastal náhodný jev A platí:

  $P(C_i|A) = \dfrac{P(A|C_i)\, P(C_i)}{P(A)}$, kde P(A) je dáno zákonem celkové pravděpodobnosti.

Nezávislost jevů

- Jev A nezávisí na tom, zda jev B nastane nebo nenastane, pak tyto jevy jsou **nezávislé**.
- Definice nezávislosti jevů:
- Je-li podmíněná pravděpodobnost jevu A za podmínky jevu B rovna nepodmíněné pravděpodobnosti jevu A, pak jevy A a B jsou nezávislé:

$$\mathrm{P(A|B)} = \mathrm{P(A)}$$

- Je to oboustranná vlastnost:

$$\mathrm{P(B|A)} = \mathrm{P(B)}$$

- Příklad: vybíráme dva jevy z množiny Z tak, že první jev „vrátíme" zpět do Z. Pak pravděpodobnost, že druhý vybraný jev bude mít určitou vlastnost, nezávisí na výběru prvního jevu. Oba jevy jsou teda nezávislé.
- Lze to zobecnit na *n* jevů.
- Z pravidla o podmíněné pravděpodobnosti platí, že pokud jev A není závislý na jevu B, tak
  - jev B není závislý na jevu A,
  - jev $\overline{\mathrm{A}}$ není závislý na jevu B.
- Tedy nezávislost je vzájemná vlastnost.

Nezávislost jevů

- Test nezávislosti dvou jevů A a B. Dva jevy A a B jsou vzájemně nezávislé, pokud platí aspoň jedna podmínka:
  - $P(A|B) = P(A)$
  - $P(B|A) = P(B)$
  - $P(A \cap B) = P(A) \cdot P(B)$
  - Výše uvedené platí i pokud jevy A nebo B jsou nahrazeny za jejich komplementy

Nezávislost jevů

- V případě dvou a více jevů je podmínka nezávislosti následující:
- Jevy $A_1, A_2, A_3, \ldots, A_n$ nazýváme nezávislé, pokud je splněna podmínka

$$P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot \ldots \cdot P(A_n)$$

- Podmínka také platí pokud jakýkoliv jev A_1 až A_n je nahrazen jeho komplementem.
- Je zřejmé, že musíme otestovat 2^n rovnic, abychom ověřili, že všechny jevy jsou nezávislé.

Závislé jevy

- Nastoupení jevu A je závislé na výsledcích předchozích experimentů.
- Typický příklad je vybírání čísel, bez jejich vracení do experimentu. Tedy vybraná čísla už nemohou být znovu náhodně vybraná.

Náhodná proměnná

Náhodná proměnná

- Prostor všech jevů Z spojený s nějakým experimentem + pravděpodobnostní funkce definovaná pro všechny možné jevy → kompletní pravděpodobnostní popis experimentu.
- Často nás zajímá jen určitá vlastnost nebo rys kompletního pravděpodobnostního popisu nebo náhodné jevy potřebujeme popsat nějakými hodnotami ať číselnými nebo slovními.
- Tzv. **náhodná proměnná (veličina)** popisuje právě nějaký konkrétní rys náhodného experimentu.
- Náhodnou veličinu dělíme na: uspořádaná × neuspořádaná.
- **Uspořádaná** se dělí na: kvantitativní × kvalitativní.
- **Kvantitativní** se dělí na: diskrétní × spojitou.

Náhodná proměnná

	ω_1					
ω_2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

obvious choice of the sample space is

$$\begin{aligned}\Omega &= \{(\omega_1, \omega_2) : \omega_1, \omega_2 \in \{1, 2, \ldots, 6\}\} \\ &= \{(1,1), (1,2), \ldots, (1,6), (2,1), \ldots, (6,5), (6,6)\}.\end{aligned}$$

However, as players of the game, we are *only* interested in the sum of the outcomes of the two throws, i.e., in the value of the function $S : \Omega \to \mathbb{R}$, given by

$$S(\omega_1, \omega_2) = \omega_1 + \omega_2 \quad \text{for } (\omega_1, \omega_2) \in \Omega.$$

diagonal. We denote the event that the function S attains the value k by $\{S = k\}$, which is an abbreviation of "the subset of those $\omega = (\omega_1, \omega_2) \in \Omega$ for which $S(\omega_1, \omega_2) = \omega_1 + \omega_2 = k$," i.e.,

$$\{S = k\} = \{(\omega_1, \omega_2) \in \Omega : S(\omega_1, \omega_2) = k\}.$$

Tedy zavádíme funkci S, která nabývá hodnot *k* podle daného předpisu.

Diskrétní náhodná proměnná

- **Definice:** nechť Z je prostor všech možných náhodných jevů. Diskrétní náhodná proměnná je funkce X : Z → **R**, která nabývá konečného počtu hodnot a_1, a_2, a_3,…, a_n nebo nekonečného počtu hodnot a_1, a_2, a_3,…
- Jinými slovy funkce mi transformuje jednu množinu náhodných jevů **Z** na jinou množinu náhodných jevů **Z´**, které nás zajímají – viz tabulka na straně 44.
- Na druhou stranu je nutné spočítat pravděpodobnost, že nastanou jevy popsané funkcí X (zjistit, jak je pravděpodobnost náhodného jevu z množiny **Z´** rozložena mezi možné hodnoty funkce X) = najít rozdělení pravděpodobnosti (pravděpodobnostní distribuci) funkce X.

Pravděpodobnostní funkce

- Jakmile je definována náhodná veličina X, tak už nepotřebuji množinu **Z**.
- Stačí znát možné hodnoty funkce X a jejich odpovídající pravděpodobnosti.
- Tyto informace jsou obsaženy v **pravděpodobnostní funkci** náhodné veličiny X.

DEFINITION. The *probability mass function* p of a discrete random variable X is the function $p : \mathbb{R} \to [0,1]$, defined by

$$p(a) = \mathrm{P}(X = a) \quad \text{for } -\infty < a < \infty.$$

Pravděpodobnostní funkce

- Příklad: použiji experiment s hodem dvěma kostkami viz tabulka na str. 44
  - a = {2, 3, 4,…, 12} → p(a) = { 1/36, 2/36, 3/36,…, 1/36}
- $p(a_i) > 0$
- $p(a_1) + p(a_2) + p(a_3) + \ldots = 1$
- p(a) = 0 pokud náhodná veličina X nenabývá hodnoty *a*.

Distribuční funkce

- Pravděpodobnostní funkce nemůže být aplikována na spojitou náhodnou veličinu.
- Zavádíme tzv. **distribuční funkci** jak pro diskrétní, tak spojitou náhodnou veličinu.

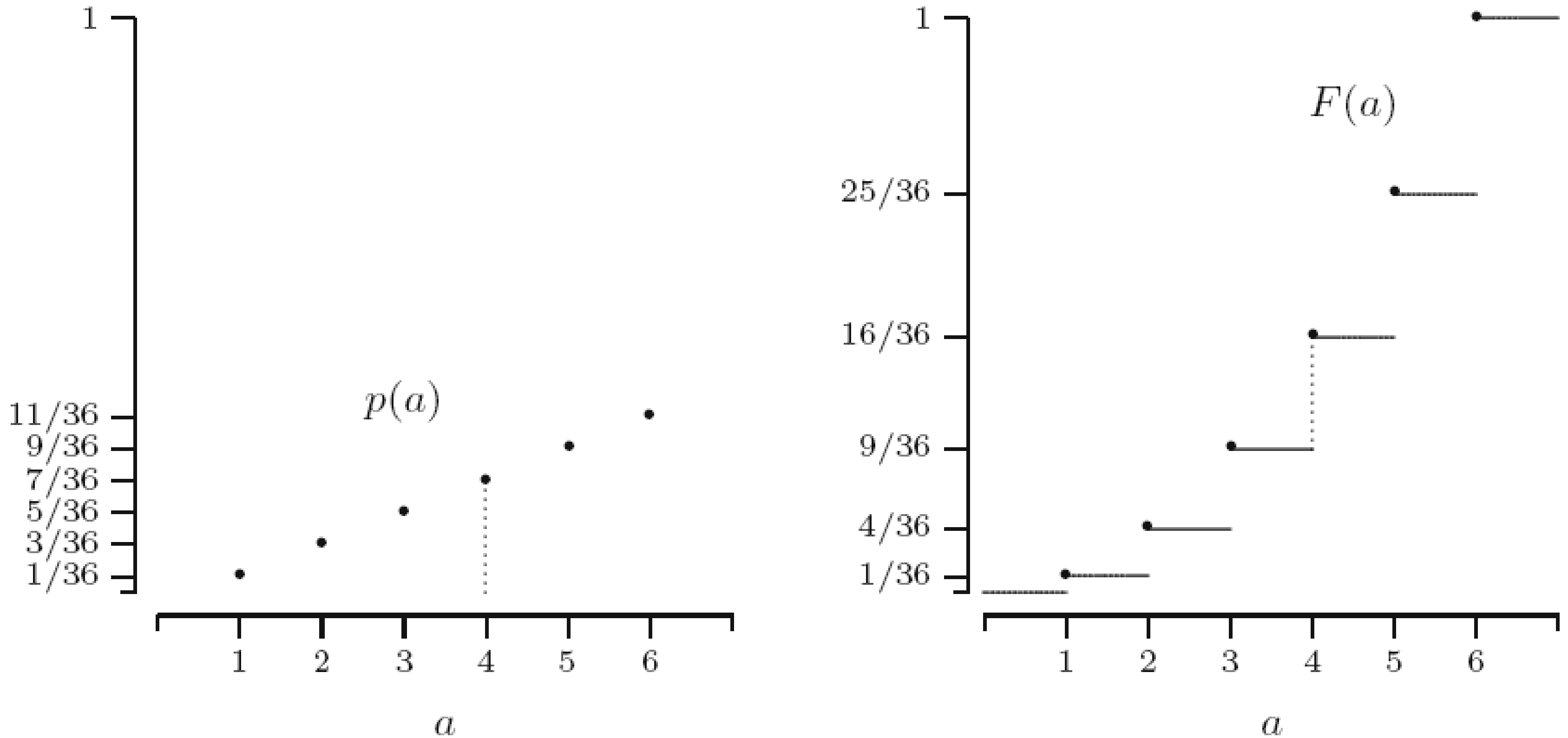
DEFINITION. The *distribution function* F of a random variable X is the function $F : \mathbb{R} \to [0,1]$, defined by

$$F(a) = \mathrm{P}(X \leq a) \quad \text{for } -\infty < a < \infty.$$

Distribuční funkce

	ω_1					
ω_2	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

- Př.: hod dvěma kostkami; náhodná veličina X je maximální číslo, jaké mi padne; a={1, 2, 3, 4, 5, 6}



Distribuční funkce

- Distribuční funkce F(a) může být vyjádřena pomocí pravděpodobnostní funkce p(a) – platí to i obráceně.

$$p(a_i) > 0, \quad p(a_1) + p(a_2) + \cdots = 1,$$

$$F(a) = \sum_{a_i \leq a} p(a_i).$$

- Funkce F(a) je zprava spojitá funkce.
- Velikost (výška) skoku (nespojitost) F(a) zleva v bodě a_i náhodné proměnné X odpovídá právě hodnotě p(a_i).

Distribuční funkce

- Pravděpodobnostní i distribuční funkce diskrétní náhodné veličiny mi jednoznačně popisují rozložení pravděpodobnosti pro všechny náhodně veličiny.
- Tři důležité vlastnosti F(a):

1. For $a \leq b$ one has that $F(a) \leq F(b)$. This property is an immediate consequence of the fact that $a \leq b$ implies that the event $\{X \leq a\}$ is contained in the event $\{X \leq b\}$.
2. Since $F(a)$ is a probability, the value of the distribution function is always between 0 and 1. Moreover,

$$\lim_{a \to +\infty} F(a) = \lim_{a \to +\infty} \mathrm{P}(X \leq a) = 1$$
$$\lim_{a \to -\infty} F(a) = \lim_{a \to -\infty} \mathrm{P}(X \leq a) = 0.$$

3. F is right-continuous, i.e., one has

$$\lim_{\varepsilon \downarrow 0} F(a + \varepsilon) = F(a).$$