

# Pravděpodobnost (pracovní verze)

# 1. Definice pojmů

Jednoduchý/náhodný pokus (*simple experiment*)

- *Akt vedoucí k jednomu výsledku - např. hod kostkou, zatočení ruletou, vytažení karty z balíčku, výběr osoby na ulici*
- *Výsledkem je výskyt jednoduchého jevu/události*

Jednoduchý jev (*elementary event*)

- *člen základní množiny*
- *výsledek jednoduchého pokusu - např. hodnota 1 na kostce, 0 na ruletě, sedmička srdcová, modrooká paní*

Jev/třída jevů (*event, event class*)

- *sada jednoduchých jevů - např. lichá čísla, „srdce“, „piky“*

Základní množina/prostor (*S*) (*Sample space*)

- *sada všech jednoduchých jevů*

Spojené jevy (*joint events*) – *nastávají když výsledek pokusu spadá pod jevy A („srdce“) i B („král“) např. „srdcový král“, popřípadě A nebo B např. „srdce“ nebo „král“*

- *Průnik ( $\cap$ ) (*intersection*) – např. průnik jevů  $A$  a  $B = A \cap B$  nebo-li  $A$  a  $B$*
- *současné nastání dvou nebo více jevů*
- *Sjednocení ( $\cup$ ) (*union*) - např. sjednocení jevů  $= A \cup B$  nebo-li  $A$  nebo  $B$*
- *sečtení dvou nebo více jednoduchých jevů bez průniku*

Doplněk ( $\sim A$ ) (*complement*)

- *doplňkem jevu  $A$  je sada všech zbývajících jevů z  $S$*

Vzájemně vylučující se/neslučitelné (*mutual exclusive*) jevy

- *nemohou nastat současně, jejich  $\cap = 0$*

Vyčerpávající (*exhaustive*) jevy

- *jevy vyplňují celý  $S$ , jejich  $\cup = S$*

Pravděpodobnost ( $p$ ) (*probability*)

- *míra jistoty nastání každého jevu ze základního prostoru - např. pravděpodobnost že padne 1 na kostce*

Podmíněná pravděpodobnost ( $p(A|B)$ ) (*conditional probability*)

- *pravděpodobnost výskytu jevu  $A$  za předpokladu, že zároveň nastane jev  $B$  –*
- *např. experiment: hod dvěma kostkami, událost: součet hodnot, otázka: jaká je pravděpodobnost výskytu události 4 když na jedné kostce padne 5?*

Statistická nezávislost (*statistical independence*)

- *nepodmíněná pravděpodobnost jevu  $A$  a podmíněná pravděpodobnost jevu  $A$  stane-li se zároveň  $B$  jsou si rovny*
- *tj.  $p(A) = p(A | B)$*
- *nebo když  $p(A \cap B) = p(A) \cdot p(B)$*

# Definice pravděpodobnosti

$p(A)$  = pravděpodobnost jevu  $A$

- za předpokladu, že máme konečný počet jednoduchých jevů v  $S$  a každý jednoduchý jev z  $S$  má stejnou pravděpodobnost nastání pak platí, že  $p(A) = n(A) / n(S)$ 
  - Př. Krabíčka 10 kuliček – 5 bílých, 3 červené, 2 černé  
experiment: bez dívání vybereme jednu kuličku  
jednoduchý jev: jedna konkrétní kulička, máme tedy 10 jednoduchých jevů tvořících  $S$ , jevy jsou vzájemně se vylučující a každý má pravděpodobnost  $1/10$   
zajímají nás události: „bílá“, „červená“, „černá“

Otázka: Jaká je  $p$  že vytáhnu „červenou“?

$$p(\text{červená}) = n(\text{červená}) / n(S) = 3 / 10 = .30$$

$$\text{Podobně } p(\text{bílá}) = .50 \text{ a } p(\text{černá}) = .20$$

$$p(\text{červená a bílá}) = 0$$

Pravděpodobnostní funkce přisuzuje každému jevu  $A$  z  $S$  číslo  $p(A)$ , pravděpodobnost jevu  $A$ , tak že platí („axiomy“/zákony):

1.  $p(A) \geq 0$  pro každé  $A$
2.  $p(S) = 1$
3. Pokud  $(A_1, A_2, \dots, A_n)$  jsou vzájemně neslučitelné,  
pak  $p(A_1 \cup A_2) = p(A_1) + p(A_2)$

# Pravidla pravděpodobnosti

1.  $p(\sim A) = 1 - p(A)$  („doplňková pravděpodobnost“)
  - Příklad. Jaká je pravděpodobnost že vyberu „ne červenou“ kuličku tj. jinou než „červenou“?
    - $p(\sim \text{červená}) = 1 - p(\text{červená}) = .7$
  
2.  $0 \leq p(A) \leq 1$  („rozsah pravděpodobnosti“)

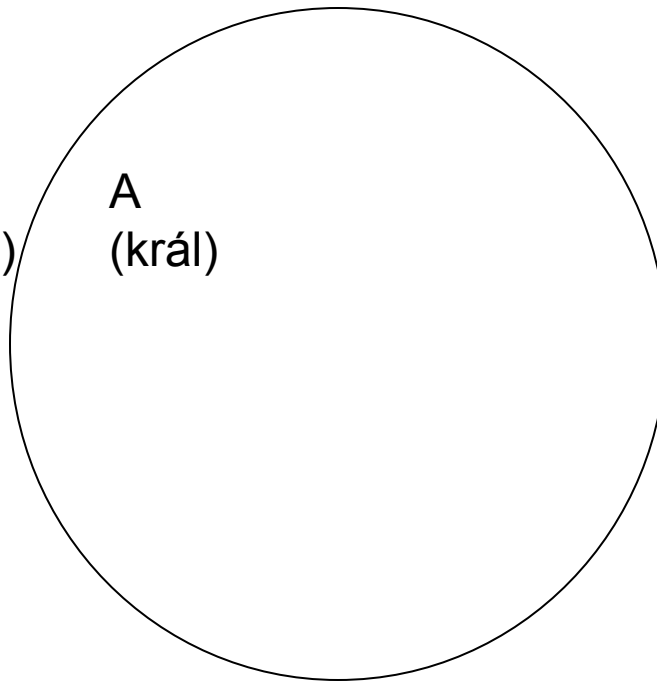
(důkaz: pokud by nějaký jev měl  $p$  větší než 1 pak by podle pravidla 1 měl doplněk jevu  $p$  zápornou a to by odporovalo axiomu 1)
  
3.  $p(\emptyset) = 0$ , pro jakékoli  $S$  („nemožný jev“)
  - Příklad. Jaká je pravděpodobnost že vyberu „bezbarvou“ kuličku tj. jinou než „červenou“ nebo „bílou“ nebo „černou“?
    - $p(\text{bez barvy}) = 0$
  
4.  $p(A \cup B) = p(A) + p(B) - p(A \cap B)$  (tzv. „nebo“ pravidlo)
  - Příklad. Balíček 52 karet. Jaká je pravděpodobnost „krále“ nebo „srdce“?
    - $P(\text{král}) = 1/13$ ,  $p(\text{srdce}) = 1/4$ ,  $p(\text{král} \cap \text{srdce}) = 1/52$  (jeden z králů je srdcový)
    - $p(\text{král nebo srdce}) = 1/13 + 1/4 - 1/52 = 16/52 = 4/13$
  - Speciální případ: když jsou jevy vzájemně se vylučující, pak  $p(A \cap B) = 0$  a proto  $p(A \cup B) = p(A) + p(B)$ 
    - Příklad.  $p(\text{červená nebo bílá}) = .30 + .50 = .80$
  
5. Pokud  $A, \dots, L$  tvoří segmenty  $S$ , pak  $p(A \cup \dots \cup L) = p(A) + \dots + p(L) = 1$ 
  - Pokud jsou jevy  $A$  až  $L$  vylučující se a vyčerpávající, pak tvoří celý prostor  $S$  a součet jejich pravděpodobností musí být 1
  - Příklad.  $p(\text{červená nebo bílá nebo černá kulička}) = .30 + .20 + .50 = 1.00$

# Ilustrace: doplněk

S(52)

$\sim A$   
(nekrál)

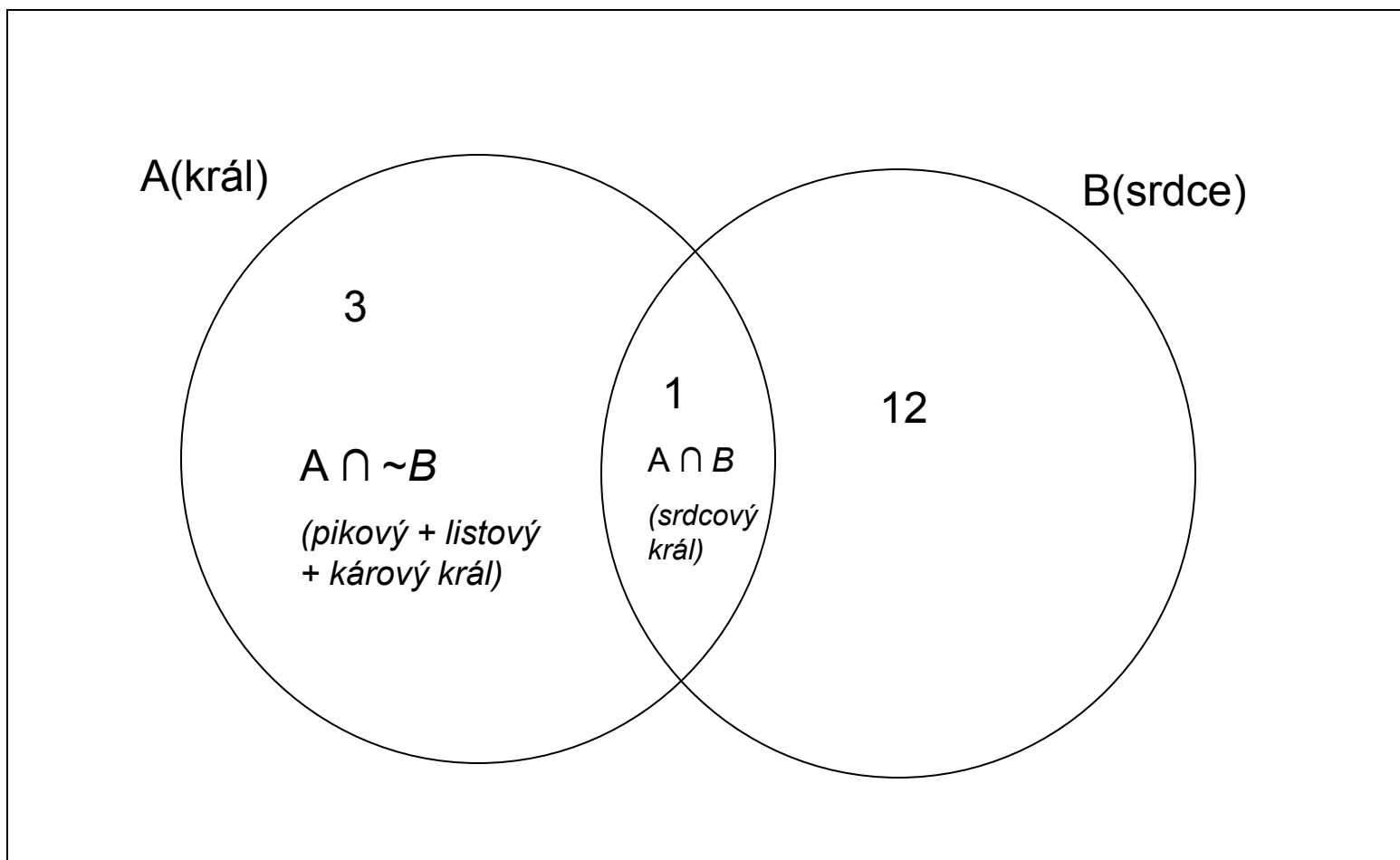
$A$   
(král)



# Průnik a sjednocení

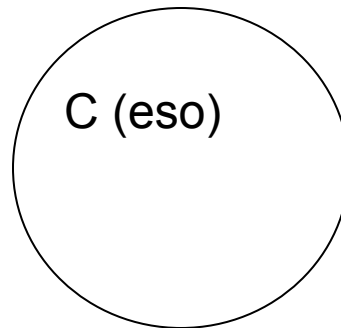
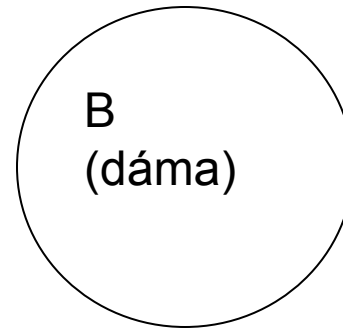
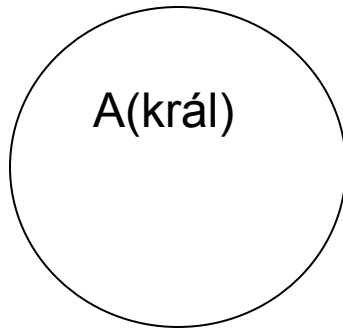
$$A \cup B = A + B - A \cap B = 13 + 4 - 1 = 16/52$$

S(52 karet)



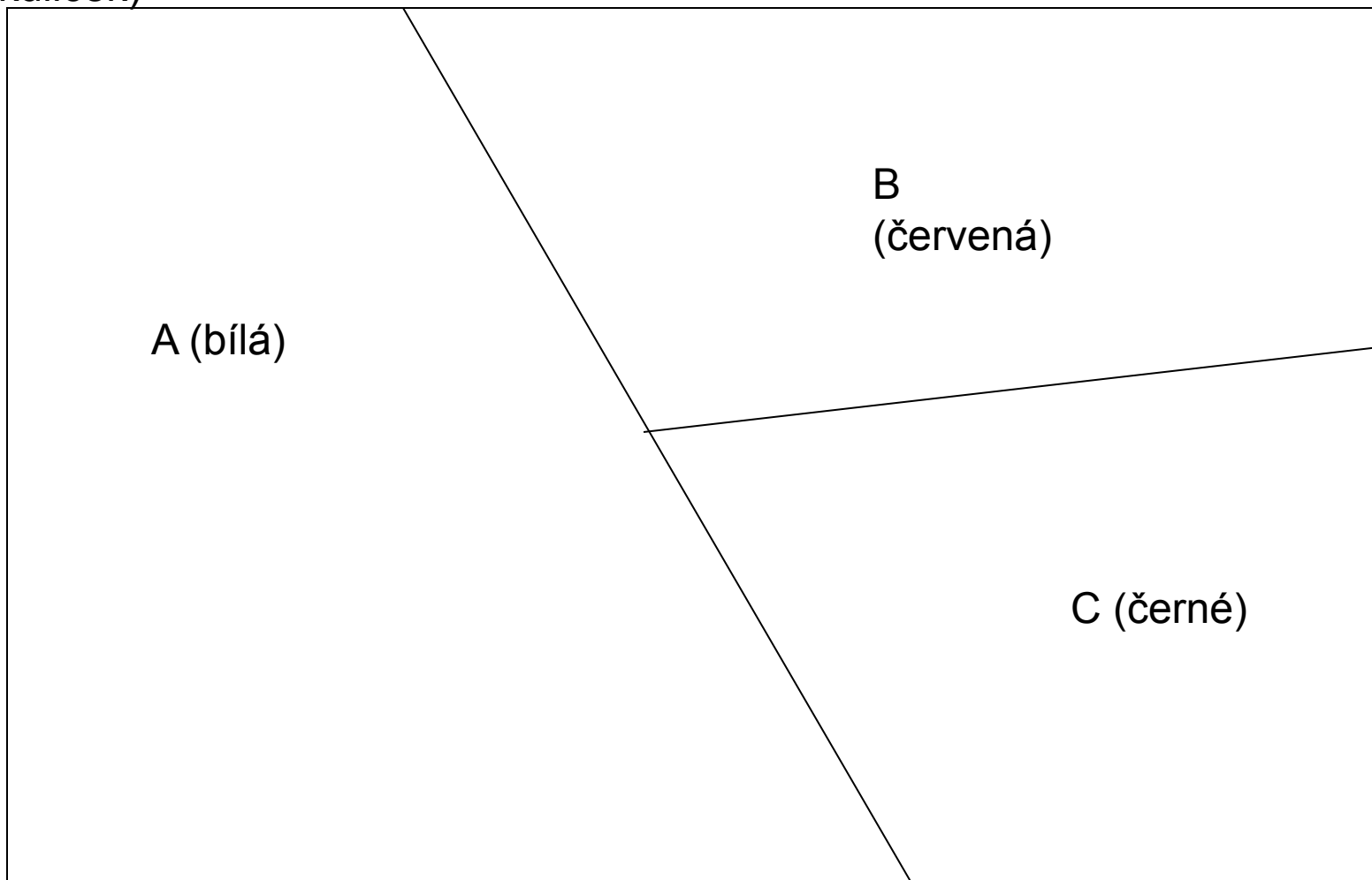
# A, B a C jsou vzájemně se vylučující jevy

S(52)



A, B a C jsou  
vzájemně se vylučující a vyčerpávající jevy

S(10 kuliček)





# Podmíněná pravděpodobnost

- $p(A | B)$  nebo-li  $p(A \text{ když } B) = n(A \cap B) / n(B)$   
 $= n(A \cap B) / n(S) / n(B) / n(S)$   
 $= p(A \cap B) / p(B)$
- Př. Výběr vzorku dětí ve škole limitujeme pouze na dívky. Vznikne nám nový základní prostor  $S_{\text{nový}}$  zahrnující pouze část jednoduchých jevů z původního  $S$ . V novém prostoru  $S_{\text{nový}}$  zahrnujícím pouze dívky jaká je pravděpodobnost výběru levorukého člověka? Přičemž víme že 51% studentů je dívek, 35% studentů je levorukých, a 10% studentů jsou levoruké dívky.

Otázka:  $p(\text{levoruký, když dívka}) = \text{počet levorukých dívek} / \text{celkový počet dívek}$

$p(\text{levoruká a dívka}) = \text{počet levorukých dívek} / \text{celkový počet studentů} = .10$

$p(\text{dívka}) = \text{celkový počet dívek} / \text{celkový počet studentů} = .51$

$p(\text{levoruký, když dívka}) = p(\text{levoruký a dívka}) / p(\text{dívka}) = .10 / .51 = .196$

$p(\text{dívka, když levoruký}) = .10 / .35 = .29$

# Statistická nezávislost

- Výskyty jevů A a B jsou statisticky nezávislé pokud výskyt jevu A neovlivňuje výskyt jevu B a opačně
  - Tedy když pravděpodobnost výskytu jevu A a pravděpodobnost výskytu jevu A když B jsou si rovny – tj.  $p(A) = p(A | B)$ 
    - př.  $p(\text{dívka}) = .51$   
 $P(\text{dívka, když levoruký}) = .29$   
Výsledek:  $.51 \neq .29$ , proto jevy nejsou nezávislé (souvisí spolu)
  - nebo alternativně když  $p(A \cap B) = p(A) * p(B)$ 
    - př.  $p(\text{dívka a levoruký}) = .10$   
 $p(\text{dívka}) = .51$   
 $p(\text{levoruký}) = .35$   
výsledek:  $.10 \neq .51 * .35$ , proto jevy nejsou nezávislé

## **Pozor! Nezávislost $\neq$ vzájemná vylučnost (neslučitelnost)**

(Pro nadšence matematický důkaz: pokud A a B jsou vzájemně neslučitelné, pak  $p(A \cap B) = 0$ . Pokud by A a B byly zároveň nezávislé, pak  $p(A \cap B) = p(A) * p(B)$ , a protože  $p(A \cap B) = 0$ , tak to nemůže být pravda pokud  $p(A)$  nebo  $p(B)$  není nula)

(Pro nadšence intuitivní důkaz: Předpokládejme že všichni muži jsou buďto „plešatíci“ nebo „hustovlasí“ (dva vzájemně se vylučující jevy). Pokud by „plešatíci“ a „hustovlasí“ byli na sobě nezávislé jevy, pak by mezi „plešatíci“ byla stejná proporce „hustovlasých“ jako je celková proporce „hustovlasých“ mezi muži, tedy  $p(A | B) = p(A)$ , což je těžké si představit ☺)

# Prezentace spojených jevů v tabulkách

- Příklad. Základní množina  $S$  = studenti kampusu v Bohunicích, každý student je buďto dívka nebo chlapec, a odpověděl buďto „ano“ nebo „ne“ na otázku zda chodí každé ráno rád do školy.

Chlapec	(chlapec a „ano“)	(chlapec a „ne“)
Dívka	(dívka a „ano“)	(dívka a „ne“)
	Ano	Ne

- Necht' v tomto kampusu pravděpodobnosti vybrat chlapce je .55 a dívku .45, a pravděpodobnost odpovědi ano je .40 a „ne“ .60.

$P(\text{chlapec}) = .55$	(chlapec a „ano“)	(chlapec a „ne“)
$P(\text{Dívka}) = .45$	(dívka a „ano“)	(dívka a „ne“)
	$P(\text{Ano}) = .40$	$P(\text{Ne}) = .60$

- $p(\text{chlapec}), p(\text{dívka}), p(\text{ano}), p(\text{ne})$  = marginální pravděpodobnosti  
(neboť se vyskytují na okraji tabulky)
- Každá marginální  $p$  = suma ( $\Sigma$ ) všech spojených pravděpodobností v určitém konkrétním řádku nebo sloupci tabulky
  - Příklad.  $p(\text{chlapec}) = p(\text{chlapec a „ano“}) + p(\text{chlapec a „ne“})$   
 $p(\text{dívka}) = p(\text{dívka a „ano“}) + p(\text{dívka a „ne“})$   
 $p(\text{ano}) = p(\text{chlapec a „ano“}) + p(\text{dívka a „ano“})$   
 $p(\text{ne}) = p(\text{chlapec a „ne“}) + p(\text{dívka a „ne“})$

P (chlapec)= .55	(chlapec a „ano“)	(chlapec a „ne“)
P(Dívka) = .45	(dívka a „ano“)	(dívka a „ne“)
	P(Ano) = .40	P(Ne) = .60

- Jevy/třídy jevů vyskytující se podél okraje jsou vzájemně neslučitelné a vyčerpávající (sada těchto jevů tedy formuje S) = dimenze
  - Příklad 2 dimenze: jevy „ano“ a „ne“ a jevy „chlapec“ a „dívka“
- Statistická nezávislost : každá kategorie nebo jev podél jednoho okraje musí být nezávislý na každém jevu podél druhého okraje = pravděpodobnost každého spojeného jevu se musí rovnat součinu pravděpodobností korespondujících (v řádku a sloupci) marginálních jevů -  $p(A \cap B) = p(A) * p(B)$ 
  - Příklad. Pokud dimenze „pohlaví“ a „odpověď ano/ne“ jsou nezávislé, pak  $p(\text{chlapec a „ano“}) = p(\text{chlapec}) * p(\text{„ano“}) = .55 * .40 = .22$   
Stejně postupujeme v ostatních případech a vznikne tabulka:

P (chlapec)= .55	$p(\text{chlapec})p(\text{„ano“})$ $= .55 * .40 = .22$	$p(\text{chlapec})p(\text{„ne“})$ $= .55 * .60 = .33$
P(Dívka) = .45	$p(\text{dívka})p(\text{„ano“})$ $= .45 * .40 = .18$	$p(\text{dívka})p(\text{„ne“})$ $= .45 * .60 = .27$
	P(Ano) = .40	P(Ne) = .60

- Pro závislé dimenze
  - pravděpodobnosti spojených jevů  $\neq$  součiny marginálních pravděpodobností:

$P(\text{chlapec}) = .55$	$p(\text{chlapec})p(\text{„ano“}) = .10$	$p(\text{chlapec})p(\text{„ne“}) = .45$
$P(\text{Dívka}) = .45$	$p(\text{dívka})p(\text{„ano“}) = .30$	$p(\text{dívka})p(\text{„ne“}) = .15$
	$P(\text{Ano}) = .40$	$P(\text{Ne}) = .60$

- Pro tuto základní množinu S (např. odpovědi dívek poté co dostali od svého milého kytici růží) pohlaví a odpověď spolu souvisejí
- Prozkoumáme-li podmíněné pravděpodobnosti,
  - $P(\text{ano} | \text{chlapec}) = .10 / .55 = .18$
  - $P(\text{ano} | \text{dívka}) = .30 / .45 = .67$
 pak vidíme že **dívky mnohem pravděpodobněji/častěji odpovídají ano v porovnání s chlapci**
- Pokud by ale obě dimenze byly nezávislé (jako v 1.tabulce) pak by tyto podmíněné pravděpodobnosti měly být stejné,
  - $P(\text{ano} | \text{chlapec}) = .22 / .55 = .40$
  - $P(\text{ano} | \text{dívka}) = .18 / .45 = .40$
 takže **znám-li pohlaví nedává mi to informaci o tom jak bude student odpovídat**