

Statistika

(MD360P03Z, MD360P03U)
ak. rok 2007/2008

Karel Zvára

karel.zvara@mff.cuni.cz
<http://www.karlin.mff.cuni.cz/~zvara>

23. října 2007



možné příští úlohy statistické indukce

- ▶ na hracích kostkách A a B padala šestka nestejně často:
na kostce A v 17 ze 100 pokusů
na kostce B v 41 ze 100 pokusů
- ▶ je pravděpodobnost šestky rovna $1/6$?
 - ▶ teorie pravděpodobnosti odvodí teoretickou hodnotu
 - ▶ matematická statistika odhadne, prověří představu teorie
- ▶ je kostka symetrická, tj. mají všechny stěny kostky stejnou pravděpodobnost?
- ▶ kolik potřebujeme nezávislých hodů, abychom s požadovanou spolehlivostí poznali, že je kostka nesymetrická?
- ▶ liší se mezi sebou kostky A a B?
- ▶ vše založeno na modelu **populace – výběr** [population, sample]

možné příští úlohy statistické indukce

- ▶ na hracích kostkách A a B padala šestka nestejně často:
na kostce A v 17 ze 100 pokusů
na kostce B v 41 ze 100 pokusů
- ▶ je pravděpodobnost šestky rovna $1/6$?
 - ▶ teorie pravděpodobnosti odvodí teoretickou hodnotu
 - ▶ matematická statistika odhadne, prověří představu teorie
- ▶ je kostka symetrická, tj. mají všechny stěny kostky stejnou pravděpodobnost?
- ▶ kolik potřebujeme nezávislých hodů, abychom s požadovanou spolehlivostí poznali, že je kostka nesymetrická?
- ▶ liší se mezi sebou kostky A a B?
- ▶ vše založeno na modelu **populace – výběr** [population, sample]

možné příští úlohy statistické indukce

- ▶ na hracích kostkách A a B padala šestka nestejně často:
na kostce A v 17 ze 100 pokusů
na kostce B v 41 ze 100 pokusů
- ▶ je pravděpodobnost šestky rovna $1/6$?
 - ▶ teorie pravděpodobnosti odvodí teoretickou hodnotu
 - ▶ matematická statistika odhadne, prověří představu teorie
- ▶ je kostka symetrická, tj. mají všechny stěny kostky stejnou pravděpodobnost?
- ▶ kolik potřebujeme nezávislých hodů, abychom s požadovanou spolehlivostí poznali, že je kostka nesymetrická?
- ▶ liší se mezi sebou kostky A a B?
- ▶ vše založeno na modelu **populace – výběr** [population, sample]

možné příští úlohy statistické indukce

- ▶ na hracích kostkách A a B padala šestka nestejně často:
na kostce A v 17 ze 100 pokusů
na kostce B v 41 ze 100 pokusů
- ▶ je pravděpodobnost šestky rovna $1/6$?
 - ▶ teorie pravděpodobnosti odvodí teoretickou hodnotu
 - ▶ matematická statistika odhadne, prověří představu teorie
- ▶ je kostka symetrická, tj. mají všechny stěny kostky stejnou pravděpodobnost?
- ▶ kolik potřebujeme nezávislých hodů, abychom s požadovanou spolehlivostí poznali, že je kostka nesymetrická?
- ▶ liší se mezi sebou kostky A a B?
- ▶ vše založeno na modelu **populace – výběr** [population, sample]

možné příští úlohy statistické indukce

- ▶ na hracích kostkách A a B padala šestka nestejně často:
na kostce A v 17 ze 100 pokusů
na kostce B v 41 ze 100 pokusů
- ▶ je pravděpodobnost šestky rovna $1/6$?
 - ▶ teorie pravděpodobnosti odvodí teoretickou hodnotu
 - ▶ matematická statistika odhadne, prověří představu teorie
- ▶ je kostka symetrická, tj. mají všechny stěny kostky stejnou pravděpodobnost?
- ▶ kolik potřebujeme nezávislých hodů, abychom s požadovanou spolehlivostí poznali, že je kostka nesymetrická?
- ▶ liší se mezi sebou kostky A a B?
- ▶ vše založeno na modelu **populace – výběr** [population, sample]

možné příští úlohy statistické indukce

- ▶ na hracích kostkách A a B padala šestka nestejně často:
na kostce A v 17 ze 100 pokusů
na kostce B v 41 ze 100 pokusů
- ▶ je pravděpodobnost šestky rovna $1/6$?
 - ▶ teorie pravděpodobnosti odvodí teoretickou hodnotu
 - ▶ matematická statistika odhadne, prověří představu teorie
- ▶ je kostka symetrická, tj. mají všechny stěny kostky stejnou pravděpodobnost?
- ▶ kolik potřebujeme nezávislých hodů, abychom s požadovanou spolehlivostí poznali, že je kostka nesymetrická?
- ▶ liší se mezi sebou kostky A a B?
- ▶ vše založeno na modelu **populace – výběr** [population, sample]

možné příští úlohy statistické indukce

- ▶ na hracích kostkách A a B padala šestka nestejně často:
na kostce A v 17 ze 100 pokusů
na kostce B v 41 ze 100 pokusů
- ▶ je pravděpodobnost šestky rovna $1/6$?
 - ▶ teorie pravděpodobnosti odvodí teoretickou hodnotu
 - ▶ matematická statistika odhadne, prověří představu teorie
- ▶ je kostka symetrická, tj. mají všechny stěny kostky stejnou pravděpodobnost?
- ▶ kolik potřebujeme nezávislých hodů, abychom s požadovanou spolehlivostí poznali, že je kostka nesymetrická?
- ▶ liší se mezi sebou kostky A a B?
- ▶ vše založeno na modelu **populace – výběr** [population, sample]

možné příští úlohy statistické indukce

- ▶ na hracích kostkách A a B padala šestka nestejně často:
na kostce A v 17 ze 100 pokusů
na kostce B v 41 ze 100 pokusů
- ▶ je pravděpodobnost šestky rovna $1/6$?
 - ▶ teorie pravděpodobnosti odvodí teoretickou hodnotu
 - ▶ matematická statistika odhadne, prověří představu teorie
- ▶ je kostka symetrická, tj. mají všechny stěny kostky stejnou pravděpodobnost?
- ▶ kolik potřebujeme nezávislých hodů, abychom s požadovanou spolehlivostí poznali, že je kostka nesymetrická?
- ▶ liší se mezi sebou kostky A a B?
- ▶ vše založeno na modelu **populace – výběr** [population, sample]

populace a výběr

- ▶ model **populace** – **výběr** umožňuje zobecnění na celou populaci z hodnot zjištěných na vybraných statistických jednotkách (výběr)
- ▶ **populace (základní soubor)** – velký soubor, jehož je zpracováván soubor (**výběr**) reprezentativním vzorkem
- ▶ **reprezentativnost** – frekvence výskytu důležitých doprovodných znaků ve výběru odpovídá jejich frekvenci v populaci
- ▶ reprezentativnosti nejlépe dosáhneme tak, že použijeme **prostý náhodný výběr**, kdy každá n -tice prvků populace má stejnou šanci (pravděpodobnost) do výběru se dostat
- ▶ na základě výběru tvrdíme něco o populaci

populace a výběr

- ▶ model **populace** – **výběr** umožňuje zobecnění na celou populaci z hodnot zjištěných na vybraných statistických jednotkách (výběr)
- ▶ **populace (základní soubor)** – velký soubor, jehož je zpracováván soubor (**výběr**) reprezentativním vzorkem
- ▶ **reprezentativnost** – frekvence výskytu důležitých doprovodných znaků ve výběru odpovídá jejich frekvenci v populaci
- ▶ reprezentativnosti nejlépe dosáhneme tak, že použijeme **prostý náhodný výběr**, kdy každá n -tice prvků populace má stejnou šanci (pravděpodobnost) do výběru se dostat
- ▶ na základě výběru tvrdíme něco o populaci

populace a výběr

- ▶ model **populace** – **výběr** umožňuje zobecnění na celou populaci z hodnot zjištěných na vybraných statistických jednotkách (výběr)
- ▶ **populace (základní soubor)** – velký soubor, jehož je zpracováván soubor (**výběr**) reprezentativním vzorkem
- ▶ **reprezentativnost** – frekvence výskytu důležitých doprovodných znaků ve výběru odpovídá jejich frekvenci v populaci
- ▶ reprezentativnosti nejlépe dosáhneme tak, že použijeme **prostý náhodný výběr**, kdy každá n -tice prvků populace má stejnou šanci (pravděpodobnost) do výběru se dostat
- ▶ na základě výběru tvrdíme něco o populaci

populace a výběr

- ▶ model **populace – výběr** umožňuje zobecnění na celou populaci z hodnot zjištěných na vybraných statistických jednotkách (výběr)
- ▶ **populace (základní soubor)** – velký soubor, jehož je zpracováván soubor (**výběr**) reprezentativním vzorkem
- ▶ **reprezentativnost** – frekvence výskytu důležitých doprovodných znaků ve výběru odpovídá jejich frekvenci v populaci
- ▶ reprezentativnosti nejlépe dosáhneme tak, že použijeme **prostý náhodný výběr**, kdy každá n -tice prvků populace má stejnou šanci (pravděpodobnost) do výběru se dostat
- ▶ na základě výběru tvrdíme něco o populaci

populace a výběr

- ▶ model **populace** – **výběr** umožňuje zobecnění na celou populaci z hodnot zjištěných na vybraných statistických jednotkách (výběr)
- ▶ **populace (základní soubor)** – velký soubor, jehož je zpracováván soubor (**výběr**) reprezentativním vzorkem
- ▶ **reprezentativnost** – frekvence výskytu důležitých doprovodných znaků ve výběru odpovídá jejich frekvenci v populaci
- ▶ reprezentativnosti nejlépe dosáhneme tak, že použijeme **prostý náhodný výběr**, kdy každá n -tice prvků populace má stejnou šanci (pravděpodobnost) do výběru se dostat
- ▶ na základě výběru tvrdíme něco o populaci

parametry – odhady, statistiky

- ▶ podle toho, jakou roli hraje hodnocený soubor, rozlišujeme **charakteristiky**
 - ▶ **populační**: vztažené k populaci, mnohdy jen ideální, námi představované, jsou to **parametry** modelu
 - ▶ **výběrové**: vztažené k výběru z nějaké populace, jsou to **statistiky** spočítané z výběru
- ▶ **statistika** – z výběru spočítaná hodnota (např. součet napozorovaných hodnot, průměr, Giniho index ...)
- ▶ speciálním případem statistik jsou **odhady** odpovídajících populačních **parametrů**,
- ▶ příkladem dvojice odhad – parametr je dvojice relativní četnost – pravděpodobnost (např. 17/100 vers. 1/6)
- ▶ statistiky se používají při **statistické indukci** (statistickém rozhodování) [statistical inference (decisions)]

parametry – odhady, statistiky

- ▶ podle toho, jakou roli hraje hodnocený soubor, rozlišujeme **charakteristiky**
 - ▶ **populační**: vztažené k populaci, mnohdy jen ideální, námi představované, jsou to **parametry** modelu
 - ▶ **výběrové**: vztažené k výběru z nějaké populace, jsou to **statistiky** spočítané z výběru
- ▶ **statistika** – z výběru spočítaná hodnota (např. součet napozorovaných hodnot, průměr, Giniho index ...)
- ▶ speciálním případem statistik jsou **odhady** odpovídajících populačních **parametrů**,
- ▶ příkladem dvojice odhad – parametr je dvojice relativní četnost – pravděpodobnost (např. 17/100 vers. 1/6)
- ▶ statistiky se používají při **statistické indukci** (statistickém rozhodování) [statistical inference (decisions)]

parametry – odhady, statistiky

- ▶ podle toho, jakou roli hraje hodnocený soubor, rozlišujeme **charakteristiky**
 - ▶ **populační**: vztažené k populaci, mnohdy jen ideální, námi představované, jsou to **parametry** modelu
 - ▶ **výběrové**: vztažené k výběru z nějaké populace, jsou to **statistiky** spočítané z výběru
- ▶ **statistika** – z výběru spočítaná hodnota (např. součet napozorovaných hodnot, průměr, Giniho index ...)
- ▶ speciálním případem statistik jsou **odhady** odpovídajících populačních **parametrů**,
- ▶ příkladem dvojice odhad – parametr je dvojice relativní četnost – pravděpodobnost (např. 17/100 vers. 1/6)
- ▶ statistiky se používají při **statistické indukci** (statistickém rozhodování) [statistical inference (decisions)]

parametry – odhady, statistiky

- ▶ podle toho, jakou roli hraje hodnocený soubor, rozlišujeme **charakteristiky**
 - ▶ **populační**: vztažené k populaci, mnohdy jen ideální, námi představované, jsou to **parametry** modelu
 - ▶ **výběrové**: vztažené k výběru z nějaké populace, jsou to **statistiky** spočítané z výběru
- ▶ **statistika** – z výběru spočítaná hodnota (např. součet napozorovaných hodnot, průměr, Giniho index . . .)
- ▶ speciálním případem statistik jsou **odhady** odpovídajících populačních **parametrů**,
- ▶ příkladem dvojice odhad – parametr je dvojice relativní četnost – pravděpodobnost (např. 17/100 vers. 1/6)
- ▶ statistiky se používají při **statistické indukci** (statistickém rozhodování) [statistical inference (decisions)]

parametry – odhady, statistiky

- ▶ podle toho, jakou roli hraje hodnocený soubor, rozlišujeme **charakteristiky**
 - ▶ **populační**: vztažené k populaci, mnohdy jen ideální, námi představované, jsou to **parametry** modelu
 - ▶ **výběrové**: vztažené k výběru z nějaké populace, jsou to **statistiky** spočítané z výběru
- ▶ **statistika** – z výběru spočítaná hodnota (např. součet napozorovaných hodnot, průměr, Giniho index . . .)
- ▶ speciálním případem statistik jsou **odhady** odpovídajících populačních **parametrů**,
- ▶ příkladem dvojice odhad – parametr je dvojice relativní četnost – pravděpodobnost (např. 17/100 vers. 1/6)
- ▶ statistiky se používají při **statistické indukci** (statistickém rozhodování) [statistical inference (decisions)]

parametry – odhady, statistiky

- ▶ podle toho, jakou roli hraje hodnocený soubor, rozlišujeme **charakteristiky**
 - ▶ **populační**: vztažené k populaci, mnohdy jen ideální, námi představované, jsou to **parametry** modelu
 - ▶ **výběrové**: vztažené k výběru z nějaké populace, jsou to **statistiky** spočítané z výběru
- ▶ **statistika** – z výběru spočítaná hodnota (např. součet napozorovaných hodnot, průměr, Giniho index . . .)
- ▶ speciálním případem statistik jsou **odhady** odpovídajících populačních **parametrů**,
- ▶ příkladem dvojice odhad – parametr je dvojice relativní četnost – pravděpodobnost (např. 17/100 vers. 1/6)
- ▶ statistiky se používají při **statistické indukci** (statistickém rozhodování) [statistical inference (decisions)]

parametry – odhady, statistiky

- ▶ podle toho, jakou roli hraje hodnocený soubor, rozlišujeme **charakteristiky**
 - ▶ **populační**: vztažené k populaci, mnohdy jen ideální, námi představované, jsou to **parametry** modelu
 - ▶ **výběrové**: vztažené k výběru z nějaké populace, jsou to **statistiky** spočítané z výběru
- ▶ **statistika** – z výběru spočítaná hodnota (např. součet napozorovaných hodnot, průměr, Giniho index . . .)
- ▶ speciálním případem statistik jsou **odhady** odpovídajících populačních **parametrů**,
- ▶ příkladem dvojice odhad – parametr je dvojice relativní četnost – pravděpodobnost (např. 17/100 vers. 1/6)
- ▶ statistiky se používají při **statistické indukci** (statistickém rozhodování) [statistical inference (decisions)]

základní pojmy

- ▶ **pokus** – dobře definovaná situace (postup), která končí jedním z řady možných výsledků (vržená kostka spadne na zem)
- ▶ **náhodný pokus** – pokus, u něhož předem nevíme, který výsledek nastane (která strana kostky padne přistě?); předpokládá se stabilita relativních četností možných výsledků
- ▶ **náhodný jev** – tvrzení o výsledku náhodného pokusu
- ▶ **pravděpodobnost** náhodného jevu A – číselné vyjádření očekávání, že výsledkem náhodného pokusu bude právě A
- ▶ racionální představa: při velkém počtu opakování pokusu se relativní četnost jevu blíží k pravděpodobnosti tohoto jevu

základní pojmy

- ▶ **pokus** – dobře definovaná situace (postup), která končí jedním z řady možných výsledků (vržená kostka spadne na zem)
- ▶ **náhodný pokus** – pokus, u něhož předem nevíme, který výsledek nastane (která strana kostky padne příště?); předpokládá se stabilita relativních četností možných výsledků
- ▶ **náhodný jev** – tvrzení o výsledku náhodného pokusu
- ▶ **pravděpodobnost** náhodného jevu A – číselné vyjádření očekávání, že výsledkem náhodného pokusu bude právě A
- ▶ racionální představa: při velkém počtu opakování pokusu se relativní četnost jevu blíží k pravděpodobnosti tohoto jevu

základní pojmy

- ▶ **pokus** – dobře definovaná situace (postup), která končí jedním z řady možných výsledků (vržená kostka spadne na zem)
- ▶ **náhodný pokus** – pokus, u něhož předem nevíme, který výsledek nastane (která strana kostky padne příště?); předpokládá se stabilita relativních četností možných výsledků
- ▶ **náhodný jev** – tvrzení o výsledku náhodného pokusu
- ▶ **pravděpodobnost** náhodného jevu A – číselné vyjádření očekávání, že výsledkem náhodného pokusu bude právě A
- ▶ racionální představa: při velkém počtu opakování pokusu se relativní četnost jevu blíží k pravděpodobnosti tohoto jevu

základní pojmy

- ▶ **pokus** – dobře definovaná situace (postup), která končí jedním z řady možných výsledků (vržená kostka spadne na zem)
- ▶ **náhodný pokus** – pokus, u něhož předem nevíme, který výsledek nastane (která strana kostky padne příště?); předpokládá se stabilita relativních četností možných výsledků
- ▶ **náhodný jev** – tvrzení o výsledku náhodného pokusu
- ▶ **pravděpodobnost** náhodného jevu A – číselné vyjádření očekávání, že výsledkem náhodného pokusu bude právě A
- ▶ racionální představa: při velkém počtu opakování pokusu se relativní četnost jevu blíží k pravděpodobnosti tohoto jevu

základní pojmy

- ▶ **pokus** – dobře definovaná situace (postup), která končí jedním z řady možných výsledků (vržená kostka spadne na zem)
- ▶ **náhodný pokus** – pokus, u něhož předem nevíme, který výsledek nastane (která strana kostky padne příště?); předpokládá se stabilita relativních četností možných výsledků
- ▶ **náhodný jev** – tvrzení o výsledku náhodného pokusu
- ▶ **pravděpodobnost** náhodného jevu A – číselné vyjádření očekávání, že výsledkem náhodného pokusu bude právě A
- ▶ racionální představa: při velkém počtu opakování pokusu se relativní četnost jevu blíží k pravděpodobnosti tohoto jevu

klasická pravděpodobnost (Laplace)

- ▶ **jistý jev** (nastává vždy) lze rozdělit na M *stejně pravděpodobných* neslučitelných (disjunktních) **elementárních jevů** (symetrie)
- ▶ každý jev lze složit z těchto **elementárních jevů**
- ▶ je celkem M_A **příznivých** jevu A (je z nich složen)
- ▶ **klasická definice pravděpodobnosti** (metoda výpočtu)

$$P(A) = \frac{M_A}{M}$$

- ▶ **klasickou pst lze použít jen někdy!** (Sportka, Sazka)

klasická pravděpodobnost (Laplace)

- ▶ **jistý jev** (nastává vždy) lze rozdělit na M *stejně pravděpodobných* neslučitelných (disjunktních) **elementárních jevů** (symetrie)
- ▶ každý jev lze složit z těchto **elementárních jevů**
- ▶ je celkem M_A **příznivých** jevu A (je z nich složen)
- ▶ **klasická definice pravděpodobnosti** (metoda výpočtu)

$$P(A) = \frac{M_A}{M}$$

- ▶ **klasickou pst lze použít jen někdy!** (Sportka, Sazka)

klasická pravděpodobnost (Laplace)

- ▶ **jistý jev** (nastává vždy) lze rozdělit na M *stejně pravděpodobných* neslučitelných (disjunktních) **elementárních jevů** (symetrie)
- ▶ každý jev lze složit z těchto **elementárních jevů**
- ▶ je celkem M_A **příznivých** jevu A (je z nich složen)
- ▶ **klasická definice pravděpodobnosti** (metoda výpočtu)

$$P(A) = \frac{M_A}{M}$$

- ▶ **klasickou pst lze použít jen někdy!** (Sportka, Sazka)

klasická pravděpodobnost (Laplace)

- ▶ **jistý jev** (nastává vždy) lze rozdělit na M *stejně pravděpodobných* neslučitelných (disjunktních) **elementárních jevů** (symetrie)
- ▶ každý jev lze složit z těchto **elementárních jevů**
- ▶ je celkem M_A **příznivých** jevu A (je z nich složen)
- ▶ **klasická definice pravděpodobnosti** (metoda výpočtu)

$$P(A) = \frac{M_A}{M}$$

- ▶ klasickou pst lze použít jen někdy! (Sportka, Sazka)

klasická pravděpodobnost (Laplace)

- ▶ **jistý jev** (nastává vždy) lze rozdělit na M *stejně pravděpodobných* neslučitelných (disjunktních) **elementárních jevů** (symetrie)
- ▶ každý jev lze složit z těchto **elementárních jevů**
- ▶ je celkem M_A **příznivých** jevu A (je z nich složen)
- ▶ **klasická definice pravděpodobnosti** (metoda výpočtu)

$$P(A) = \frac{M_A}{M}$$

- ▶ **klasickou pst lze použít jen někdy!** (Sportka, Sazka)

příklad: hrací kostka

- ▶ idealizovaná symetrická hrací kostka
 - ▶ homogenní
 - ▶ přesná krychle
 - ▶ těžiště uprostřed
 - ▶ každá strana má stejnou pravděpodobnost
- ▶ A – padne šestka, B – padne sudé číslo
- ▶ $M = 6$
- ▶ $M_A = 1$, tedy $P(A) = 1/6$
- ▶ $M_B = 3$, tedy $P(B) = 3/6 = 1/2$

příklad: hrací kostka

- ▶ idealizovaná symetrická hrací kostka
 - ▶ homogenní
 - ▶ přesná krychle
 - ▶ těžiště uprostřed
 - ▶ každá strana má stejnou pravděpodobnost
- ▶ A – padne šestka, B – padne sudé číslo
- ▶ $M = 6$
- ▶ $M_A = 1$, tedy $P(A) = 1/6$
- ▶ $M_B = 3$, tedy $P(B) = 3/6 = 1/2$

příklad: hrací kostka

- ▶ idealizovaná symetrická hrací kostka
 - ▶ homogenní
 - ▶ přesná krychle
 - ▶ těžiště uprostřed
 - ▶ každá strana má stejnou pravděpodobnost
- ▶ A – padne šestka, B – padne sudé číslo
- ▶ $M = 6$
- ▶ $M_A = 1$, tedy $P(A) = 1/6$
- ▶ $M_B = 3$, tedy $P(B) = 3/6 = 1/2$

příklad: hrací kostka

- ▶ idealizovaná symetrická hrací kostka
 - ▶ homogenní
 - ▶ přesná krychle
 - ▶ těžiště uprostřed
 - ▶ každá strana má stejnou pravděpodobnost
- ▶ A – padne šestka, B – padne sudé číslo
- ▶ $M = 6$
- ▶ $M_A = 1$, tedy $P(A) = 1/6$
- ▶ $M_B = 3$, tedy $P(B) = 3/6 = 1/2$

příklad: hrací kostka

- ▶ idealizovaná symetrická hrací kostka
 - ▶ homogenní
 - ▶ přesná krychle
 - ▶ těžiště uprostřed
 - ▶ každá strana má stejnou pravděpodobnost
- ▶ A – padne šestka, B – padne sudé číslo
- ▶ $M = 6$
- ▶ $M_A = 1$, tedy $P(A) = 1/6$
- ▶ $M_B = 3$, tedy $P(B) = 3/6 = 1/2$

příklad: hrací kostka

- ▶ idealizovaná symetrická hrací kostka
 - ▶ homogenní
 - ▶ přesná krychle
 - ▶ těžiště uprostřed
 - ▶ každá strana má stejnou pravděpodobnost
- ▶ A – padne šestka, B – padne sudé číslo
- ▶ $M = 6$
- ▶ $M_A = 1$, tedy $P(A) = 1/6$
- ▶ $M_B = 3$, tedy $P(B) = 3/6 = 1/2$

příklad: hrací kostka

- ▶ idealizovaná symetrická hrací kostka
 - ▶ homogenní
 - ▶ přesná krychle
 - ▶ těžiště uprostřed
 - ▶ každá strana má stejnou pravděpodobnost
- ▶ A – padne šestka, B – padne sudé číslo
- ▶ $M = 6$
- ▶ $M_A = 1$, tedy $P(A) = 1/6$
- ▶ $M_B = 3$, tedy $P(B) = 3/6 = 1/2$

příklad: hrací kostka

- ▶ idealizovaná symetrická hrací kostka
 - ▶ homogenní
 - ▶ přesná krychle
 - ▶ těžiště uprostřed
 - ▶ každá strana má stejnou pravděpodobnost
- ▶ A – padne šestka, B – padne sudé číslo
- ▶ $M = 6$
- ▶ $M_A = 1$, tedy $P(A) = 1/6$
- ▶ $M_B = 3$, tedy $P(B) = 3/6 = 1/2$

příklad: hrací kostka

- ▶ idealizovaná symetrická hrací kostka
 - ▶ homogenní
 - ▶ přesná krychle
 - ▶ těžiště uprostřed
 - ▶ každá strana má stejnou pravděpodobnost
- ▶ A – padne šestka, B – padne sudé číslo
- ▶ $M = 6$
- ▶ $M_A = 1$, tedy $P(A) = 1/6$
- ▶ $M_B = 3$, tedy $P(B) = 3/6 = 1/2$

faktoriál

[FAKTORIÁL(n)]

[factorial(n)]

- ▶ **faktoriál** $n! = n \cdot (n - 1) \cdots 2 \cdot 1$ $0! = 1$
- ▶ kolika způsoby lze uspořádat za sebou n rozlišitelných prvků
- ▶ příklady:
 - ▶ $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$
 - ▶ $1! = 1$
- ▶ kolika způsoby lze uspořádat za sebou 14 krajů ČR:
 $14! = 14 \cdot 13 \cdot 12 \cdots 2 \cdot 1 = 87\,178\,291\,200 = 8,7 \cdot 10^{10}$

faktoriál

[FAKTORIÁL(n)]

[factorial(n)]

- ▶ **faktoriál** $n! = n \cdot (n - 1) \cdots 2 \cdot 1$ $0! = 1$
- ▶ kolika způsoby lze uspořádat za sebou n rozlišitelných prvků
- ▶ příklady:
 - ▶ $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$
 - ▶ $1! = 1$
- ▶ kolika způsoby lze uspořádat za sebou 14 krajů ČR:
 $14! = 14 \cdot 13 \cdot 12 \cdots 2 \cdot 1 = 87\,178\,291\,200 = 8,7 \cdot 10^{10}$

faktoriál

[FAKTORIÁL(n)]

[factorial(n)]

- ▶ **faktoriál** $n! = n \cdot (n - 1) \cdots 2 \cdot 1$ $0! = 1$
- ▶ kolika způsoby lze uspořádat za sebou n rozlišitelných prvků
- ▶ příklady:
 - ▶ $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$
 - ▶ $1! = 1$
- ▶ kolika způsoby lze uspořádat za sebou 14 krajů ČR:
 $14! = 14 \cdot 13 \cdot 12 \cdots 2 \cdot 1 = 87\,178\,291\,200 = 8,7 \cdot 10^{10}$

faktoriál

[FAKTORIÁL(n)]

[factorial(n)]

- ▶ **faktoriál** $n! = n \cdot (n - 1) \cdots 2 \cdot 1$ $0! = 1$
- ▶ kolika způsoby lze uspořádat za sebou n rozlišitelných prvků
- ▶ příklady:
 - ▶ $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$
 - ▶ $1! = 1$
- ▶ kolika způsoby lze uspořádat za sebou 14 krajů ČR:
 $14! = 14 \cdot 13 \cdot 12 \cdots 2 \cdot 1 = 87\,178\,291\,200 = 8,7 \cdot 10^{10}$

faktoriál

[FAKTORIÁL(n)]

[factorial(n)]

- ▶ **faktoriál** $n! = n \cdot (n - 1) \cdots 2 \cdot 1$ $0! = 1$
- ▶ kolika způsoby lze uspořádat za sebou n rozlišitelných prvků
- ▶ příklady:
 - ▶ $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$
 - ▶ $1! = 1$
- ▶ kolika způsoby lze uspořádat za sebou 14 krajů ČR:
 $14! = 14 \cdot 13 \cdot 12 \cdots 2 \cdot 1 = 87\,178\,291\,200 = 8,7 \cdot 10^{10}$

faktoriál

[FAKTORIÁL(n)]

[factorial(n)]

- ▶ **faktoriál** $n! = n \cdot (n - 1) \cdots 2 \cdot 1$ $0! = 1$
- ▶ kolika způsoby lze uspořádat za sebou n rozlišitelných prvků
- ▶ příklady:
 - ▶ $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$
 - ▶ $1! = 1$
- ▶ kolika způsoby lze uspořádat za sebou 14 krajů ČR:
 $14! = 14 \cdot 13 \cdot 12 \cdots 2 \cdot 1 = 87\,178\,291\,200 = 8,7 \cdot 10^{10}$

počet kombinací

[KOMBINACE(n ; k)]

[choose(n , k)]

- ▶ **kombinační číslo** $\binom{n}{k}$ (čti „ n nad k “)
- ▶ počet k -prvkových podmnožin množiny o n prvcích nezávisle na jejich pořadí

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k \cdot (k-1) \cdots 2 \cdot 1}$$

- ▶ kolika způsoby si mohu z pěti knížek vybrat dvě na dovolenou:

$$\binom{5}{2} = \frac{5!}{2!3!} = \frac{5 \cdot 4}{2 \cdot 1} = 10$$

- ▶ kolika způsoby si z oněch pěti mohu vybrat tři knihy? (10)

počet kombinací

[KOMBINACE(n ; k)][choose(n , k)]

- ▶ **kombinační číslo** $\binom{n}{k}$ (čti „ n nad k “)
- ▶ počet k -prvkových podmnožin množiny o n prvcích nezávisle na jejich pořadí

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k \cdot (k-1) \cdots 2 \cdot 1}$$

- ▶ kolika způsoby si mohou z pěti knížek vybrat dvě na dovolenou:

$$\binom{5}{2} = \frac{5!}{2!3!} = \frac{5 \cdot 4}{2 \cdot 1} = 10$$

- ▶ kolika způsoby si z oněch pěti mohou vybrat tři knihy? (10)

počet kombinací

[KOMBINACE(n ; k)]

[choose(n , k)]

- ▶ **kombinační číslo** $\binom{n}{k}$ (čti „ n nad k “)
- ▶ počet k -prvkových podmnožin množiny o n prvcích nezávisle na jejich pořadí

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k \cdot (k-1) \cdots 2 \cdot 1}$$

- ▶ kolika způsoby si mohou z pěti knížek vybrat dvě na dovolenou:

$$\binom{5}{2} = \frac{5!}{2!3!} = \frac{5 \cdot 4}{2 \cdot 1} = 10$$

- ▶ kolika způsoby si z oněch pěti mohou vybrat tři knihy? (10)

počet kombinací

[KOMBINACE(n ; k)]

[choose(n , k)]

- ▶ **kombinační číslo** $\binom{n}{k}$ (čti „ n nad k “)
- ▶ počet k -prvkových podmnožin množiny o n prvcích nezávisle na jejich pořadí

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k \cdot (k-1) \cdots 2 \cdot 1}$$

- ▶ kolika způsoby si mohu z pěti knížek vybrat dvě na dovolenou:

$$\binom{5}{2} = \frac{5!}{2!3!} = \frac{5 \cdot 4}{2 \cdot 1} = 10$$

- ▶ kolika způsoby si z oněch pěti mohu vybrat tři knihy? (10)

příklad: losování otázek (1)

- ▶ student *neumí* 5 otázek, *umí* 10 otázek
- ▶ losuje se dvojice otázek z oněch 15 otázek
- ▶ pravděpodobnost $P(A)$, že student nezná ani jednu z vylosovaných:
- ▶ elementární jevy: první losovaná otázka – 15 možností, druhá jen 14 možností, nezáleží na pořadí, tedy dělit 2 (tedy počet kombinací)

$$M = \binom{5 + 10}{2} = \binom{15}{2} = \frac{15!}{2!13!} = \frac{15 \cdot 14}{2 \cdot 1} = 105$$

- ▶ příznivé elementární jevy: vylosuje obě z pěti, které neumí

$$M_A = \binom{5}{2} \binom{10}{0} = \frac{5 \cdot 4}{2 \cdot 1} \cdot 1 = 10 \Rightarrow P(A) = \frac{10}{105} = 9,5 \%$$

příklad: losování otázek (1)

- ▶ student *neumí* 5 otázek, *umí* 10 otázek
- ▶ losuje se dvojice otázek z oněch 15 otázek
- ▶ pravděpodobnost $P(A)$, že student nezná ani jednu z vylosovaných:
- ▶ elementární jevy: první losovaná otázka – 15 možností, druhá jen 14 možností, nezáleží na pořadí, tedy dělit 2 (tedy počet kombinací)

$$M = \binom{5 + 10}{2} = \binom{15}{2} = \frac{15!}{2!13!} = \frac{15 \cdot 14}{2 \cdot 1} = 105$$

- ▶ příznivé elementární jevy: vylosuje obě z pěti, které neumí

$$M_A = \binom{5}{2} \binom{10}{0} = \frac{5 \cdot 4}{2 \cdot 1} \cdot 1 = 10 \Rightarrow P(A) = \frac{10}{105} = 9,5 \%$$

příklad: losování otázek (1)

- ▶ student *neumí* 5 otázek, *umí* 10 otázek
- ▶ losuje se dvojice otázek z oněch 15 otázek
- ▶ pravděpodobnost $P(A)$, že student nezná ani jednu z vylosovaných:
- ▶ elementární jevy: první losovaná otázka – 15 možností, druhá jen 14 možností, nezáleží na pořadí, tedy dělit 2 (tedy počet kombinací)

$$M = \binom{5 + 10}{2} = \binom{15}{2} = \frac{15!}{2!13!} = \frac{15 \cdot 14}{2 \cdot 1} = 105$$

- ▶ příznivé elementární jevy: vylosuje obě z pěti, které neumí

$$M_A = \binom{5}{2} \binom{10}{0} = \frac{5 \cdot 4}{2 \cdot 1} \cdot 1 = 10 \Rightarrow P(A) = \frac{10}{105} = 9,5 \%$$

příklad: losování otázek (1)

- ▶ student *neumí* 5 otázek, *umí* 10 otázek
- ▶ losuje se dvojice otázek z oněch 15 otázek
- ▶ pravděpodobnost $P(A)$, že student nezná ani jednu z vylosovaných:
- ▶ elementární jevy: první losovaná otázka – 15 možností, druhá jen 14 možností, nezáleží na pořadí, tedy dělit 2 (tedy počet kombinací)

$$M = \binom{5 + 10}{2} = \binom{15}{2} = \frac{15!}{2!13!} = \frac{15 \cdot 14}{2 \cdot 1} = 105$$

- ▶ příznivé elementární jevy: vylosuje obě z pěti, které neumí

$$M_A = \binom{5}{2} \binom{10}{0} = \frac{5 \cdot 4}{2 \cdot 1} \cdot 1 = 10 \Rightarrow P(A) = \frac{10}{105} = 9,5 \%$$

příklad: losování otázek (1)

- ▶ student *neumí* 5 otázek, *umí* 10 otázek
- ▶ losuje se dvojice otázek z oněch 15 otázek
- ▶ pravděpodobnost $P(A)$, že student nezná ani jednu z vylosovaných:
- ▶ elementární jevy: první losovaná otázka – 15 možností, druhá jen 14 možností, nezáleží na pořadí, tedy dělit 2 (tedy počet kombinací)

$$M = \binom{5 + 10}{2} = \binom{15}{2} = \frac{15!}{2!13!} = \frac{15 \cdot 14}{2 \cdot 1} = 105$$

- ▶ příznivé elementární jevy: vylosuje obě z pěti, které neumí

$$M_A = \binom{5}{2} \binom{10}{0} = \frac{5 \cdot 4}{2 \cdot 1} \cdot 1 = 10 \Rightarrow P(A) = \frac{10}{105} = 9,5 \%$$

příklad: losování otázek (2)

- ▶ pravděpodobnost $P(B)$, že zná *právě* jednu otázku

$$M_B = \binom{5}{1} \cdot \binom{10}{1} = 5 \cdot 10 = 50 \Rightarrow P(B) = \frac{50}{105} = 47,6 \%$$

- ▶ pravděpodobnost $P(C)$, že zná *obě* otázky (*právě dvě*)

$$M_C = \binom{5}{0} \cdot \binom{10}{2} = 1 \cdot \frac{10 \cdot 9}{2 \cdot 1} = 45 \Rightarrow P(C) = \frac{45}{105} = 42,9 \%$$

- ▶ pravděpodobnost $P(D)$, že zná *aspoň jednu* otázku

$$M_D = M_B + M_C = 50 + 45 = 95 \Rightarrow P(D) = \frac{95}{105} = 90,5 \%$$

- ▶ kontrola: $M_D + M_A = M$

příklad: losování otázek (2)

- ▶ pravděpodobnost $P(B)$, že zná *právě* jednu otázku

$$M_B = \binom{5}{1} \cdot \binom{10}{1} = 5 \cdot 10 = 50 \Rightarrow P(B) = \frac{50}{105} = 47,6 \%$$

- ▶ pravděpodobnost $P(C)$, že zná *obě* otázky (*právě dvě*)

$$M_C = \binom{5}{0} \cdot \binom{10}{2} = 1 \cdot \frac{10 \cdot 9}{2 \cdot 1} = 45 \Rightarrow P(C) = \frac{45}{105} = 42,9 \%$$

- ▶ pravděpodobnost $P(D)$, že zná *aspoň jednu* otázku

$$M_D = M_B + M_C = 50 + 45 = 95 \Rightarrow P(D) = \frac{95}{105} = 90,5 \%$$

- ▶ kontrola: $M_D + M_A = M$

příklad: losování otázek (2)

- ▶ pravděpodobnost $P(B)$, že zná *právě* jednu otázku

$$M_B = \binom{5}{1} \cdot \binom{10}{1} = 5 \cdot 10 = 50 \Rightarrow P(B) = \frac{50}{105} = 47,6 \%$$

- ▶ pravděpodobnost $P(C)$, že zná *obě* otázky (*právě dvě*)

$$M_C = \binom{5}{0} \cdot \binom{10}{2} = 1 \cdot \frac{10 \cdot 9}{2 \cdot 1} = 45 \Rightarrow P(C) = \frac{45}{105} = 42,9 \%$$

- ▶ pravděpodobnost $P(D)$, že zná *aspoň jednu* otázku

$$M_D = M_B + M_C = 50 + 45 = 95 \Rightarrow P(D) = \frac{95}{105} = 90,5 \%$$

- ▶ kontrola: $M_D + M_A = M$

příklad: losování otázek (2)

- ▶ pravděpodobnost $P(B)$, že zná *právě* jednu otázku

$$M_B = \binom{5}{1} \cdot \binom{10}{1} = 5 \cdot 10 = 50 \Rightarrow P(B) = \frac{50}{105} = 47,6 \%$$

- ▶ pravděpodobnost $P(C)$, že zná *obě* otázky (*právě dvě*)

$$M_C = \binom{5}{0} \cdot \binom{10}{2} = 1 \cdot \frac{10 \cdot 9}{2 \cdot 1} = 45 \Rightarrow P(C) = \frac{45}{105} = 42,9 \%$$

- ▶ pravděpodobnost $P(D)$, že zná *aspoň jednu* otázku

$$M_D = M_B + M_C = 50 + 45 = 95 \Rightarrow P(D) = \frac{95}{105} = 90,5 \%$$

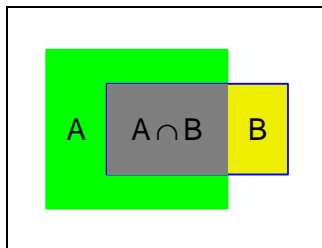
- ▶ kontrola: $M_D + M_A = M$

pravidla pro pravděpodobnost (1)

- ▶ **sjednocení** jevů $A \cup B$: platí **A nebo B** (aspoň jeden z jevů A, B)
- ▶ **průnik** $A \cap B$: platí **A a současně B** (oba jevy A, B současně)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- ▶ Vennův diagram



$A \cup B$ = celá vybarvená plocha

$P(A) = 0,42$ = zelená + šedivá plocha

$P(B) = 0,24$ = žlutá + šedivá plocha

$P(A \cap B) = 0,16$ = šedivá plocha

$P(A) + P(B)$ = zelená + žlutá

+ 2 · šedivá plocha

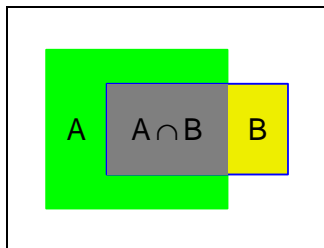
$P(A \cup B) = 0,42 + 0,24 - 0,16 = 0,50$

pravidla pro pravděpodobnost (1)

- ▶ **sjednocení** jevů $A \cup B$: platí **A nebo B** (aspoň jeden z jevů A, B)
- ▶ **průnik** $A \cap B$: platí **A a současně B** (oba jevy A, B současně)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- ▶ Vennův diagram



$A \cup B$ = celá vybarvená plocha

$P(A) = 0,42$ = zelená + šedivá plocha

$P(B) = 0,24$ = žlutá + šedivá plocha

$P(A \cap B) = 0,16$ = šedivá plocha

$P(A) + P(B)$ = zelená + žlutá

+ 2 · šedivá plocha

$P(A \cup B) = 0,42 + 0,24 - 0,16 = 0,50$

pravidla pro pravděpodobnost (2)

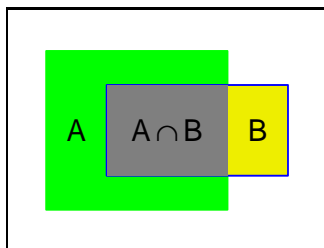
- ▶ **neslučitelné jevy**: nemohou nastat nikdy současně, navzájem se vylučují; pro neslučitelné jevy platí

$$P(A \cup B) = P(A) + P(B)$$

- ▶ **podmíněná pravděpodobnost** pravděpodobnost jevu A , když už jev B nastal:

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

- ▶ Vennův diagram



$P(B) = 0,24 =$ žlutá + šedivá plocha

$P(A \cap B) = 0,16 =$ šedivá plocha

$P(A|B) =$ šedivá vzhledem k (žlutá + šedivá)

$P(A|B) = 0,16/0,24 = 0,67$, ale $P(A) = 0,42$

pravidla pro pravděpodobnost (2)

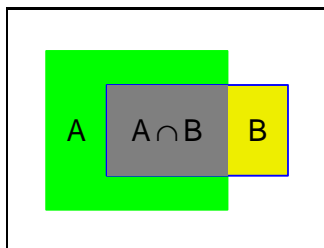
- ▶ **neslučitelné jevy**: nemohou nastat nikdy současně, navzájem se vylučují; pro neslučitelné jevy platí

$$P(A \cup B) = P(A) + P(B)$$

- ▶ **podmíněná pravděpodobnost** pravděpodobnost jevu A , když už jev B nastal:

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

- ▶ Vennův diagram



$P(B) = 0,24 =$ žlutá + šedivá plocha

$P(A \cap B) = 0,16 =$ šedivá plocha

$P(A|B) =$ šedivá vzhledem k (žlutá + šedivá)

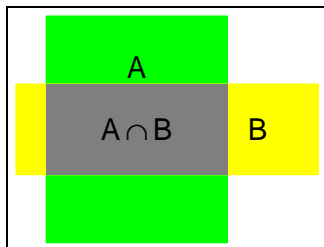
$P(A|B) = 0,16/0,24 = 0,67$, ale $P(A) = 0,42$

nezávislost náhodných jevů

- ▶ **nezávislé jevy**: výskyt jednoho jevu **neovlivní** pravděpodobnost výskytu druhého (definice **nezávislosti** náhodných jevů):

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

- ▶ Vennův diagram



$$P(A) = 0,60 = \text{zelená} + \text{šedivá}$$

$$P(B) = 0,40 = \text{žlutá} + \text{šedivá plocha}$$

$$P(A \cap B) = 0,24 = \text{šedivá plocha}$$

$$P(A|B) = \text{šedivá vzhledem k (žlutá + šedivá)}$$

$$P(A|B) = 0,24/0,40 = 0,60$$

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

$\Rightarrow A$ a B jsou nezávislé

idealizovaný příklad

- ▶ A – jednička ze statistiky, $P(A) = 0,3$
- ▶ B – jednička z matematiky, $P(B) = 0,2$
- ▶ $A \cap B$ – jednička z obou předmětů, $P(A \cap B) = 0,1$
- ▶ jsou jevy A, B nezávislé? (jsou jedničky ze dvou předmětů nezávislé?) NE, protože $0,3 \cdot 0,2 \neq 0,1$
- ▶ jaká je pst jedničky ze statistiky, když už je z matematiky?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,1}{0,2} = 0,5$$

- ▶ pst jedničky z matematiky, když už je ze statistiky:
 $P(B|A) = 0,1/0,3 = 1/3$
- ▶ pravděpodobnost, že aspoň jedna jednička:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,3 + 0,2 - 0,1 = 0,4$$

idealizovaný příklad

- ▶ A – jednička ze statistiky, $P(A) = 0,3$
- ▶ B – jednička z matematiky, $P(B) = 0,2$
- ▶ $A \cap B$ – jednička z obou předmětů, $P(A \cap B) = 0,1$
- ▶ jsou jevy A, B nezávislé? (jsou jedničky ze dvou předmětů nezávislé?) NE, protože $0,3 \cdot 0,2 \neq 0,1$
- ▶ jaká je pst jedničky ze statistiky, když už je z matematiky?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,1}{0,2} = 0,5$$

- ▶ pst jedničky z matematiky, když už je ze statistiky:
 $P(B|A) = 0,1/0,3 = 1/3$
- ▶ pravděpodobnost, že aspoň jedna jednička:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,3 + 0,2 - 0,1 = 0,4$$

idealizovaný příklad

- ▶ A – jednička ze statistiky, $P(A) = 0,3$
- ▶ B – jednička z matematiky, $P(B) = 0,2$
- ▶ $A \cap B$ – jednička z obou předmětů, $P(A \cap B) = 0,1$
- ▶ jsou jevy A, B nezávislé? (jsou jedničky ze dvou předmětů nezávislé?) NE, protože $0,3 \cdot 0,2 \neq 0,1$
- ▶ jaká je pst jedničky ze statistiky, když už je z matematiky?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,1}{0,2} = 0,5$$

- ▶ pst jedničky z matematiky, když už je ze statistiky:
 $P(B|A) = 0,1/0,3 = 1/3$
- ▶ pravděpodobnost, že aspoň jedna jednička:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,3 + 0,2 - 0,1 = 0,4$$

idealizovaný příklad

- ▶ A – jednička ze statistiky, $P(A) = 0,3$
- ▶ B – jednička z matematiky, $P(B) = 0,2$
- ▶ $A \cap B$ – jednička z obou předmětů, $P(A \cap B) = 0,1$
- ▶ jsou jevy A, B nezávislé? (jsou jedničky ze dvou předmětů nezávislé?)
NE, protože $0,3 \cdot 0,2 \neq 0,1$
- ▶ jaká je pst jedničky ze statistiky, když už je z matematiky?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,1}{0,2} = 0,5$$

- ▶ pst jedničky z matematiky, když už je ze statistiky:
 $P(B|A) = 0,1/0,3 = 1/3$
- ▶ pravděpodobnost, že aspoň jedna jednička:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,3 + 0,2 - 0,1 = 0,4$$

idealizovaný příklad

- ▶ A – jednička ze statistiky, $P(A) = 0,3$
- ▶ B – jednička z matematiky, $P(B) = 0,2$
- ▶ $A \cap B$ – jednička z obou předmětů, $P(A \cap B) = 0,1$
- ▶ jsou jevy A, B nezávislé? (jsou jedničky ze dvou předmětů nezávislé?)
NE, protože $0,3 \cdot 0,2 \neq 0,1$
- ▶ jaká je pst jedničky ze statistiky, když už je z matematiky?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,1}{0,2} = 0,5$$

- ▶ pst jedničky z matematiky, když už je ze statistiky:
 $P(B|A) = 0,1/0,3 = 1/3$
- ▶ pravděpodobnost, že aspoň jedna jednička:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,3 + 0,2 - 0,1 = 0,4$$

idealizovaný příklad

- ▶ A – jednička ze statistiky, $P(A) = 0,3$
- ▶ B – jednička z matematiky, $P(B) = 0,2$
- ▶ $A \cap B$ – jednička z obou předmětů, $P(A \cap B) = 0,1$
- ▶ jsou jevy A, B nezávislé? (jsou jedničky ze dvou předmětů nezávislé?) NE, protože $0,3 \cdot 0,2 \neq 0,1$
- ▶ jaká je pst jedničky ze statistiky, když už je z matematiky?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,1}{0,2} = 0,5$$

- ▶ pst jedničky z matematiky, když už je ze statistiky:
 $P(B|A) = 0,1/0,3 = 1/3$
- ▶ pravděpodobnost, že aspoň jedna jednička:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,3 + 0,2 - 0,1 = 0,4$$

idealizovaný příklad

- ▶ A – jednička ze statistiky, $P(A) = 0,3$
- ▶ B – jednička z matematiky, $P(B) = 0,2$
- ▶ $A \cap B$ – jednička z obou předmětů, $P(A \cap B) = 0,1$
- ▶ jsou jevy A, B nezávislé? (jsou jedničky ze dvou předmětů nezávislé?) NE, protože $0,3 \cdot 0,2 \neq 0,1$
- ▶ jaká je pst jedničky ze statistiky, když už je z matematiky?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,1}{0,2} = 0,5$$

- ▶ pst jedničky z matematiky, když už je ze statistiky:
 $P(B|A) = 0,1/0,3 = 1/3$
- ▶ pravděpodobnost, že aspoň jedna jednička:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,3 + 0,2 - 0,1 = 0,4$$

rozdělení náhodné veličiny

- ▶ **náhodná veličina** – číselně vyjádřený výsledek náhodného pokusu
- ▶ **diskrétní rozdělení** (pro četnosti) určeno seznamem možných hodnot a jejich pravděpodobnostmi:

$$x_1, x_2, \dots$$

$$P(X = x_1), P(X = x_2), \dots$$

- ▶ **spojité rozdělení** (pro spojité měřítko) určeno **distribuční funkcí**

$$F_X(x) = P(X \leq x)$$

nebo **hustotou**

$$f_X(x) = \frac{d}{dx} F_X(x), \quad F_X(x) = \int_{-\infty}^x f_X(t) dt$$

rozdělení náhodné veličiny

- ▶ **náhodná veličina** – číselně vyjádřený výsledek náhodného pokusu
- ▶ **diskrétní rozdělení** (pro četnosti) určeno seznamem možných hodnot a jejich pravděpodobnostmi:

$$x_1, x_2, \dots$$

$$P(X = x_1), P(X = x_2), \dots$$

- ▶ **spojité rozdělení** (pro spojité měřítko) určeno **distribuční funkcí**

$$F_X(x) = P(X \leq x)$$

nebo **hustotou**

$$f_X(x) = \frac{d}{dx} F_X(x), \quad F_X(x) = \int_{-\infty}^x f_X(t) dt$$

rozdělení náhodné veličiny

- ▶ **náhodná veličina** – číselně vyjádřený výsledek náhodného pokusu
- ▶ **diskrétní rozdělení** (pro četnosti) určeno seznamem možných hodnot a jejich pravděpodobnostmi:

$$x_1, x_2, \dots$$

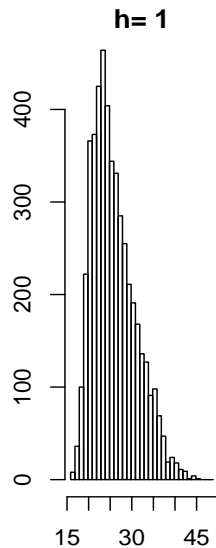
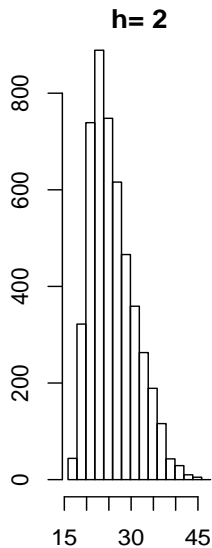
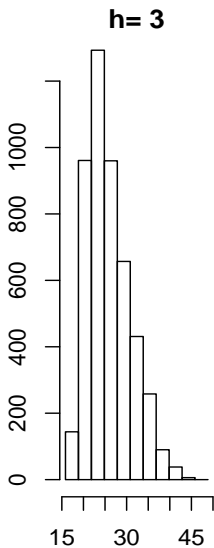
$$P(X = x_1), P(X = x_2), \dots$$

- ▶ **spojité rozdělení** (pro spojité měřítko) určeno **distribuční funkcí**

$$F_X(x) = P(X \leq x)$$

nebo **hustotou**

$$f_X(x) = \frac{d}{dx} F_X(x), \quad F_X(x) = \int_{-\infty}^x f_X(t) dt$$

věk matek ($n=4838$)

- ▶ velká populace, spojitá veličina – intervaly pro třídění mohou být krátké, obálce histogramu relativních četností odpovídá **hustota** $f_X(x)$ [density]
- ▶ podobně kumulativním relativním četnostem odpovídá **distribuční funkce** [distribution function]
- ▶ bezprostředním výběrovým protějškem distribuční funkce je **empirická distribuční funkce**

$$F_n(x) = \frac{\#(x_i \leq x)}{n}$$

- ▶ $x_1^* < x_2^* < \dots < x_m^*$ existující různé hodnoty n_1, n_2, \dots, n_m jejich četnosti ($n = \sum_j n_j$)
 $F_n(x)$ je schodovitá funkce, v bodě x_j^* má skok n_j/n

- ▶ velká populace, spojitá veličina – intervaly pro třídění mohou být krátké, obálce histogramu relativních četností odpovídá **hustota** $f_X(x)$ [density]
- ▶ podobně kumulativním relativním četnostem odpovídá **distribuční funkce** [distribution function]
- ▶ bezprostředním výběrovým protějškem distribuční funkce je **empirická distribuční funkce**

$$F_n(x) = \frac{\#(x_i \leq x)}{n}$$

- ▶ $x_1^* < x_2^* < \dots < x_m^*$ existující různé hodnoty n_1, n_2, \dots, n_m jejich četnosti ($n = \sum_j n_j$)
 $F_n(x)$ je schodovitá funkce, v bodě x_j^* má skok n_j/n

- ▶ velká populace, spojitá veličina – intervaly pro třídění mohou být krátké, obálce histogramu relativních četností odpovídá **hustota** $f_X(x)$ [density]
- ▶ podobně kumulativním relativním četnostem odpovídá **distribuční funkce** [distribution function]
- ▶ bezprostředním výběrovým protějškem distribuční funkce je **empirická distribuční funkce**

$$F_n(x) = \frac{\#(x_i \leq x)}{n}$$

- ▶ $x_1^* < x_2^* < \dots < x_m^*$ existující různé hodnoty n_1, n_2, \dots, n_m jejich četnosti ($n = \sum_j n_j$)
 $F_n(x)$ je schodovitá funkce, v bodě x_j^* má skok n_j/n

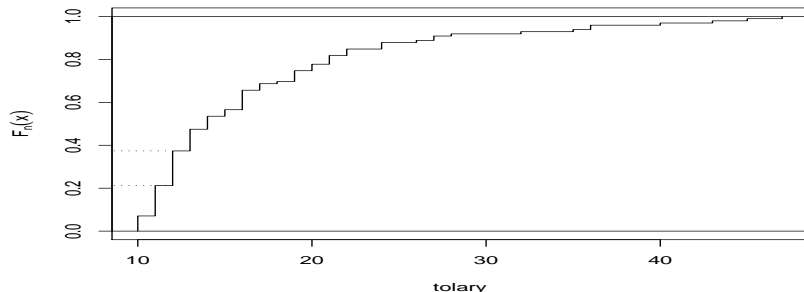
- ▶ velká populace, spojitá veličina – intervaly pro třídění mohou být krátké, obálce histogramu relativních četností odpovídá **hustota** $f_X(x)$ [density]
- ▶ podobně kumulativním relativním četnostem odpovídá **distribuční funkce** [distribution function]
- ▶ bezprostředním výběrovým protějškem distribuční funkce je **empirická distribuční funkce**

$$F_n(x) = \frac{\#(x_i \leq x)}{n}$$

- ▶ $x_1^* < x_2^* < \dots < x_m^*$ existující různé hodnoty n_1, n_2, \dots, n_m jejich četnosti ($n = \sum_j n_j$)
 $F_n(x)$ je schodovitá funkce, v bodě x_j^* má skok n_j/n

kumulativní distribuční funkce (tolary)

skoky odpovídají četnostem, např. ve 12 je skok z 0,21 na 0,37 o $16/99=0,16$



x_j^*	10	11	12	13	14	15	16	17	18	19	20		
n_j	7	14	16	10	6	3	9	3	1	5	3		
N_j	7	21	37	47	53	56	65	68	69	74	77		
x_j^*	21	22	24	26	27	28	32	35	36	40	43	45	47
n_j	4	3	3	1	2	1	1	1	2	1	1	1	1
N_j	81	84	87	88	90	91	92	93	95	96	97	98	99

příklad diskrétního rozdělení: známky u zkoušky

X, Y známky ze dvou předmětů

známka k	1	2	3	4
$P(X = k)$	0,3	0,4	0,2	0,1
$P(Y = k)$	0,3	0,3	0,2	0,2

- ▶ z tabulky *nic* nepoznáme o případné závislosti X, Y
- ▶ jak jedním číslem charakterizovat úroveň známek?
- ▶ obyčejný průměr možných hodnot by X, Y nerozlišil
- ▶ použijme **vážený průměr**, kde vahami známek jsou **pravděpodobnosti možných hodnot**
- ▶ dostaneme tak **střední hodnoty** X a Y (**populační průměry**)

$$\mu_X = 1 \cdot 0,3 + 2 \cdot 0,4 + 3 \cdot 0,2 + 4 \cdot 0,1 = 2,1$$

$$\mu_Y = 1 \cdot 0,3 + 2 \cdot 0,3 + 3 \cdot 0,2 + 4 \cdot 0,2 = 2,3$$

příklad diskrétního rozdělení: známky u zkoušky

X, Y známky ze dvou předmětů

známka k	1	2	3	4
$P(X = k)$	0,3	0,4	0,2	0,1
$P(Y = k)$	0,3	0,3	0,2	0,2

- ▶ z tabulky *nic* nepoznáme o případné závislosti X, Y
- ▶ jak jedním číslem charakterizovat úroveň známek?
- ▶ obyčejný průměr možných hodnot by X, Y nerozlišil
- ▶ použijme **vážený průměr**, kde vahami známek jsou **pravděpodobnosti možných hodnot**
- ▶ dostaneme tak **střední hodnoty** X a Y (**populační průměry**)

$$\mu_X = 1 \cdot 0,3 + 2 \cdot 0,4 + 3 \cdot 0,2 + 4 \cdot 0,1 = 2,1$$

$$\mu_Y = 1 \cdot 0,3 + 2 \cdot 0,3 + 3 \cdot 0,2 + 4 \cdot 0,2 = 2,3$$

příklad diskrétního rozdělení: známky u zkoušky

X, Y známky ze dvou předmětů

známka k	1	2	3	4
$P(X = k)$	0,3	0,4	0,2	0,1
$P(Y = k)$	0,3	0,3	0,2	0,2

- ▶ z tabulky *nic* nepoznáme o případné závislosti X, Y
- ▶ jak jedním číslem charakterizovat úroveň známek?
- ▶ obyčejný průměr možných hodnot by X, Y nerozlišil
- ▶ použijme **vážený průměr**, kde vahami známek jsou **pravděpodobnosti možných hodnot**
- ▶ dostaneme tak **střední hodnoty** X a Y (**populační průměry**)

$$\mu_X = 1 \cdot 0,3 + 2 \cdot 0,4 + 3 \cdot 0,2 + 4 \cdot 0,1 = 2,1$$

$$\mu_Y = 1 \cdot 0,3 + 2 \cdot 0,3 + 3 \cdot 0,2 + 4 \cdot 0,2 = 2,3$$

příklad diskrétního rozdělení: známky u zkoušky

X, Y známky ze dvou předmětů

známka k	1	2	3	4
$P(X = k)$	0,3	0,4	0,2	0,1
$P(Y = k)$	0,3	0,3	0,2	0,2

- ▶ z tabulky *nic* nepoznáme o případné závislosti X, Y
- ▶ jak jedním číslem charakterizovat úroveň známek?
- ▶ obyčejný průměr možných hodnot by X, Y nerozlišil
- ▶ použijme **vážený průměr**, kde vahami známek jsou **pravděpodobnosti možných hodnot**
- ▶ dostaneme tak **střední hodnoty** X a Y (**populační průměry**)

$$\mu_X = 1 \cdot 0,3 + 2 \cdot 0,4 + 3 \cdot 0,2 + 4 \cdot 0,1 = 2,1$$

$$\mu_Y = 1 \cdot 0,3 + 2 \cdot 0,3 + 3 \cdot 0,2 + 4 \cdot 0,2 = 2,3$$

příklad diskrétního rozdělení: známky u zkoušky

X, Y známky ze dvou předmětů

známka k	1	2	3	4
$P(X = k)$	0,3	0,4	0,2	0,1
$P(Y = k)$	0,3	0,3	0,2	0,2

- ▶ z tabulky *nic* nepoznáme o případné závislosti X, Y
- ▶ jak jedním číslem charakterizovat úroveň známek?
- ▶ obyčejný průměr možných hodnot by X, Y nerozlišil
- ▶ použijme **vážený průměr**, kde vahami známek jsou **pravděpodobnosti možných hodnot**
- ▶ dostaneme tak **střední hodnoty** X a Y (**populační průměry**)

$$\mu_X = 1 \cdot 0,3 + 2 \cdot 0,4 + 3 \cdot 0,2 + 4 \cdot 0,1 = 2,1$$

$$\mu_Y = 1 \cdot 0,3 + 2 \cdot 0,3 + 3 \cdot 0,2 + 4 \cdot 0,2 = 2,3$$

charakteristiky rozdělení náhodné veličiny (1)

- ▶ **střední hodnota** náhodné veličiny X (populační průměr)
- ▶ je to **vážený průměr možných hodnot**
- ▶ vahami jsou pravděpodobnosti hodnot

$$\mu_X = E X = x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \dots = \sum_j x_j \cdot P(X = x_j)$$

- ▶ operátor E (expectation) aplikovaný na náhodnou veličinu X spočítá vážený průměr jejích hodnot, vahami jsou u diskrétního rozdělení pravděpodobnosti těchto hodnot
- ▶ pro spojité rozdělení

$$\mu_X = E X = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

charakteristiky rozdělení náhodné veličiny (1)

- ▶ **střední hodnota** náhodné veličiny X (populační průměr)
- ▶ je to **vážený průměr možných hodnot**
- ▶ vahami jsou pravděpodobnosti hodnot

$$\mu_X = E X = x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \dots = \sum_j x_j \cdot P(X = x_j)$$

- ▶ operátor E (expectation) aplikovaný na náhodnou veličinu X spočítá vážený průměr jejích hodnot, vahami jsou u diskrétního rozdělení pravděpodobnosti těchto hodnot
- ▶ pro spojité rozdělení

$$\mu_X = E X = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

charakteristiky rozdělení náhodné veličiny (1)

- ▶ **střední hodnota** náhodné veličiny X (populační průměr)
- ▶ je to **vážený průměr možných hodnot**
- ▶ vahami jsou pravděpodobnosti hodnot

$$\mu_X = E X = x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \dots = \sum_j x_j \cdot P(X = x_j)$$

- ▶ operátor E (expectation) aplikovaný na náhodnou veličinu X spočítá vážený průměr jejích hodnot, vahami jsou u diskrétního rozdělení pravděpodobnosti těchto hodnot
- ▶ pro spojité rozdělení

$$\mu_X = E X = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

charakteristiky rozdělení náhodné veličiny (1)

- ▶ **střední hodnota** náhodné veličiny X (populační průměr)
- ▶ je to **vážený průměr možných hodnot**
- ▶ vahami jsou pravděpodobnosti hodnot

$$\mu_X = E X = x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \dots = \sum_j x_j \cdot P(X = x_j)$$

- ▶ operátor E (expectation) aplikovaný na náhodnou veličinu X spočítá vážený průměr jejích hodnot, vahami jsou u diskrétního rozdělení pravděpodobnosti těchto hodnot
- ▶ pro spojité rozdělení

$$\mu_X = E X = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

charakteristiky rozdělení náhodné veličiny (1)

- ▶ **střední hodnota** náhodné veličiny X (populační průměr)
- ▶ je to **vážený průměr možných hodnot**
- ▶ vahami jsou pravděpodobnosti hodnot

$$\mu_X = E X = x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \dots = \sum_j x_j \cdot P(X = x_j)$$

- ▶ operátor E (expectation) aplikovaný na náhodnou veličinu X spočítá vážený průměr jejích hodnot, vahami jsou u diskrétního rozdělení pravděpodobnosti těchto hodnot
- ▶ pro spojité rozdělení

$$\mu_X = E X = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

- ▶ **střední hodnota funkce** $Y = g(X)$ náhodné veličiny X
vážený průměr **funkčních hodnot**

$$E Y = E g(X) = \sum_k g(x_k)P(X = x_k)$$

resp. pro spojité rozdělení

$$E Y = E g(X) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

- ▶ **populační medián** $\tilde{\mu}$ spojitého rozdělení

$$F_X(\tilde{\mu}) = P(X \leq \tilde{\mu}) = 0,5$$

\tilde{x} číslo, které dělí možné hodnoty náhodné veličiny na dva stejně pravděpodobné intervaly hodnot větších a menších

- ▶ **střední hodnota funkce** $Y = g(X)$ náhodné veličiny X
vážený průměr **funkčních hodnot**

$$E Y = E g(X) = \sum_k g(x_k)P(X = x_k)$$

resp. pro spojité rozdělení

$$E Y = E g(X) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

- ▶ **populační medián** $\tilde{\mu}$ spojitého rozdělení

$$F_X(\tilde{\mu}) = P(X \leq \tilde{\mu}) = 0,5$$

\tilde{x} číslo, které dělí možné hodnoty náhodné veličiny na dva stejně pravděpodobné intervaly hodnot větších a menších

příklad diskrétního rozdělení: známka u zkoušky

známka k	1	2	3	4	μ	σ^2	σ
$P(X = k)$	0,3	0,4	0,2	0,1	2,1	0,89	0,943
$P(Y = k)$	0,3	0,3	0,2	0,2	2,3	1,21	1,100

- ▶ jedním číslem charakterizovat kolísání známek (**variabilitu**)
- ▶ (**populační**) rozptyl = vážený průměr čtverců vzdáleností od střední hodnoty
- ▶ vahami jsou pravděpodobnosti

$$\sigma_X^2 = (1 - 2,1)^2 \cdot 0,3 + (2 - 2,1)^2 \cdot 0,4 + (3 - 2,1)^2 \cdot 0,2 + (4 - 2,1)^2 \cdot 0,1 = 0,89 = 0,943^2$$

$$\sigma_Y^2 = (1 - 2,3)^2 \cdot 0,3 + (2 - 2,3)^2 \cdot 0,3 + (3 - 2,3)^2 \cdot 0,2 + (4 - 2,3)^2 \cdot 0,2 = 1,21 = 1,1^2$$

příklad diskrétního rozdělení: známka u zkoušky

známka k	1	2	3	4	μ	σ^2	σ
$P(X = k)$	0,3	0,4	0,2	0,1	2,1	0,89	0,943
$P(Y = k)$	0,3	0,3	0,2	0,2	2,3	1,21	1,100

- ▶ jedním číslem charakterizovat kolísání známek (**variabilitu**)
- ▶ **(populační) rozptyl = vážený průměr čtverců vzdáleností od střední hodnoty**
- ▶ vahami jsou pravděpodobnosti

$$\sigma_X^2 = (1 - 2,1)^2 \cdot 0,3 + (2 - 2,1)^2 \cdot 0,4 + (3 - 2,1)^2 \cdot 0,2 + (4 - 2,1)^2 \cdot 0,1 = 0,89 = 0,943^2$$

$$\sigma_Y^2 = (1 - 2,3)^2 \cdot 0,3 + (2 - 2,3)^2 \cdot 0,3 + (3 - 2,3)^2 \cdot 0,2 + (4 - 2,3)^2 \cdot 0,2 = 1,21 = 1,1^2$$

příklad diskrétního rozdělení: známka u zkoušky

známka k	1	2	3	4	μ	σ^2	σ
$P(X = k)$	0,3	0,4	0,2	0,1	2,1	0,89	0,943
$P(Y = k)$	0,3	0,3	0,2	0,2	2,3	1,21	1,100

- ▶ jedním číslem charakterizovat kolísání známek (**variabilitu**)
- ▶ **(populační) rozptyl = vážený průměr čtverců vzdáleností od střední hodnoty**
- ▶ vahami jsou pravděpodobnosti

$$\sigma_X^2 = (1 - 2,1)^2 \cdot 0,3 + (2 - 2,1)^2 \cdot 0,4 + (3 - 2,1)^2 \cdot 0,2 + (4 - 2,1)^2 \cdot 0,1 = 0,89 = 0,943^2$$

$$\sigma_Y^2 = (1 - 2,3)^2 \cdot 0,3 + (2 - 2,3)^2 \cdot 0,3 + (3 - 2,3)^2 \cdot 0,2 + (4 - 2,3)^2 \cdot 0,2 = 1,21 = 1,1^2$$

(populační) rozptyl náhodné veličiny X

- ▶ vážený průměr čtverců vzdáleností možných hodnot od střední hodnoty

$$\begin{aligned}\sigma_X^2 &= E(X - \mu_X)^2 \\ &= (x_1 - \mu_X)^2 P(X = x_1) + (x_2 - \mu_X)^2 P(X = x_2) + \dots \\ &= \sum_j (x_j - \mu_X)^2 P(X = x_j)\end{aligned}$$

$$\sigma_X^2 = E(X - \mu_X)^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$$

- ▶ (populační) směrodatná odchylka odmocnina z (populačního) rozptylu

$$\sigma_X = \sqrt{\sigma_X^2}$$

(populační) rozptyl náhodné veličiny X

- ▶ vážený průměr čtverců vzdáleností možných hodnot od střední hodnoty

$$\begin{aligned}\sigma_X^2 &= E(X - \mu_X)^2 \\ &= (x_1 - \mu_X)^2 P(X = x_1) + (x_2 - \mu_X)^2 P(X = x_2) + \dots \\ &= \sum_j (x_j - \mu_X)^2 P(X = x_j)\end{aligned}$$

$$\sigma_X^2 = E(X - \mu_X)^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$$

- ▶ **(populační) směrodatná odchylka** odmocnina z (populačního) rozptylu

$$\sigma_X = \sqrt{\sigma_X^2}$$

vlastnosti střední hodnoty a rozptylu

X, Y – náhodné veličiny, a, b konstanty, $b > 0$

$$\mu_{a+X} = E(a + X) = a + EX = a + \mu_X$$

$$\mu_{b \cdot X} = E(b \cdot X) = b \cdot EX = b \cdot \mu_X$$

$$\mu_{X+Y} = E(X + Y) = EX + EY = \mu_X + \mu_Y$$

► Návrat k průměru $\sigma_{a+X}^2 = \sigma_X^2, \quad \sigma_{a+X} = \sigma_X$

$$\sigma_{b \cdot X}^2 = b^2 \sigma_X^2, \quad \sigma_{b \cdot X} = |b| \sigma_X$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

► Návrat k rozptylu $\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y)$ **kovariance** X, Y

$$= (x_1 - \mu_X)(y_1 - \mu_Y)P(X = x_1, Y = y_1)$$

$$+ (x_1 - \mu_X)(y_2 - \mu_Y)P(X = x_1, Y = y_2) + \dots$$

(sčítá se přes všechny možné dvojice)

vlastnosti střední hodnoty a rozptylu

X, Y – náhodné veličiny, a, b konstanty, $b > 0$

$$\mu_{a+X} = E(a + X) = a + EX = a + \mu_X$$

$$\mu_{b \cdot X} = E(b \cdot X) = b \cdot EX = b \cdot \mu_X$$

$$\mu_{X+Y} = E(X + Y) = EX + EY = \mu_X + \mu_Y$$

► Návrat k průměru $\sigma_{a+X}^2 = \sigma_X^2, \quad \sigma_{a+X} = \sigma_X$

$$\sigma_{b \cdot X}^2 = b^2 \sigma_X^2, \quad \sigma_{b \cdot X} = |b| \sigma_X$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

► Návrat k rozptylu $\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y)$ **kovariance** X, Y

$$= (x_1 - \mu_X)(y_1 - \mu_Y)P(X = x_1, Y = y_1)$$

$$+ (x_1 - \mu_X)(y_2 - \mu_Y)P(X = x_1, Y = y_2) + \dots$$

(sčítá se přes všechny možné dvojice)

vlastnosti střední hodnoty a rozptylu

X, Y – náhodné veličiny, a, b konstanty, $b > 0$

$$\mu_{a+X} = E(a + X) = a + EX = a + \mu_X$$

$$\mu_{b \cdot X} = E(b \cdot X) = b \cdot EX = b \cdot \mu_X$$

$$\mu_{X+Y} = E(X + Y) = EX + EY = \mu_X + \mu_Y$$

► Návrat k průměru $\sigma_{a+X}^2 = \sigma_X^2, \quad \sigma_{a+X} = \sigma_X$

$$\sigma_{b \cdot X}^2 = b^2 \sigma_X^2, \quad \sigma_{b \cdot X} = |b| \sigma_X$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

► Návrat k rozptylu $\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y)$ **kovariance** X, Y

$$= (x_1 - \mu_X)(y_1 - \mu_Y)P(X = x_1, Y = y_1)$$

$$+ (x_1 - \mu_X)(y_2 - \mu_Y)P(X = x_1, Y = y_2) + \dots$$

(sčítá se přes všechny možné dvojice)

vlastnosti střední hodnoty a rozptylu

X, Y – náhodné veličiny, a, b konstanty, $b > 0$

$$\mu_{a+X} = E(a + X) = a + EX = a + \mu_X$$

$$\mu_{b \cdot X} = E(b \cdot X) = b \cdot EX = b \cdot \mu_X$$

$$\mu_{X+Y} = E(X + Y) = EX + EY = \mu_X + \mu_Y$$

► Návrat k průměru $\sigma_{a+X}^2 = \sigma_X^2, \quad \sigma_{a+X} = \sigma_X$

$$\sigma_{b \cdot X}^2 = b^2 \sigma_X^2, \quad \sigma_{b \cdot X} = |b| \sigma_X$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

► Návrat k rozptylu $\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y)$ **kovariance** X, Y

$$= (x_1 - \mu_X)(y_1 - \mu_Y)P(X = x_1, Y = y_1)$$

$$+ (x_1 - \mu_X)(y_2 - \mu_Y)P(X = x_1, Y = y_2) + \dots$$

(sčítá se přes všechny možné dvojice)

vlastnosti střední hodnoty a rozptylu

X, Y – náhodné veličiny, a, b konstanty, $b > 0$

$$\mu_{a+X} = E(a + X) = a + EX = a + \mu_X$$

$$\mu_{b \cdot X} = E(b \cdot X) = b \cdot EX = b \cdot \mu_X$$

$$\mu_{X+Y} = E(X + Y) = EX + EY = \mu_X + \mu_Y$$

► Návrat k průměru $\sigma_{a+X}^2 = \sigma_X^2, \quad \sigma_{a+X} = \sigma_X$

$$\sigma_{b \cdot X}^2 = b^2 \sigma_X^2, \quad \sigma_{b \cdot X} = |b| \sigma_X$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

► Návrat k rozptylu $\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y)$ **kovariance** X, Y

$$= (x_1 - \mu_X)(y_1 - \mu_Y)P(X = x_1, Y = y_1)$$

$$+ (x_1 - \mu_X)(y_2 - \mu_Y)P(X = x_1, Y = y_2) + \dots$$

(sčítá se přes všechny možné dvojice)

vlastnosti střední hodnoty a rozptylu

X, Y – náhodné veličiny, a, b konstanty, $b > 0$

$$\mu_{a+X} = E(a + X) = a + EX = a + \mu_X$$

$$\mu_{b \cdot X} = E(b \cdot X) = b \cdot EX = b \cdot \mu_X$$

$$\mu_{X+Y} = E(X + Y) = EX + EY = \mu_X + \mu_Y$$

► Návrat k průměru $\sigma_{a+X}^2 = \sigma_X^2, \quad \sigma_{a+X} = \sigma_X$

$$\sigma_{b \cdot X}^2 = b^2 \sigma_X^2, \quad \sigma_{b \cdot X} = |b| \sigma_X$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

► Návrat k rozptylu $\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y)$ **kovariance** X, Y

$$= (x_1 - \mu_X)(y_1 - \mu_Y)P(X = x_1, Y = y_1)$$

$$+ (x_1 - \mu_X)(y_2 - \mu_Y)P(X = x_1, Y = y_2) + \dots$$

(sčítá se přes všechny možné dvojice)

vlastnosti střední hodnoty a rozptylu

X, Y – náhodné veličiny, a, b konstanty, $b > 0$

$$\mu_{a+X} = E(a + X) = a + EX = a + \mu_X$$

$$\mu_{b \cdot X} = E(b \cdot X) = b \cdot EX = b \cdot \mu_X$$

$$\mu_{X+Y} = E(X + Y) = EX + EY = \mu_X + \mu_Y$$

► Návrat k průměru $\sigma_{a+X}^2 = \sigma_X^2, \quad \sigma_{a+X} = \sigma_X$

$$\sigma_{b \cdot X}^2 = b^2 \sigma_X^2, \quad \sigma_{b \cdot X} = |b| \sigma_X$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

► Návrat k rozptylu $\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y)$ **kovariance** X, Y

$$= (x_1 - \mu_X)(y_1 - \mu_Y)P(X = x_1, Y = y_1)$$

$$+ (x_1 - \mu_X)(y_2 - \mu_Y)P(X = x_1, Y = y_2) + \dots$$

(sčítá se přes všechny možné dvojice)

vlastnosti střední hodnoty a rozptylu

X, Y – náhodné veličiny, a, b konstanty, $b > 0$

$$\mu_{a+X} = E(a + X) = a + EX = a + \mu_X$$

$$\mu_{b \cdot X} = E(b \cdot X) = b \cdot EX = b \cdot \mu_X$$

$$\mu_{X+Y} = E(X + Y) = EX + EY = \mu_X + \mu_Y$$

▶ Návrat k průměru $\sigma_{a+X}^2 = \sigma_X^2, \quad \sigma_{a+X} = \sigma_X$

$$\sigma_{b \cdot X}^2 = b^2 \sigma_X^2, \quad \sigma_{b \cdot X} = |b| \sigma_X$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

▶ Návrat k rozptylu $\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y)$ **kovariance** X, Y

$$= (x_1 - \mu_X)(y_1 - \mu_Y)P(X = x_1, Y = y_1)$$

$$+ (x_1 - \mu_X)(y_2 - \mu_Y)P(X = x_1, Y = y_2) + \dots$$

(sčítá se přes všechny možné dvojice)

vlastnosti střední hodnoty a rozptylu

X, Y – náhodné veličiny, a, b konstanty, $b > 0$

$$\mu_{a+X} = E(a + X) = a + EX = a + \mu_X$$

$$\mu_{b \cdot X} = E(b \cdot X) = b \cdot EX = b \cdot \mu_X$$

$$\mu_{X+Y} = E(X + Y) = EX + EY = \mu_X + \mu_Y$$

► Návrat k průměru $\sigma_{a+X}^2 = \sigma_X^2, \quad \sigma_{a+X} = \sigma_X$

$$\sigma_{b \cdot X}^2 = b^2 \sigma_X^2, \quad \sigma_{b \cdot X} = |b| \sigma_X$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

► Návrat k rozptylu $\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y)$ **kovariance** X, Y

$$= (x_1 - \mu_X)(y_1 - \mu_Y)P(X = x_1, Y = y_1)$$

$$+ (x_1 - \mu_X)(y_2 - \mu_Y)P(X = x_1, Y = y_2) + \dots$$

(sčítá se přes všechny možné dvojice)

vlastnosti střední hodnoty a rozptylu

X, Y – náhodné veličiny, a, b konstanty, $b > 0$

$$\mu_{a+X} = E(a + X) = a + EX = a + \mu_X$$

$$\mu_{b \cdot X} = E(b \cdot X) = b \cdot EX = b \cdot \mu_X$$

$$\mu_{X+Y} = E(X + Y) = EX + EY = \mu_X + \mu_Y$$

▶ Návrat k průměru $\sigma_{a+X}^2 = \sigma_X^2, \quad \sigma_{a+X} = \sigma_X$

$$\sigma_{b \cdot X}^2 = b^2 \sigma_X^2, \quad \sigma_{b \cdot X} = |b| \sigma_X$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

▶ Návrat k rozptylu $\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y)$ **kovariance** X, Y

$$= (x_1 - \mu_X)(y_1 - \mu_Y)P(X = x_1, Y = y_1)$$

$$+ (x_1 - \mu_X)(y_2 - \mu_Y)P(X = x_1, Y = y_2) + \dots$$

(sčítá se přes všechny možné dvojice)

vlastnosti střední hodnoty a rozptylu

X, Y – náhodné veličiny, a, b konstanty, $b > 0$

$$\mu_{a+X} = E(a + X) = a + EX = a + \mu_X$$

$$\mu_{b \cdot X} = E(b \cdot X) = b \cdot EX = b \cdot \mu_X$$

$$\mu_{X+Y} = E(X + Y) = EX + EY = \mu_X + \mu_Y$$

► Návrat k průměru $\sigma_{a+X}^2 = \sigma_X^2, \quad \sigma_{a+X} = \sigma_X$

$$\sigma_{b \cdot X}^2 = b^2 \sigma_X^2, \quad \sigma_{b \cdot X} = |b| \sigma_X$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

► Návrat k rozptylu $\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y)$ **kovariance** X, Y

$$= (x_1 - \mu_X)(y_1 - \mu_Y)P(X = x_1, Y = y_1)$$

$$+ (x_1 - \mu_X)(y_2 - \mu_Y)P(X = x_1, Y = y_2) + \dots$$

(sčítá se přes všechny možné dvojice)

nezávislé náhodné veličiny

- ▶ připomeňme: náhodné jevy A, B jsou nezávislé, když

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

- ▶ náhodné veličiny X, Y jsou **nezávislé**, když pro **všechny dvojice** možných hodnot (x_i, y_j) platí

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

- ▶ X a Y jsou tedy nezávislé, jsou-li nezávislé jevy $A = \{\text{tvrzení o } X\}$ a $B = \{\text{tvrzení o } Y\}$
- ▶ jsou-li X, Y nezávislé, pak

$$\sigma_{X,Y} = 0, \quad \text{tedy} \quad \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

- ▶ pro **nezávislé** náhodné veličiny platí:
rozptyl součtu = součet rozptylů

nezávislé náhodné veličiny

- ▶ připomeňme: náhodné jevy A, B jsou nezávislé, když

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

- ▶ náhodné veličiny X, Y jsou **nezávislé**, když pro **všechny dvojice** možných hodnot (x_i, y_j) platí

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

- ▶ X a Y jsou tedy nezávislé, jsou-li nezávislé jevy $A = \{\text{tvrzení o } X\}$ a $B = \{\text{tvrzení o } Y\}$
- ▶ jsou-li X, Y nezávislé, pak

$$\sigma_{X,Y} = 0, \quad \text{tedy} \quad \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

- ▶ pro **nezávislé** náhodné veličiny platí:
rozptyl součtu = součet rozptylů

nezávislé náhodné veličiny

- ▶ připomeňme: náhodné jevy A, B jsou nezávislé, když

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

- ▶ náhodné veličiny X, Y jsou **nezávislé**, když pro **všechny dvojice** možných hodnot (x_i, y_j) platí

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

- ▶ X a Y jsou tedy nezávislé, jsou-li nezávislé jevy $A = \{\text{tvrzení o } X\}$ a $B = \{\text{tvrzení o } Y\}$
- ▶ jsou-li X, Y nezávislé, pak

$$\sigma_{X,Y} = 0, \quad \text{tedy} \quad \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

- ▶ pro **nezávislé** náhodné veličiny platí:
rozptyl součtu = součet rozptylů

nezávislé náhodné veličiny

- ▶ připomeňme: náhodné jevy A, B jsou nezávislé, když

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

- ▶ náhodné veličiny X, Y jsou **nezávislé**, když pro **všechny dvojice** možných hodnot (x_i, y_j) platí

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

- ▶ X a Y jsou tedy nezávislé, jsou-li nezávislé jevy $A = \{\text{tvrzení o } X\}$ a $B = \{\text{tvrzení o } Y\}$
- ▶ jsou-li X, Y nezávislé, pak

$$\sigma_{X,Y} = 0, \quad \text{tedy} \quad \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

- ▶ pro **nezávislé** náhodné veličiny platí:
rozptyl součtu = součet rozptylů

nezávislé náhodné veličiny

- ▶ připomeňme: náhodné jevy A, B jsou nezávislé, když

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

- ▶ náhodné veličiny X, Y jsou **nezávislé**, když pro **všechny dvojice** možných hodnot (x_i, y_j) platí

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

- ▶ X a Y jsou tedy nezávislé, jsou-li nezávislé jevy $A = \{\text{tvrzení o } X\}$ a $B = \{\text{tvrzení o } Y\}$
- ▶ jsou-li X, Y nezávislé, pak

$$\sigma_{X,Y} = 0, \quad \text{tedy} \quad \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

- ▶ pro **nezávislé** náhodné veličiny platí:
rozptyl součtu = součet rozptylů

(populační) korelační koeficient

- ▶ Pearsonův korelační koeficient

$$r_{x,y} = \frac{s_{xy}}{s_x s_y}$$

- ▶ výběrová kovariance dána vztahem (str. 59)

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- ▶ populační protějšek

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- ▶ ρ_{XY} má stejné vlastnosti jako r_{xy} , zejména platí $|\rho_{XY}| \leq 1$
- ▶ pro **nezávislé** náhodné veličiny X, Y je vždy $\rho_{XY} = 0$

(populační) korelační koeficient

- ▶ Pearsonův korelační koeficient

$$r_{x,y} = \frac{s_{xy}}{s_x s_y}$$

- ▶ výběrová kovariance dána vztahem (str. 59)

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- ▶ populační protějšek

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- ▶ ρ_{XY} má stejné vlastnosti jako r_{xy} , zejména platí $|\rho_{XY}| \leq 1$
- ▶ pro **nezávislé** náhodné veličiny X, Y je vždy $\rho_{XY} = 0$

(populační) korelační koeficient

- ▶ Pearsonův korelační koeficient

$$r_{x,y} = \frac{s_{xy}}{s_x s_y}$$

- ▶ výběrová kovariance dána vztahem (str. 59)

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- ▶ **populační protějšek**

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- ▶ ρ_{XY} má stejné vlastnosti jako r_{xy} , zejména platí $|\rho_{XY}| \leq 1$
- ▶ pro **nezávislé** náhodné veličiny X, Y je vždy $\rho_{XY} = 0$

(populační) korelační koeficient

- ▶ Pearsonův korelační koeficient

$$r_{x,y} = \frac{s_{xy}}{s_x s_y}$$

- ▶ výběrová kovariance dána vztahem (str. 59)

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- ▶ **populační protějšek**

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- ▶ ρ_{XY} má stejné vlastnosti jako r_{xy} , zejména platí $|\rho_{XY}| \leq 1$
- ▶ pro **nezávislé** náhodné veličiny X, Y je vždy $\rho_{XY} = 0$

(populační) korelační koeficient

- ▶ Pearsonův korelační koeficient

$$r_{x,y} = \frac{s_{xy}}{s_x s_y}$$

- ▶ výběrová kovariance dána vztahem (str. 59)

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- ▶ **populační protějšek**

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- ▶ ρ_{XY} má stejné vlastnosti jako r_{xy} , zejména platí $|\rho_{XY}| \leq 1$
- ▶ pro **nezávislé** náhodné veličiny X, Y je vždy $\rho_{XY} = 0$

idealizovaný příklad: známky u zkoušky

sdružené a **marginální** pravděpodobnosti

X	Y				P(X = k)
	1	2	3	4	
1	0,15	0,10	0,05	0,00	0,3
2	0,10	0,15	0,10	0,05	0,4
3	0,05	0,05	0,05	0,05	0,2
4	0,00	0,00	0,00	0,10	0,1
	0,3	0,3	0,2	0,2	1,0

$$\sigma_{X,Y} = (1 - 2,1)(1 - 2,3) \cdot 0,15 + (1 - 2,1)(2 - 2,3) \cdot 0,10 + \dots \\ + (4 - 2,1)(3 - 2,3) \cdot 0,00 + (4 - 2,1)(4 - 2,3) \cdot 0,10 = 0,57$$

$$\rho_{X,Y} = \frac{0,57}{0,943 \cdot 1,1} = 0,55 \quad \Rightarrow \quad X \text{ a } Y \text{ jsou závislé}$$

idealizovaný příklad: známky u zkoušky

sdružené a **marginální** pravděpodobnosti

X	Y				P(X = k)
	1	2	3	4	
1	0,15	0,10	0,05	0,00	0,3
2	0,10	0,15	0,10	0,05	0,4
3	0,05	0,05	0,05	0,05	0,2
4	0,00	0,00	0,00	0,10	0,1
	0,3	0,3	0,2	0,2	1,0

$$\sigma_{X,Y} = (1 - 2,1)(1 - 2,3) \cdot 0,15 + (1 - 2,1)(2 - 2,3) \cdot 0,10 + \dots \\ + (4 - 2,1)(3 - 2,3) \cdot 0,00 + (4 - 2,1)(4 - 2,3) \cdot 0,10 = 0,57$$

$$\rho_{X,Y} = \frac{0,57}{0,943 \cdot 1,1} = 0,55 \quad \Rightarrow \quad X \text{ a } Y \text{ jsou závislé}$$

idealizovaný příklad: známky u zkoušky

sdružené a **marginální** pravděpodobnosti

X	Y				P(X = k)
	1	2	3	4	
1	0,15	0,10	0,05	0,00	0,3
2	0,10	0,15	0,10	0,05	0,4
3	0,05	0,05	0,05	0,05	0,2
4	0,00	0,00	0,00	0,10	0,1
	0,3	0,3	0,2	0,2	1,0

$$\sigma_{X,Y} = (1 - 2,1)(1 - 2,3) \cdot 0,15 + (1 - 2,1)(2 - 2,3) \cdot 0,10 + \dots \\ + (4 - 2,1)(3 - 2,3) \cdot 0,00 + (4 - 2,1)(4 - 2,3) \cdot 0,10 = 0,57$$

$$\rho_{X,Y} = \frac{0,57}{0,943 \cdot 1,1} = 0,55 \quad \Rightarrow \quad X \text{ a } Y \text{ jsou závislé}$$

idealizovaný příklad: známky u zkoušky

sdružené a **marginální** pravděpodobnosti

X	Y				P(X = k)
	1	2	3	4	
1	0,15	0,10	0,05	0,00	0,3
2	0,10	0,15	0,10	0,05	0,4
3	0,05	0,05	0,05	0,05	0,2
4	0,00	0,00	0,00	0,10	0,1
	0,3	0,3	0,2	0,2	1,0

$$\sigma_{X,Y} = (1 - 2,1)(1 - 2,3) \cdot 0,15 + (1 - 2,1)(2 - 2,3) \cdot 0,10 + \dots \\ + (4 - 2,1)(3 - 2,3) \cdot 0,00 + (4 - 2,1)(4 - 2,3) \cdot 0,10 = 0,57$$

$$\rho_{X,Y} = \frac{0,57}{0,943 \cdot 1,1} = 0,55 \quad \Rightarrow \quad X \text{ a } Y \text{ jsou závislé}$$