Úvod do praktické fyziky, cvičení 5
Binomické a Poissonovo rozdělení

Jan Matoušek

27. 10. 2020

UNIVERZITA KARLOVA
Matematicko-fyzikální
fakulta

Opakování: Momenty rozdělení diskrétní náhodné proměnné

Diskrétní náhodná proměnná k může nabývat hodnot všech celých čísel od nuly do nekonečna a má rozdělení popsané posloupností pravděpodobností $P_k = \frac{1}{ek!}$. Vypočítejte střední hodnotu μ a standardní odchylku σ této náhodné proměnné. Jaká je pravděpodobnost, že $k > 4$?

- Víme, že rozdělení pravděpodobností musí být normalizované, tj.

$$\sum_{k=0}^{\infty} P_k = \sum_{k=0}^{\infty} \frac{1}{ek!} = 1.$$

  (můžeme to snadno ověřit, vzpomeneme-li si na rozvoj $e^x = \sum_0^\infty \frac{x^k}{k!}$ a dosadíme $x = 1$)

- Střední hodnota (definovali jsme $l = k - 1$): :

$$\mu = E[k] = \sum_{k=0}^{\infty} kP_k = 0 + \sum_{k=1}^{\infty} \frac{k}{ek!} = \sum_{k=1}^{\infty} \frac{1}{e(k-1)!} = \sum_{l=0}^{\infty} \frac{1}{el!} = 1.$$

- Druhý moment:

$$E[k^2] = \sum_{k=0}^{\infty} \frac{k^2}{ek!} = \sum_{k=0}^{\infty} \frac{k(k-1)+k}{ek!} = 0 + 0 + \sum_{k=2}^{\infty} \frac{1}{e(k-2)!} + \sum_{k=0}^{\infty} \frac{k}{ek!} = 1 + 1 = 2.$$

- Standardní odchylka: $\sigma = \sqrt{E[k^2] - \mu^2} = \sqrt{2-1} = 1.$
- $P(k > 4) = 1 - P(k \leq 4) = 1 - \sum_0^4 P_k = 1 - \frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24}\right) \approx 0.0037$

Opakování: Momenty rozdělení diskrétní náhodné proměnné

Diskrétní náhodná proměnná k může nabývat hodnot všech celých čísel od nuly do nekonečna a má rozdělení popsané posloupností pravděpodobností $P_k = \frac{1}{ek!}$. Vypočítejte střední hodnotu μ a standardní odchylku σ této náhodné proměnné. Jaká je pravděpodobnost, že $k > 4$?

- Víme, že rozdělení pravděpodobností musí být normalizované, tj.

$$\sum_{k=0}^{\infty} P_k = \sum_{k=0}^{\infty} \frac{1}{ek!} = 1.$$

(můžeme to snadno ověřit, vzpomeneme-li si na rozvoj $e^x = \sum_0^\infty \frac{x^k}{k!}$ a dosadíme $x = 1$)

- Střední hodnota (definovali jsme $l = k - 1$): :

$$\mu = E[k] = \sum_{k=0}^{\infty} kP_k = 0 + \sum_{k=1}^{\infty} \frac{k}{ek!} = \sum_{k=1}^{\infty} \frac{1}{e(k-1)!} = \sum_{l=0}^{\infty} \frac{1}{el!} = 1.$$

- Druhý moment:

$$E[k^2] = \sum_{k=0}^{\infty} \frac{k^2}{ek!} = \sum_{k=0}^{\infty} \frac{k(k-1)+k}{ek!} = 0 + 0 + \sum_{k=2}^{\infty} \frac{1}{e(k-2)!} + \sum_{k=0}^{\infty} \frac{k}{ek!} = 1 + 1 = 2.$$

- Standardní odchylka: $\sigma = \sqrt{E[k^2] - \mu^2} = \sqrt{2-1} = 1.$
- $P(k > 4) = 1 - P(k \le 4) = 1 - \sum_0^4 P_k = 1 - \frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24}\right) \approx 0.0037$

Diskrétní náhodná proměnná k může nabývat hodnot všech celých čísel od nuly do nekonečna a má rozdělení popsané posloupností pravděpodobností $P_k = \frac{1}{ek!}$. Vypočítejte střední hodnotu μ a standardní odchylku σ této náhodné proměnné. Jaká je pravděpodobnost, že $k > 4$?

- Víme, že rozdělení pravděpodobností musí být normalizované, tj.

$$\sum_{k=0}^{\infty} P_k = \sum_{k=0}^{\infty} \frac{1}{ek!} = 1.$$

(můžeme to snadno ověřit, vzpomeneme-li si na rozvoj $e^x = \sum_0^\infty \frac{x^k}{k!}$ a dosadíme $x = 1$)

- Střední hodnota (definovali jsme $l = k - 1$): :

$$\mu = E[k] = \sum_{k=0}^{\infty} kP_k = 0 + \sum_{k=1}^{\infty} \frac{k}{ek!} = \sum_{k=1}^{\infty} \frac{1}{e(k-1)!} = \sum_{l=0}^{\infty} \frac{1}{el!} = 1.$$

- Druhý moment:

$$E[k^2] = \sum_{k=0}^{\infty} \frac{k^2}{ek!} = \sum_{k=0}^{\infty} \frac{k(k-1)+k}{ek!} = 0 + 0 + \sum_{k=2}^{\infty} \frac{1}{e(k-2)!} + \sum_{k=0}^{\infty} \frac{k}{ek!} = 1 + 1 = 2.$$

- Standardní odchylka: $\sigma = \sqrt{E[k^2] - \mu^2} = \sqrt{2-1} = 1.$
- $P(k > 4) = 1 - P(k \leq 4) = 1 - \sum_0^4 P_k = 1 - \frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24}\right) \approx 0.0037$

Opakování: Momenty rozdělení diskrétní náhodné proměnné

Diskrétní náhodná proměnná k může nabývat hodnot všech celých čísel od nuly do nekonečna a má rozdělení popsané posloupností pravděpodobností $P_k = \frac{1}{ek!}$. Vypočítejte střední hodnotu μ a standardní odchylku σ této náhodné proměnné. Jaká je pravděpodobnost, že $k > 4$?

- Víme, že rozdělení pravděpodobností musí být normalizované, tj.

$$\sum_{k=0}^{\infty} P_k = \sum_{k=0}^{\infty} \frac{1}{ek!} = 1.$$

(můžeme to snadno ověřit, vzpomeneme-li si na rozvoj $e^x = \sum_0^\infty \frac{x^k}{k!}$ a dosadíme $x = 1$)

- Střední hodnota (definovali jsme $l = k - 1$): :

$$\mu = E[k] = \sum_{k=0}^{\infty} k P_k = 0 + \sum_{k=1}^{\infty} \frac{k}{ek!} = \sum_{k=1}^{\infty} \frac{1}{e(k-1)!} = \sum_{l=0}^{\infty} \frac{1}{el!} = 1.$$

- Druhý moment:

$$E[k^2] = \sum_{k=0}^{\infty} \frac{k^2}{ek!} = \sum_{k=0}^{\infty} \frac{k(k-1)+k}{ek!} = 0 + 0 + \sum_{k=2}^{\infty} \frac{1}{e(k-2)!} + \sum_{k=0}^{\infty} \frac{k}{ek!} = 1 + 1 = 2.$$

- Standardní odchylka: $\sigma = \sqrt{E[k^2] - \mu^2} = \sqrt{2-1} = 1.$

- $P(k > 4) = 1 - P(k \leq 4) = 1 - \sum_0^4 P_k = 1 - \frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24}\right) \approx 0.0037$

Opakování: Momenty rozdělení diskrétní náhodné proměnné

Diskrétní náhodná proměnná k může nabývat hodnot všech celých čísel od nuly do nekonečna a má rozdělení popsané posloupností pravděpodobností $P_k = \frac{1}{ek!}$. Vypočítejte střední hodnotu μ a standardní odchylku σ této náhodné proměnné. Jaká je pravděpodobnost, že $k > 4$?

- Víme, že rozdělení pravděpodobností musí být normalizované, tj.

$$\sum_{k=0}^{\infty} P_k = \sum_{k=0}^{\infty} \frac{1}{ek!} = 1.$$

(můžeme to snadno ověřit, vzpomeneme-li si na rozvoj $e^x = \sum_0^\infty \frac{x^k}{k!}$ a dosadíme $x = 1$)

- Střední hodnota (definovali jsme $l = k - 1$): :

$$\mu = E[k] = \sum_{k=0}^{\infty} k P_k = 0 + \sum_{k=1}^{\infty} \frac{k}{ek!} = \sum_{k=1}^{\infty} \frac{1}{e(k-1)!} = \sum_{l=0}^{\infty} \frac{1}{el!} = 1.$$

- Druhý moment:

$$E[k^2] = \sum_{k=0}^{\infty} \frac{k^2}{ek!} = \sum_{k=0}^{\infty} \frac{k(k-1)+k}{ek!} = 0 + 0 + \sum_{k=2}^{\infty} \frac{1}{e(k-2)!} + \sum_{k=0}^{\infty} \frac{k}{ek!} = 1 + 1 = 2.$$

- Standardní odchylka: $\sigma = \sqrt{E[k^2] - \mu^2} = \sqrt{2-1} = 1.$

- $P(k > 4) = 1 - P(k \le 4) = 1 - \sum_0^4 P_k = 1 - \frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24}\right) \approx 0.0037$

Diskrétní náhodná proměnná k může nabývat hodnot všech celých čísel od nuly do nekonečna a má rozdělení popsané posloupností pravděpodobností $P_k = \frac{1}{ek!}$. Vypočítejte střední hodnotu μ a standardní odchylku σ této náhodné proměnné. Jaká je pravděpodobnost, že $k > 4$?

- Víme, že rozdělení pravděpodobností musí být normalizované, tj.

$$\sum_{k=0}^{\infty} P_k = \sum_{k=0}^{\infty} \frac{1}{ek!} = 1.$$

  (můžeme to snadno ověřit, vzpomeneme-li si na rozvoj $e^x = \sum_0^{\infty} \frac{x^k}{k!}$ a dosadíme $x = 1$)
- Střední hodnota (definovali jsme $l = k - 1$): :

$$\mu = E[k] = \sum_{k=0}^{\infty} kP_k = 0 + \sum_{k=1}^{\infty} \frac{k}{ek!} = \sum_{k=1}^{\infty} \frac{1}{e(k-1)!} = \sum_{l=0}^{\infty} \frac{1}{el!} = 1.$$

- Druhý moment:

$$E[k^2] = \sum_{k=0}^{\infty} \frac{k^2}{ek!} = \sum_{k=0}^{\infty} \frac{k(k-1)+k}{ek!} = 0 + 0 + \sum_{k=2}^{\infty} \frac{1}{e(k-2)!} + \sum_{k=0}^{\infty} \frac{k}{ek!} = 1 + 1 = 2.$$

- Standardní odchylka: $\sigma = \sqrt{E[k^2] - \mu^2} = \sqrt{2-1} = 1.$
- $P(k > 4) = 1 - P(k \leq 4) = 1 - \sum_0^4 P_k = 1 - \frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24}\right) \approx 0.0037$

Ke 26. 10. je v Praze evidováno 696 osob s prokázaným onemocněním COVID-19 na 100 000 obyvatel [MZ ČR]. Předpokládejme, že stejný počet lidí o svém onemocnění zatím neví. Kolik lidí se může sejít, aby pravděpodobnost že alespoň jeden bude nemocný byla menší než 5 %?

- Pravděpodobnost, že náhodný člověk je nakažený a neví o tom: $p = 7 \cdot 10^{-3}$.
- Pravděpodobnost, že n náhodných lidí není nakažených:

$$P(0|n, p) = (1-p)^n,$$

(*n* nezávislých jevů o pravděpodobnosti $1 - p$, platí pokud nákaza jednotlivých osob je nezávislá, např. nepotkali se nedávno).

- Pravděpodobnost, že alespoň jeden je nakažený:

$$P(k \geq 1|n, p) = 1 - P(0|n, p) = 1 - (1-p)^n.$$

$$0.05 = 1 - (1-p)^n$$

$$(1-p)^n = 0.95$$

$$n \ln(1-p) = \ln 0.95$$

$$n = \frac{\ln 0.95}{\ln(1-p)} = \frac{-0.05}{-0.007} \approx 7.$$

Binomické rozdělení

Ke 26. 10. je v Praze evidováno 696 osob s prokázaným onemocněním COVID-19 na 100 000 obyvatel [MZ ČR]. Předpokládejme, že stejný počet lidí o svém onemocnění zatím neví. Kolik lidí se může sejít, aby pravděpodobnost že alespoň jeden bude nemocný byla menší než 5 %?

- Pravděpodobnost, že náhodný člověk je nakažený a neví o tom: $p = 7 \cdot 10^{-3}$.
- Pravděpodobnost, že n náhodných lidí není nakažených:

$$P(0|n,p) = (1-p)^n,$$

(n nezávislých jevů o pravděpodobnosti $1-p$, platí pokud nákaza jednotlivých osob je nezávislá, např. nepotkali se nedávno).

- Pravděpodobnost, že alespoň jeden je nakažený:

$$P(k \geq 1|n,p) = 1 - P(0|n,p) = 1 - (1-p)^n.$$

$$\begin{aligned} 0.05 &= 1 - (1-p)^n \\ (1-p)^n &= 0.95 \\ n \ln(1-p) &= \ln 0.95 \\ n &= \frac{\ln 0.95}{\ln(1-p)} = \frac{-0.05}{-0.007} \approx 7. \end{aligned}$$

Binomické rozdělení

Podle otevřených dat MZ ČR [osoby.csv] pochází z Prahy 39 510 osob pozitivně testovaných na COVID-19 z celkového počtu 246 951 osob. Jaká je pravděpodobnost, že z 10 pozitivně testovaných budou z Prahy nula, jeden, dva, všichni a obecně k z N?

- Pravděpodobnost, že náhodný pozitivně testovaný je z Prahy: $p = \frac{39\,510}{246\,951} \approx 0.16$.
- Pravděpodobnost, že ani jeden z 10 vybraných není z Prahy: $P(0|10, 0.16) = (1 - 0.16)^{10} \approx 0.17$ (souběh 10 nezávislých jevů s prav. $(1 - p)$).
- Pravděpodobnost, že všichni z 10 vybraných jsou z Prahy: $P(10|10, 0.16) = 0.16^{10} \approx 10^{-8}$ (souběh 10 jevů 'ano').
- Pravděpodobnost, že jeden z 10 vybraných je z Prahy:
  - 10 nezávislých jevů (1 ano a 9 ne) $\rightarrow 0.16 \cdot (1 - 0.16)^9$.
  - navíc máme 10 možností jak toho jednoho vybrat:

ano	ne	ne	ne	ne	ne	ne	ne	ne	ne
ne	ano	ne	ne	ne	ne	ne	ne	ne	ne
.									
.									
.									
ne	ne	ne	ne	ne	ne	ne	ne	ne	ano

$\rightarrow P(1|10, 0.16) = 10 \cdot 0.16 \cdot (1 - 0.16)^9 \approx 0.33$.

Podle otevřených dat MZ ČR [osoby.csv] pochází z Prahy 39 510 osob pozitivně testovaných na COVID-19 z celkového počtu 246 951 osob. Jaká je pravděpodobnost, že z 10 pozitivně testovaných budou z Prahy nula, jeden, dva, všichni a obecně k z N?

- Pravděpodobnost, že náhodný pozitivně testovaný je z Prahy: $p = \frac{39\,510}{246\,951} \approx 0.16$.
- Pravděpodobnost, že ani jeden z 10 vybraných není z Prahy: $P(0|10, 0.16) = (1 - 0.16)^{10} \approx 0.17$ (souběh 10 nezávislých jevů s prav. $(1 - p)$).
- Pravděpodobnost, že všichni z 10 vybraných jsou z Prahy: $P(10|10, 0.16) = 0.16^{10} \approx 10^{-8}$ (souběh 10 jevů 'ano').
- Pravděpodobnost, že jeden z 10 vybraných je z Prahy:
  - 10 nezávislých jevů (1 ano a 9 ne) $\rightarrow 0.16 \cdot (1 - 0.16)^9$.
  - navíc máme 10 možností jak toho jednoho vybrat:

ano	ne	ne	ne	ne	ne	ne	ne	ne	ne
ne	ano	ne	ne	ne	ne	ne	ne	ne	ne
.									
.									
.									
ne	ne	ne	ne	ne	ne	ne	ne	ne	ano

$\rightarrow P(1|10, 0.16) = 10 \cdot 0.16 \cdot (1 - 0.16)^9 \approx 0.33$.

Binomické rozdělení

- Pravděpodobnost, že dva z 10 vybraných jsou z Prahy:
  - 10 nezávislých jevů (2 ano, 8 ne) $\rightarrow 0.16^2(1-0.16)^8$.
  - Máme 10 možností jak vybrat prvního.
  - Máme 9 možností, jak vybrat druhého.
  - Musíme ještě dělit 2 = počtem permutací mezi jevy 'ano'!

1. ano	2. ano	ne	ne	ne	ne	ne	ne	ne	ne
2. ano	1. ano	ne	ne	ne	ne	ne	ne	ne	ne

$$\rightarrow P(2|10, 0.16) = 10 \cdot 9 \cdot \frac{1}{2} \cdot 0.16^2(1-0.16)^8 \approx 0.29.$$

Obecně k "úspěchů" z N pokusů: Binomické rozdělení

$$P(k|N,p) = \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k}.$$

$$E[k] = \mu = Np$$

$$V[k] = \sigma^2 = E[k^2] - \mu^2 = Np(1-p).$$

- Jeden experiment: „úspěch" s prav. p, „neúspěch" s prav. $1-p$.
- $p^k(1-p)^{N-k}$ je prav. souběhu k úspěchů a $N-k$ neúspěchů.
- $\frac{N!}{k!(N-k)!}$ počet způsobů výběru k úspěchů z N pokusů bez ohledu na pořadí.

(Důkaz střední hodnoty a rozptylu: stejný princip jako u prního dnešního příkladu, viz přednáška prof. Čížka.

Binomické rozdělení

- Pravděpodobnost, že dva z 10 vybraných jsou z Prahy:
  - 10 nezávislých jevů (2 ano, 8 ne) $\rightarrow 0.16^2(1-0.16)^8$.
  - Máme 10 možností jak vybrat prvního.
  - Máme 9 možností, jak vybrat druhého.
  - Musíme ještě dělit 2 = počtem permutací mezi jevy 'ano'!

1. ano	2. ano	ne	ne	ne	ne	ne	ne	ne	ne
2. ano	1. ano	ne	ne	ne	ne	ne	ne	ne	ne

$\rightarrow P(2|10, 0.16) = 10 \cdot 9 \cdot \frac{1}{2} \cdot 0.16^2(1-0.16)^8 \approx 0.29$.

Obecně k "úspěchů" z N pokusů: Binomické rozdělení

$$P(k|N,p) = \frac{N!}{k!(N-k)!}p^k(1-p)^{N-k}.$$

$$E[k] = \mu = Np$$

$$V[k] = \sigma^2 = E[k^2] - \mu^2 = Np(1-p).$$

- Jeden experiment: „úspěch“ s prav. p, „neúspěch“ s prav. $1-p$.
- $p^k(1-p)^{N-k}$ je prav. souběhu k úspěchů a $N-k$ neúspěchů.
- $\frac{N!}{k!(N-k)!}$ počet způsobů výběru k úspěchů z N pokusů bez ohledu na pořadí.

(Důkaz střední hodnoty a rozptylu: stejný princip jako u prního dnešního příkladu, viz přednáška prof. Čížka.

Binomické rozdělení

- Pravděpodobnost, že dva z 10 vybraných jsou z Prahy:
  - 10 nezávislých jevů (2 ano, 8 ne) $\rightarrow 0.16^2(1-0.16)^8$.
  - Máme 10 možností jak vybrat prvního.
  - Máme 9 možností, jak vybrat druhého.
  - Musíme ještě dělit 2 = počtem permutací mezi jevy 'ano'!

1. ano	2. ano	ne	ne	ne	ne	ne	ne	ne	ne
2. ano	1. ano	ne	ne	ne	ne	ne	ne	ne	ne

$\rightarrow P(2|10, 0.16) = 10 \cdot 9 \cdot \frac{1}{2} \cdot 0.16^2(1-0.16)^8 \approx 0.29$.

Obecně k "úspěchů" z N pokusů: Binomické rozdělení

$$P(k|N,p) = \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k}.$$

$$E[k] = \mu = Np$$

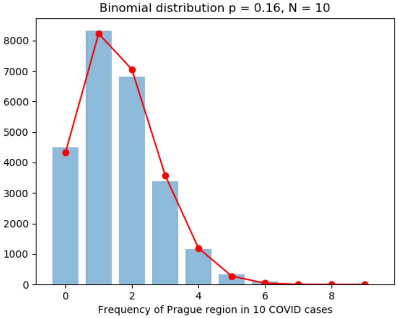
$$V[k] = \sigma^2 = E[k^2] - \mu^2 = Np(1-p).$$

- Jeden experiment: „úspěch“ s prav. p, „neúspěch“ s prav. $1-p$.
- $p^k(1-p)^{N-k}$ je prav. souběhu k úspěchů a $N-k$ neúspěchů.
- $\frac{N!}{k!(N-k)!}$ počet způsobů výběru k úspěchů z N pokusů bez ohledu na pořadí.

(Důkaz střední hodnoty a rozptylu: stejný princip jako u prního dnešního příkladu, viz přednáška prof. Čížka.

Binomické rozdělení

- Data MZ ČR, uložená do textového souboru `upf_cv5.dat`.
- Program v Pythonu `upf_cv5_praha.py`.
- Vybere 10 po sobě jdoucích osob a spočítá k, kolik z nich je z Prahy (kód kraje CZ010).
- Přičte jedničku v k-tém binu histogramu.
- Podívá se na dalších 10 osob a tak dál.
- Vykreslí histogram.
- Očekávaný počet případů $n_k = n\,P(k|10, 0.16)$, kde $n = \sum_k n_k$ (červené body).



Poissonovo rozdělení

Pravděpodobnost pozorování k velmi vzácných jevů v mnoha pokusech

Limita binomického rozdělení pro $p \to 0$ a $N \to \infty$,$\nu = pN$

$$P(k|\nu) = \frac{\nu^k}{k!} e^{-\nu}.$$
$$E[k] = \mu = \nu$$
$$V[k] = \sigma^2 = E[k^2] - \mu^2 = \nu.$$

- Data MZ ČR, uložená do textového souboru `upf_cv5.dat`.
- Počet nákaz přivezených z Německa: 304 z 246 951.
- Odhad střední hodnoty: $\nu = N \frac{304}{246\,951} = 0.00123N$.
- Očekávaný počet takových případů z 1000 je $n_k = P(k|1.23)$