Statistika

4. predavanje

Barbara Boldin

Fakulteta za matematiko, naravoslovje in informacijske tehnologije
Univerza na Primorskem

Diskretne slučajne spremenljivke

Binomska porazdelitev $B(n, p)$ (nadaljevanje)

Primer. Naj bo $0 < p < 1$ verjetnost, da posameznik po prejetju zdravila razvije stranske učinke. Če n posameznikov prejme zdravilo, potem je

$$X = \text{število posameznikov, ki ima po prejetju zdravila stranske učinke} \sim B(n, p).$$

Npr.: če je $n = 10, p = 0.1$, je $X \sim B(10, 0.1)$.

- verjetnost, da imata izmed 10 ljudi 2 stranske učinke je
- verjetnost, da imata izmed 10 ljudi največ dva stranske učinke je

Diskretne slučajne spremenljivke

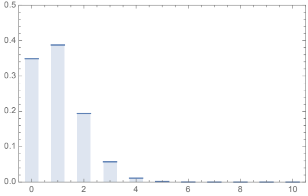
Binomska porazdelitev $B(n, p)$ (nadaljevanje)

Primer. Naj bo $0 < p < 1$ verjetnost, da posameznik po prejetju zdravila razvije stranske učinke. Če n posameznikov prejme zdravilo, potem je

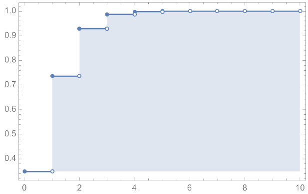
$$X = \text{število posameznikov, ki ima po prejetju zdravila stranske učinke} \sim B(n, p).$$

Npr.: če je $n = 10, p = 0.1$, je $X \sim B(10, 0.1)$.

- verjetnost, da imata izmed 10 ljudi 2 stranske učinke je $P(X = 2) = \binom{10}{2} 0.1^2 0.9^8 = 0.194$
- verjetnost, da imata izmed 10 ljudi največ dva stranske učinke je $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \binom{10}{0} 0.9^{10} + \binom{10}{1} 0.1^1 0.9^9 + \binom{10}{2} 0.1^2 0.9^8 = 0.93$

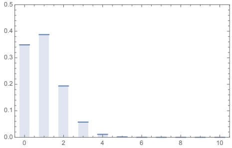


Funkcija verjetnosti za $X \sim B(10, 0.1)$

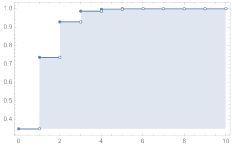


Porazdelitvena funkcija za $X \sim B(10, 0.1)$

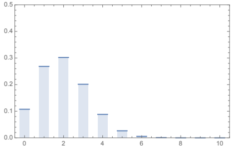
Binomske porazdelitve $B(n, p)$ za različne vrednosti p



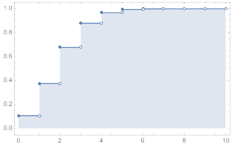
Funkcija verjetnosti za $B(10, 0.1)$



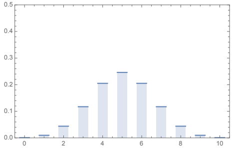
Porazdelitvena funkcija za $B(10, 0.1)$



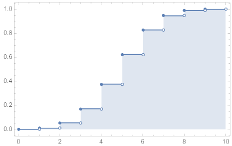
Funkcija verjetnosti za $B(10, 0.2)$



Porazdelitvena funkcija za $B(10, 0.2)$



Funkcija verjetnosti za $B(10, 0.5)$



Porazdelitvena funkcija za $B(10, 0.5)$

Pričakovana vrednost diskretne slučajne spremenljivke X, ki zavzame vrednosti x_i $(i = 1, 2, \ldots)$ in ima funkcijo verjetnosti $p(x)$ je

$$\boxed{E(X) = \sum_{i=1}^{\infty} p(x_i) x_i}$$

Primer. Izračunajmo pričakovano vrednost za slučajno spremenljivko X, za katero je:

k	p(X = k)
-2	$\frac{1}{8}$
-1	$\frac{2}{8}$
1	$\frac{3}{8}$
2	$\frac{2}{8}$

$$E(X) = (-2) \cdot \frac{1}{8} + (-1) \cdot \frac{2}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{2}{8} = \frac{3}{8}$$

Pričakovana vrednost diskretne slučajne spremenljivke X, ki zavzame vrednosti x_i $(i = 1, 2, \ldots)$ in ima funkcijo verjetnosti $p(x)$ je

$$\boxed{E(X) = \sum_{i=1}^{\infty} p(x_i) x_i}$$

Primer. Izračunajmo pričakovano vrednost za slučajno spremenljivko X, za katero je:

k	p(X = k)
-2	$\frac{1}{8}$
-1	$\frac{2}{8}$
1	$\frac{3}{8}$
2	$\frac{2}{8}$

$$E(X) = (-2) \cdot \frac{1}{8} + (-1) \cdot \frac{2}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{2}{8} = \frac{3}{8}$$

Pričakovana vrednost diskretne slučajne spremenljivke X, ki zavzame vrednosti x_i ($i = 1, 2, \ldots$) in ima funkcijo verjetnosti $p(x)$ je

$$\boxed{E(X) = \sum_{i=1}^{\infty} p(x_i) x_i}$$

Primer. Izračunajmo pričakovano vrednost za slučajno spremenljivko X, za katero je:

k	p(X = k)
-2	$\frac{1}{8}$
-1	$\frac{2}{8}$
1	$\frac{3}{8}$
2	$\frac{2}{8}$

$$E(X) = (-2) \cdot \frac{1}{8} + (-1) \cdot \frac{2}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{2}{8} = \frac{3}{8}$$

Za binomsko slučajno spremenljivko $X \sim B(n, p)$, je

$$\boxed{E(X) = np}$$

Npr.:

- pričakovana vrednost števila prejemnikov zdravila s stranskimi učinki $X \sim B(10, 0.1)$ je $E(X) = 1$
- pričakovana vrednost števila grbov pri 50 metih kovanca $Y \sim B(50, 0.5)$ je $E(Y) = 25$

Za binomsko slučajno spremenljivko $X \sim B(n, p)$, je

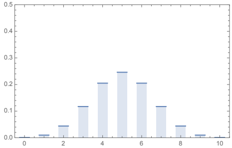
$$\boxed{E(X) = np}$$

Npr.:

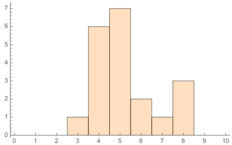
- pričakovana vrednost števila prejemnikov zdravila s stranskimi učinki $X \sim B(10, 0.1)$ je $E(X) = 1$
- pričakovana vrednost števila grbov pri 50 metih kovanca $Y \sim B(50, 0.5)$ je $E(Y) = 25$

Ali je pričakovana vrednost X enaka povprečju naključnega vzorca? Ne!

Vendar, ko velikost vzorca narašča, je povprečje vzorca vse bližje $E(X)$



Funkcija verjetnosti za $X \sim B(10, 0.5), E(X) = 5$



Histogram 20 naključno izbranih vrednosti $X \sim B(10, 0.5), \bar{x} = 5.25$

Histogram 10^6 naključno izbranih vrednosti $X \sim B(10, 0.5), \bar{x} = 5.0003$

Varianca slučajne spremenljivke X s pričakovano vrednostjo $\mu = E(X)$ je

$$\boxed{Var(X) = E\big((X-\mu)^2\big) = \sum_{i=1}^{\infty} p(x_i)(x_i-\mu)^2}$$

Drugače zapisano: $\boxed{Var(X) = E(X^2) - E^2(X)}$

Primer. Izračunajmo varianco za slučajno spremenljivko X podano v tabeli.

k	$p(X=k)$
-2	$\frac{1}{8}$
-1	$\frac{2}{8}$
1	$\frac{3}{8}$
2	$\frac{2}{8}$

Varianca slučajne spremenljivke X s pričakovano vrednostjo $\mu = E(X)$ je

$$\boxed{Var(X) = E\big((X-\mu)^2\big) = \sum_{i=1}^{\infty} p(x_i)(x_i-\mu)^2}$$

Drugače zapisano: $\boxed{Var(X) = E(X^2) - E^2(X)}$

Primer. Izračunajmo varianco za slučajno spremenljivko X podano v tabeli.

k	$p(X=k)$
-2	$\frac{1}{8}$
-1	$\frac{2}{8}$
1	$\frac{3}{8}$
2	$\frac{2}{8}$

Varianca slučajne spremenljivke X s pričakovano vrednostjo $\mu = E(X)$ je

$$\boxed{Var(X) = E\Big((X-\mu)^2\Big) = \sum_{i=1}^{\infty} p(x_i)(x_i-\mu)^2}$$

Drugače zapisano: $\boxed{Var(X) = E(X^2) - E^2(X)}$

Primer. Izračunajmo varianco za slučajno spremenljivko X podano v tabeli.

k	$p(X=k)$
-2	$\frac{1}{8}$
-1	$\frac{2}{8}$
1	$\frac{3}{8}$
2	$\frac{2}{8}$

k	$p(X^2=k)$
1	$\frac{5}{8}$
4	$\frac{3}{8}$

Vemo $E(X) = \frac{3}{8}$. Izračunamo še $E(X^2) = 1 \cdot \frac{5}{8} + 4 \cdot \frac{3}{8} = \frac{17}{8}$. Torej

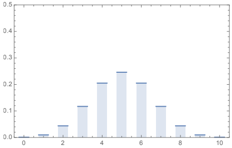
$$Var(X) = E(X^2) - E^2(X) = \frac{17}{8} - \frac{9}{64} = \frac{127}{64}.$$

Za binomsko slučajno spremenljivko $X \sim B(n, p)$ je

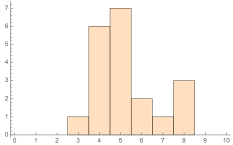
$$\boxed{Var(X) = np(1-p)}$$

Ali je varianca X enaka vzorčni varianci? Ne!

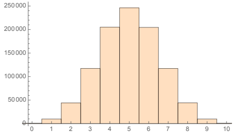
Vendar, ko velikost vzorca narašča, je vzorčna varianca vse boljši približek $Var(X)$



Funkcija verjetnosti za $X \sim B(10, 0.5)$, $Var(X) = 2.5$



Histogram 20 naključno izbranih vrednosti $X \sim B(10, 0.5)$, $s^2 = 2.2$



Histogram 10^6 naključno izbranih vrednosti $X \sim B(10, 0.5)$, $s^2 = 2.503$

Diskretne slučajne spremenljivke

Poissonova slučajna spremenljivka $Poisson(\lambda)$

Poissonova slučajna spremenljivka X: število dogodkov v fiksnem časovnem intervalu, če se ti dogodki zgodijo neodvisno drug od drugega in z dano stopnjo λ.

Poissonovo porazdelitev dobimo iz binomske če $n \to \infty, p \to 0$ in $\lambda = np = konstanta$.

Poissonova slučajna spremenljivka $X \sim Poisson(\lambda)$ ima funkcijo verjetnosti

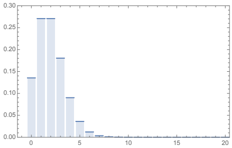
$$\boxed{P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots}$$

Za $X \sim Poisson(\lambda)$ je

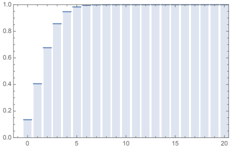
$$\boxed{E(X) = \lambda}$$

$$\boxed{Var(X) = \lambda}$$

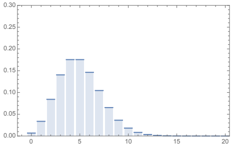
Funkcije verjetnosti in porazdelitvene funkcije *Poisson*(λ) za različne λ:



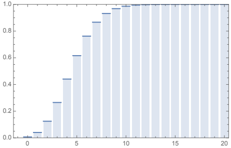
Funkcija verjetnosti za $X \sim$ *Poisson*(2)



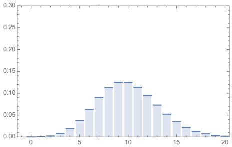
Zbirna funkcija verjetnosti za $X \sim$ *Poisson*(2)



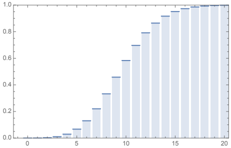
Funkcija verjetnosti za $X \sim$ *Poisson*(5)



Zbirna funkcija verjetnosti za $X \sim$ *Poisson*(5)



Funkcija verjetnosti za $X \sim$ *Poisson*(10)



Zbirna funkcija verjetnosti za $X \sim$ *Poisson*(10)

Primer. Denimo, da je število potresov v enem letu v parku Yellowstone z magnitudo več kot 3 Poissonova slučajna spremenljivka X s pričakovano vrednostjo 5.

- Kakšna je verjetnost, da bosta v parku naslednje leto dva potresa z magnitudo več kot 3?
- Kakšna je verjetnost, da bosta v parku naslednje leto največ dva potresa z magnitudo več kot 3?
- Kakšna je verjetnost, da bodo v parku naslednje leto vsaj trije potresi z magnitudo več kot 3?

Primer. Denimo, da je število potresov v enem letu v parku Yellowstone z magnitudo več kot 3 Poissonova slučajna spremenljivka X s pričakovano vrednostjo 5.

- Kakšna je verjetnost, da bosta v parku naslednje leto dva potresa z magnitudo več kot 3?

$$P(X = 2) = \frac{e^{-5}5^2}{2!} = 0.08$$

- Kakšna je verjetnost, da bosta v parku naslednje leto največ dva potresa z magnitudo več kot 3?
- Kakšna je verjetnost, da bodo v parku naslednje leto vsaj trije potresi z magnitudo več kot 3?

Primer. Denimo, da je število potresov v enem letu v parku Yellowstone z magnitudo več kot 3 Poissonova slučajna spremenljivka X s pričakovano vrednostjo 5.

- Kakšna je verjetnost, da bosta v parku naslednje leto dva potresa z magnitudo več kot 3?

$$P(X = 2) = \frac{e^{-5}5^2}{2!} = 0.08$$

- Kakšna je verjetnost, da bosta v parku naslednje leto največ dva potresa z magnitudo več kot 3?

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \frac{e^{-5}5^0}{0!} + \frac{e^{-5}5^1}{1!} + \frac{e^{-5}5^2}{2!} = 0.125 \end{aligned}$$

- Kakšna je verjetnost, da bodo v parku naslednje leto vsaj trije potresi z magnitudo več kot 3?

Primer. Denimo, da je število potresov v enem letu v parku Yellowstone z magnitudo več kot 3 Poissonova slučajna spremenljivka X s pričakovano vrednostjo 5.

- Kakšna je verjetnost, da bosta v parku naslednje leto dva potresa z magnitudo več kot 3?

$$P(X = 2) = \frac{e^{-5}5^2}{2!} = 0.08$$

- Kakšna je verjetnost, da bosta v parku naslednje leto največ dva potresa z magnitudo več kot 3?

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \frac{e^{-5}5^0}{0!} + \frac{e^{-5}5^1}{1!} + \frac{e^{-5}5^2}{2!} = 0.125 \end{aligned}$$

- Kakšna je verjetnost, da bodo v parku naslednje leto vsaj trije potresi z magnitudo več kot 3?

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.125 = 0.875$$

Diskretne slučajne spremenljivke

Geometrična slučajna spremenljivka $Geom(p)$

Z geometrično slučajno spremenljivko imamo v mislih eno od naslednjih slučajnih spremenljivk:

- število Bernoullijevih poskusov do prvega ”uspešnega” dogodka (X)

  Kaj pomeni ”uspeh”, je odvisno od konteksta, npr.: število metov kovanca do prvega grba
- število “neuspehov“ zaporednih Bernoullijevih poskusov do prvega uspešnega dogodka ($Y = X - 1$)

Omejimo se na prvo definicijo. Če je p verjetnost uspeha Bernoullijevega poskusa je funkcija verjetnosti za geometrično slučajno spremenljivko $X \sim Geom(p)$ dana z

$$\boxed{P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \ldots}$$

Velja

$$\boxed{E(X) = \frac{1}{p}} \qquad \boxed{Var(X) = \frac{1-p}{p^2}}$$

Npr.: število metov kovanca do prvega grba ima porazdelitev $Geom(0.5)$. Pričakovano število metov do prvega grba je 2.

Diskretne slučajne spremenljivke

Geometrična slučajna spremenljivka $Geom(p)$

Z geometrično slučajno spremenljivko imamo v mislih eno od naslednjih slučajnih spremenljivk:

- število Bernoullijevih poskusov do prvega ”uspešnega” dogodka (X)

  Kaj pomeni ”uspeh”, je odvisno od konteksta, npr.: število metov kovanca do prvega grba

- število “neuspehov“ zaporednih Bernoullijevih poskusov do prvega uspešnega dogodka ($Y = X - 1$)

Omejimo se na prvo definicijo. Če je p verjetnost uspeha Bernoullijevega poskusa je funkcija verjetnosti za geometrično slučajno spremenljivko $X \sim Geom(p)$ dana z

$$\boxed{P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \ldots}$$

Velja

$$\boxed{E(X) = \frac{1}{p}} \qquad \boxed{Var(X) = \frac{1-p}{p^2}}$$

Npr.: število metov kovanca do prvega grba ima porazdelitev $Geom(0.5)$. Pričakovano število metov do prvega grba je 2.

Diskretne slučajne spremenljivke

Geometrična slučajna spremenljivka $Geom(p)$

Z geometrično slučajno spremenljivko imamo v mislih eno od naslednjih slučajnih spremenljivk:

- število Bernoullijevih poskusov do prvega "uspešnega" dogodka (X)

  Kaj pomeni "uspeh", je odvisno od konteksta, npr.: število metov kovanca do prvega grba

- število "neuspehov" zaporednih Bernoullijevih poskusov do prvega uspešnega dogodka ($Y = X - 1$)

Omejimo se na prvo definicijo. Če je p verjetnost uspeha Bernoullijevega poskusa je funkcija verjetnosti za geometrično slučajno spremenljivko $X \sim Geom(p)$ dana z

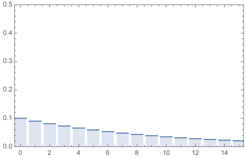
$$\boxed{P(X = k) = (1 - p)^{k-1}p, \quad k = 1, 2, \ldots}$$

Velja

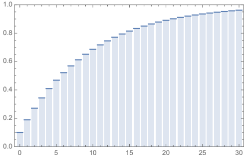
$$\boxed{E(X) = \frac{1}{p}} \qquad \boxed{Var(X) = \frac{1-p}{p^2}}$$

Npr.: število metov kovanca do prvega grba ima porazdelitev $Geom(0.5)$. Pričakovano število metov do prvega grba je 2.

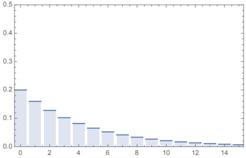
Funkcije verjetnosti in porazdelitvene funkcije $Geom(p)$ za različne p:



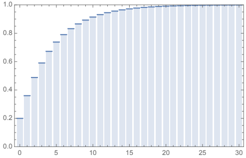
Funkcija verjetnosti za $X \sim Geom(0.1)$



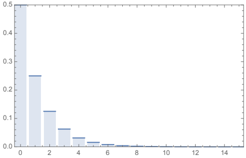
Zbirna funkcija verjetnosti za $X \sim Geom(0.1)$



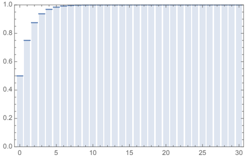
Funkcija verjetnosti za $X \sim Geom(0.2)$



Zbirna funkcija verjetnosti za $X \sim Geom(0.2)$



Funkcija verjetnosti za $X \sim Geom(0.5)$



Zbirna funkcija verjetnosti za $X \sim Geom(0.5)$

Primer. DNK je polimer, katerega osnovna enota je nukleotid. Nukleotid je sestavljen iz sladkorja (deoksiriboza), fosfatne skupine in ene od štirih dušikovih baz:

adenina (A), citozina (C), gvinina (G) ali timina (T).

$$\ldots - A - G - G - T - A - C - G - T - \ldots$$

- Na vsakem mestu nukleotidnega zaporedja je število T-jev lahko enako 1 ali 0. Število T- jev na izbranem mestu je torej Bernoullijeva slučajna spremenljivka X z

$$P(X = 1) = p = \frac{1}{4}$$
$$P(X = 0) = p = \frac{3}{4}.$$

Primer. DNK je polimer, katerega osnovna enota je nukleotid. Nukleotid je sestavljen iz sladkorja (deoksiriboza), fosfatne skupine in ene od štirih dušikovih baz:

adenina (A), citozina (C), gvinina (G) ali timina (T).

$$\ldots - A - G - G - T - A - C - G - T - \ldots$$

- Na vsakem mestu nukleotidnega zaporedja je število T-jev lahko enako 1 ali 0. Število T- jev na izbranem mestu je torej Bernoullijeva slučajna spremenljivka X z

$$P(X = 1) = p = \frac{1}{4}$$

$$P(X = 0) = p = \frac{3}{4}.$$

Primer. DNK je polimer, katerega osnovna enota je nukleotid. Nukleotid je sestavljen iz sladkorja (deoksiriboza), fosfatne skupine in ene od štirih dušikovih baz:

adenina (A), citozina (C), gvinina (G) ali timina (T).

$$\ldots - A - G - G - T - A - C - G - T - \ldots$$

- Na vsakem mestu nukleotidnega zaporedja je število T-jev lahko enako 1 ali 0. Število T- jev na izbranem mestu je torej Bernoullijeva slučajna spremenljivka X z

$$P(X = 1) = p = \frac{1}{4}$$
$$P(X = 0) = p = \frac{3}{4}.$$

- Poglejmo sedaj zaporedje petih nukleotidov.

$$\ldots - A - T - \boxed{G - G - T - A - C} - C - A - T - \ldots$$

$$\ldots - G - A - \boxed{T - C - T - A - G} - C - T - T - \ldots$$

Število T- jev na teh petih mestih je

◇ Poglejmo sedaj zaporedje petih nukleotidov.

$$\ldots - A - T - \boxed{G - G - T - A - C} - C - A - T - \ldots$$

$$\ldots - G - A - \boxed{T - C - T - A - G} - C - T - T - \ldots$$

Število T- jev na teh petih mestih je binomska slučajna spremenljivka $Y \sim B(5, 0.25)$. Funkcija verjetnosti za Y je

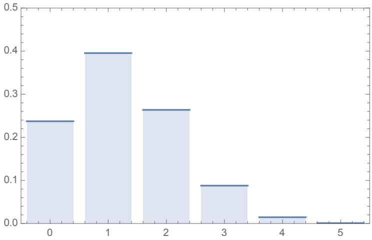
- Poglejmo sedaj zaporedje petih nukleotidov.

$$\ldots - A - T - \boxed{G - G - T - A - C} - C - A - T - \ldots$$

$$\ldots - G - A - \boxed{T - C - T - A - G} - C - T - T - \ldots$$

Število T- jev na teh petih mestih je binomska slučajna spremenljivka $Y \sim B(5, 0.25)$. Funkcija verjetnosti za Y je

k	P(Y = k)
0	0.237
1	0.395
2	0.264
3	0.088
4	0.015
5	0.001



◇

$$A - T - G - G - T - A - C - C - A - T - \ldots$$

$$G - A - T - C - T - A - G - C - T - T - \ldots$$

Število nukleotidov v zaporedju DNK do prvega T je

◇

$$A - T - G - G - T - A - C - C - A - T - \dots$$

$$G - A - T - C - T - A - G - C - T - T - \dots$$

Število nukleotidov v zaporedju DNK do prvega T je geometrična slučajna spremenljivka $Z \sim Geom(0.25)$.
Funkcija verjetnosti je

◇

$$A - T - G - G - T - A - C - C - A - T - \ldots$$

$$G - A - T - C - T - A - G - C - T - T - \ldots$$

Število nukleotidov v zaporedju DNK do prvega T je geometrična slučajna spremenljivka $Z \sim Geom(0.25)$.
Funkcija verjetnosti je

k	P(Z = k)
1	0.25
2	0.19
3	0.14
4	0.11
5	0.08
≥ 6	0.24