Téma 5.: Pravděpodobnostní funkce, hustoty a distribuční funkce v systému STATISTICA, výpočet pravděpodobností pomocí distribučních funkcí

Systém STATISTICA vytváří grafy hustot a distribučních funkcí mnoha spojitých rozložení, umí stanovit hodnotu distribuční funkce či počítat 1 - hodnota distribuční funkce. Slouží k tomu Pravděpodobnostní kalkulátor v menu Statistiky. Hodnoty pravděpodobnostních funkcí, hustot a distribučních funkcí lze počítat též pomocí funkcí implementovaných v položce „Dlouhé jméno" proměnné.
Zaměříme se na binomické rozložení, Poissonovo rozložení, exponenciální rozložení.

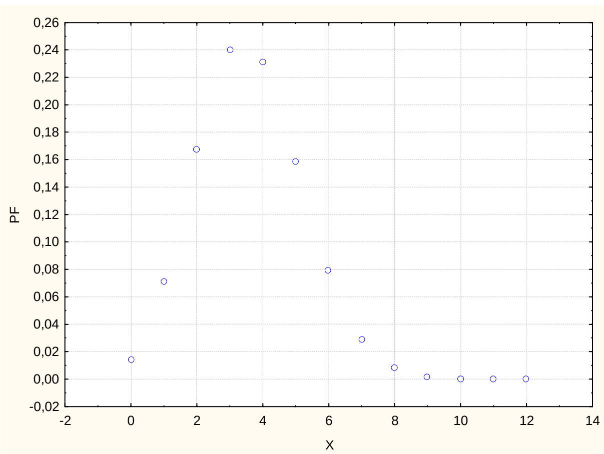
a) Binomické rozložení Bi(n, ϑ)
Náhodná veličina X udává počet úspěchů v posloupnosti n nezávislých opakovaných pokusů, přičemž pravděpodobnost úspěchu je v každém pokusu ϑ. Píšeme X ~ Bi(n, ϑ).

$$\pi(x) = \begin{cases} \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x} \text{ pro } x = 0, \dots, n \\ 0 \text{ jinak} \end{cases}, \Phi(x) = \sum_{t=0}^{x} \binom{n}{t} \vartheta^t (1-\vartheta)^{n-t}$$

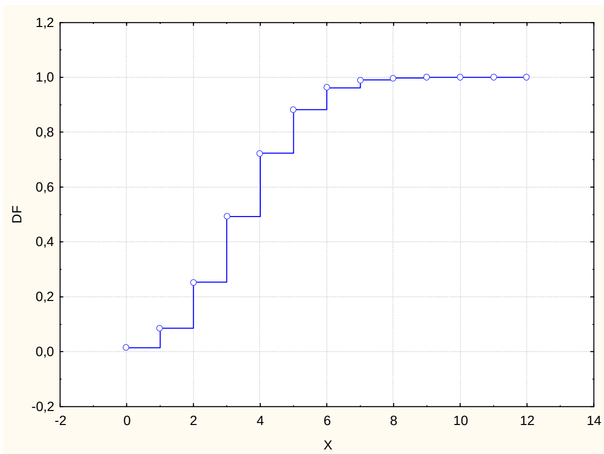
Kreslení grafů funkcí $\pi(x)$ a $\Phi(x)$ v systému STATISTICA

1. možnost: Ukážeme si, jak získat grafy pravděpodobnostní a distribuční funkce náhodné veličiny $X \sim Bi(12;0,3)$. Vytvoříme nový datový soubor o 3 proměnných a 13 případech. První proměnnou nazveme X a uložíme do ní hodnoty 0, 1, ..., 12 (do Dlouhého jména napíšeme =v0-1). Druhou proměnnou nazveme PF a uložíme do ní hodnoty pravděpodobnostní funkce (do Dlouhého jména napíšeme příkaz =Binom(x;0,3;12)). Třetí proměnnou nazveme DF a uložíme do ní hodnoty distribuční funkce (do Dlouhého jména napíšeme příkaz =IBinom(x;0,3;12)).
Graf pravděpodobnostní funkce: Grafy – Bodové grafy – Proměnné X, PF – OK – vypneme Lineární proložení – OK.
Graf distribuční funkce: Grafy – Bodové grafy – Proměnné X, DF – OK – vypneme Lineární proložení – OK – 2x klikneme na pozadí grafu – Graf:Obecné – zaškrtneme Spojnice – Typ spojnice: Schod – OK.

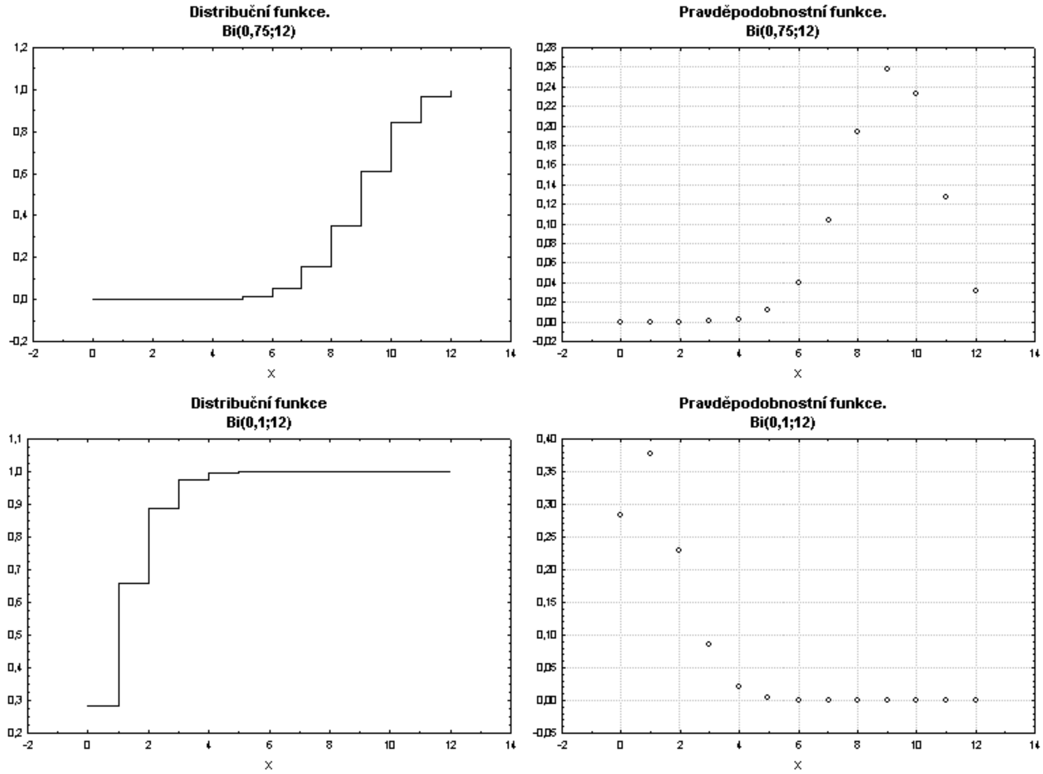
Graf funkce $p(x)$ rozložení $Bi(12;0,3)$



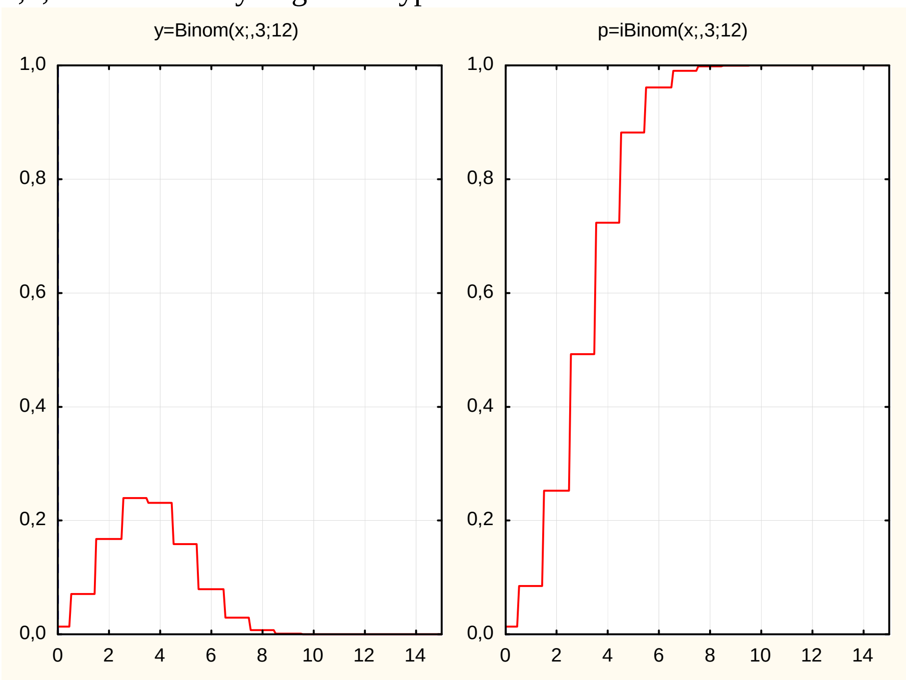
Graf funkce $F(x)$ rozložení $Bi(12;0,3)$



Analogickým způsobem můžeme získat grafy pravděpodobnostních distribučních funkcí binomického rozložení pro různá n a ϑ a sledovat vliv těchto parametrů na vzhled grafů.



2. možnost: Využijeme Pravděpodobnostní kalkulátor.
Statistiky – Pravděpodobnostní kalkulátor – Rozdělení – Binomické. Vyplníme X: 0, N: 12, p: 0,3, zaškrtneme Vytv. graf – Výpočet.



Graf pravděpodobnostní funkce není z formálního hlediska správný, protože pravděpodobnostní funkce je kladná pouze v bodech 0, 1, …, n (=12) všude jinde je nulová.

b) Poissonovo rozložení Po(λ)

Náhodná veličina X udává počet událostí, které nastanou v jednotkovém časovém intervalu (resp. v jednotkové oblasti), přičemž k událostem dochází náhodně, jednotlivě a vzájemně nezávisle. Parametr $\lambda > 0$ je střední počet těchto událostí. Píšeme $X \sim \mathrm{Po}(\lambda)$.

$$\pi(x) = \begin{cases} \dfrac{\lambda^x}{x!} e^{-\lambda} \text{ pro } x = 0, 1, \ldots \\ 0 \text{ jinak} \end{cases}, \Phi(x) = \sum_{t=0}^{x} \frac{\lambda^t}{t!} e^{-\lambda}$$

Vztah mezi pravděpodobnostní funkcí binomického a Poissonova rozložení:

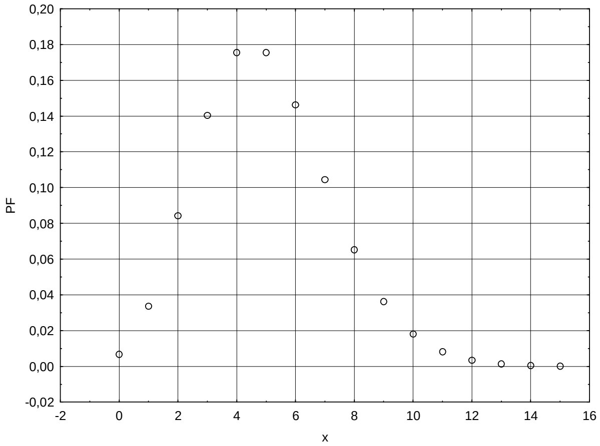
Náhodná veličina X ~ Po(λ) a náhodná veličina Y ~ Bi(n, ϑ_n). Nechť $\vartheta_n \to 0$ pro $n \to \infty$ a přitom $n\vartheta_n \to \lambda$. Pak pravděpodobnostní funkce náhodné veličiny Y konverguje k pravděpodobnostní funkci náhodné veličiny X, tj. $\lim\limits_{n\to\infty} \binom{n}{y} \vartheta_n{}^y (1-\vartheta_n)^{n-y} = \dfrac{\lambda^y}{y!} e^{-\lambda}$.

(Aproximace binomického rozložení Poissonovým je vyhovující, když n > 30 a ϑ < 0,1.)

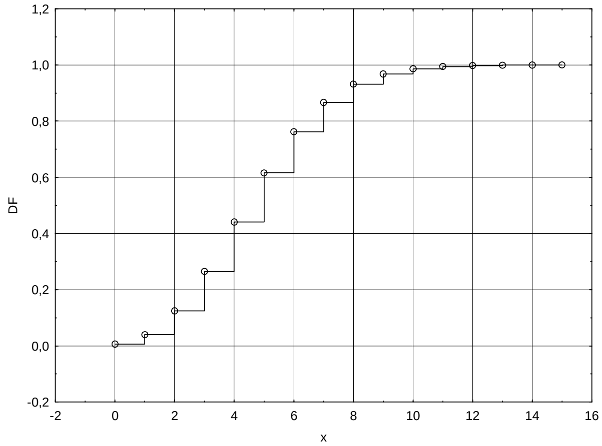
Kreslení grafů funkcí $\pi(x)$ a $\Phi(x)$ v systému STATISTICA

1. možnost: Při tvorbě grafů pravděpodobnostní a distribuční funkce náhodné veličiny s Poissonovým rozložením, např. $X \sim \mathrm{Po}(5)$, postupujeme podobně jako u binomického rozložení, ale v datovém souboru bude 16 případů a použijeme funkce Poisson(x;5) a IPoisson(x;5).

Pravděpodobnostní funkce rozložení Po(5)



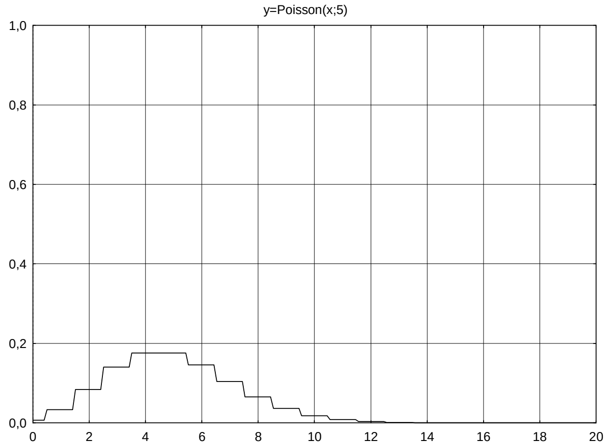
Distribuční funkce rozložení Po(5)



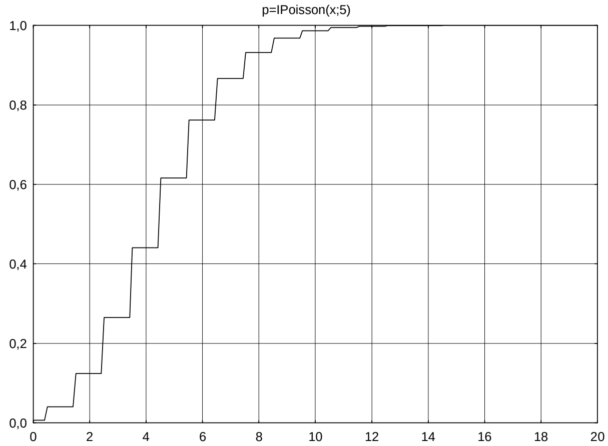
2. možnost: Využijeme Pravděpodobnostní kalkulátor.

Statistiky – Pravděpodobnostní kalkulátor – Rozdělení – Poisson. Vyplníme X: 0, Lambda 5, zaškrtneme Vytv. graf – Výpočet.

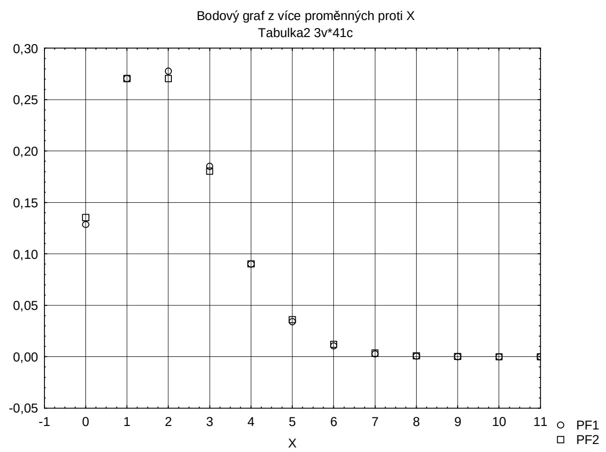
Pravděpodobnostní funkce rozložení Po(5)



Distribuční funkce rozložení Po(5)



Úkol na ilustraci vztahu mezi binomickým a Poissonovým rozložením:
Pro n = 40 a ϑ = 0,05 ilustrujte aproximaci binomického rozložení Bi(n, ϑ) Poissonovým rozložením Po(nϑ). Vypočtené hodnoty obou pravděpodobnostních funkcí v bodech x = 0, 1, …, 10 znázorněte graficky.
Ve STATISTICE: Otevřeme nový datový soubor se třemi proměnnými X, PF1, PF2 a 41 případy. Do Dlouhého jména proměnné X napíšeme =v0-1, do Dlouhého jména proměnné PF1 napíšeme =Binom(X;0,05;40) a do Dlouhého jména proměnné PF2 napíšeme =Poisson(X;2).
Grafy – Bodové grafy – vypneme Lineární proložení – zaškrtneme Typ grafu Vícenásobný – Proměnné X: X, Y: PF1, PF2 – OK. Ve vzniklém grafu změníme měřítko na vodorovné ose (od -1 do 11) a krok zmenšíme na 2.



Příklad 1.: Při provozu balicího automatu vznikají během směny náhodné poruchy, které se řídí rozložením Po(2). Jaká je pravděpodobnost, že během směny dojde k aspoň jedné poruše?
Řešení: X – počet poruch během směny, X ~ Po(2), $P(X \geq 1) = 1 – P(X < 1) = 1 – P(X = 0) =$

$$= 1 - \frac{2^0}{0!} e^{-2} = 0{,}8647.$$

Návod na výpočet pomocí systému STATISTICA: Otevřeme nový datový soubor o jedné proměnné a jednom případu. Do Dlouhého jména této proměnné napíšeme =1-IPoisson(0;2). Dostaneme výsledek 0,8647.

Příklad 2.: Telefonní ústředna zapojí během hodiny průměrně 15 hovorů. Jaká je pravděpodobnost, že během 4 minut ústředna zapojí a) právě 1 hovor, b) aspoň 2 hovory?
Řešení: X – počet zapojených hovorů během 4 minut = 1/15 hodiny, X ~ Po(tλ), kde t = 1/15 a λ = 15, tedy X ~ Po(1).

ad a) $P(X=1)=e^{-1}=0{,}36788$,

ad b) $P(X\geq 2)=1-P(X\leq 1)=1-P(X=0)-P(X=0)=1-2e^{-1}=0{,}264242$

Návod na výpočet pomocí systému STATISTICA: a) = Poisson(1;1), b) =1– IPoisson(1;1)

Příklad 3.: Semena rostlin určitého druhu jsou znečištěna malým množstvím plevele. Je známo, že na jedné jednotce plochy vyrostou po osetí v průměru 4 rostliny plevele. Lze předpokládat, že počet rostlin plevele na dané jednotce plochy se řídí Poissonovým rozložením. Vypočítejte pravděpodobnost, že na dané jednotce plochy:
a) nebude žádný plevel,
b) vyrostou nejvýše 3 rostliny plevele,
c) vyroste aspoň 5, ale nejvýše 7 rostlin plevele.
Řešení: X – počet rostlin plevele na jednotce plochy, X ~ Po(4)

Ad a) $P(X=0)=e^{-4}=0{,}0183$

Ve STATISTICE: =Poisson(0;4)

Ad b) $P(X\leq 3)=\sum_{x=0}^{3}\frac{4^x}{x!}e^{-4}=0{,}4335$

Ve STATISTICE: =IPoisson(3;4)

Ad c) $P(5\leq X\leq 7)=\sum_{x=5}^{7}\frac{4^x}{x!}e^{-4}=0{,}32$

Ve STATISTICE: =IPoisson(7;4)- IPoisson(4;4)

Příklad 4.: (Aproximace binomického rozložení Poissonovým rozložením) Dělnice v přádelně obsluhuje 800 vřeten. Pravděpodobnost toho, že se příze přetrhne během časového intervalu délky t, je pro všechna vřetena stejná a je rovna 0,005. Určete pravděpodobnost, že během intervalu délky t dojde k nejvýše 10 přetržením.
Řešení: Y – počet přetržení v časovém intervalu délky t, Y ~ Bi(800;0,005).

Přesný výpočet: $P(Y\leq 10)=\sum_{y=0}^{10}\binom{800}{y}0{,}005^y(1-0{,}005)^{800-y}=0{,}997239$

Ve STATISTICE: =IBinom(10;0,005;800).
Aproximativní výpočet: podmínky dobré aproximace jsou splněny, parametr

$\lambda = n\vartheta = 800.0{,}005 = 4$, $P(Y\leq 10)=\sum_{y=0}^{10}\frac{4^y}{y!}e^{-4}=0{,}9971602$

Ve STATISTICE: =IPoisson(10;4).

c) Exponenciální rozložení Ex(λ)

Náhodná veličina X udává dobu čekání na příchod nějaké události, která se může dostavit každým okamžikem se stejnou šancí bez ohledu na dosud pročekanou dobu. (Jde o tzv. čekání bez paměti.) Přitom $\frac{1}{\lambda}$ vyjadřuje střední dobu čekání. Náhodná veličina X ~ Ex(λ) má hustotu

$$\varphi(x) = \begin{cases} \lambda e^{-\lambda x} \text{ pro } x > 0 \\ 0 \text{ jinak} \end{cases}.$$

Použití systému STATISTICA při výpočtu hodnot distribuční funkce exponenciálního rozložení:

První možnost: Ve volbě Rozdělení vybereme Exponenciální, do okénka lambda napíšeme hodnotu parametru λ. Hodnotu distribuční funkce v bodě x zjistíme tak, že do okénka označeného X napíšeme dané x a po kliknutí na Výpočet se v okénku p objeví hodnota distribuční funkce.

Druhá možnost: Výpočet hodnoty distribuční funkce pomocí funkcí implementovaných v položce „Dlouhé jméno“: Otevřeme nový datový soubor o jedné proměnné a jednom případu. V položce „Dlouhé jméno“ této proměnné použijeme funkci IExpon(x;lambda).

Příklad 1.: Doba do ukončení opravy v opravně obuvi je náhodná veličina, která se řídí exponenciálním rozložením se střední dobou opravy 3 dny. Jaká je pravděpodobnost, že oprava bude ukončena do dvou dnů?

Řešení: $X \sim \mathrm{Ex}(1/3),\ P(X \le 2) = \int_0^2 \frac{1}{3} e^{-\frac{x}{3}} dx = \left[-e^{-\frac{x}{3}} \right]_0^2 = 1 - e^{-\frac{2}{3}} = 0{,}4866$

Návod na výpočet pomocí systému STATISTICA:

První možnost: Do okénka lambda napíšeme 0,3333, do okénka exp. napíšeme 2 a po kliknutí na Výpočet se v okénku p objeví 0,4866.

Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případu. Do dlouhého jména této proměnné napíšeme =IExpon(2;1/3). Dostaneme 0,4866.

Příklad 2.: Doba (v hodinách), která uplyne mezi dvěma naléhavými příjmy v jisté nemocnici, se řídí exponenciálním rozložením se střední dobou čekání 2 h. Jaká je pravděpodobnost, že uplyne více než 5 h bez naléhavého příjmu?

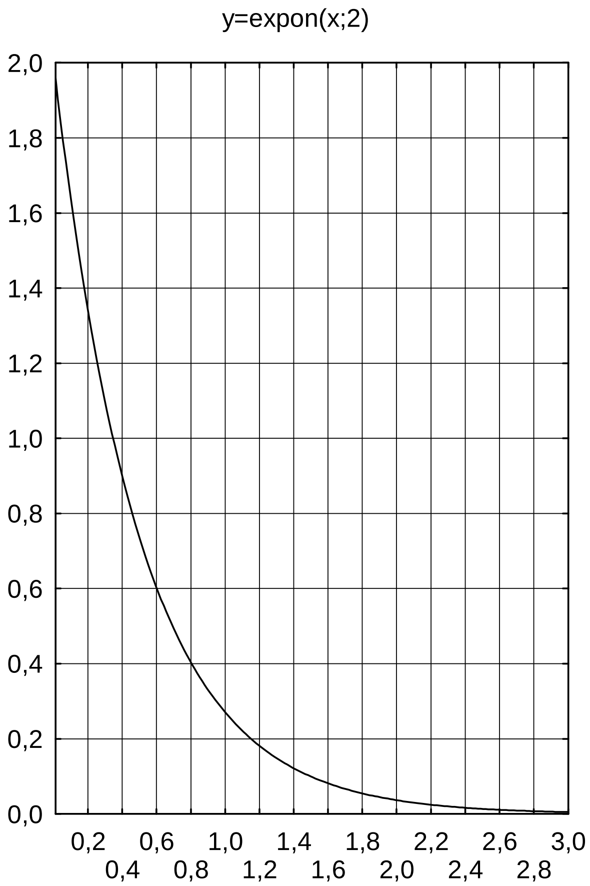
Výsledek: $X \sim \mathrm{Ex}(1/2),\ P(X > 5) = 0{,}082$

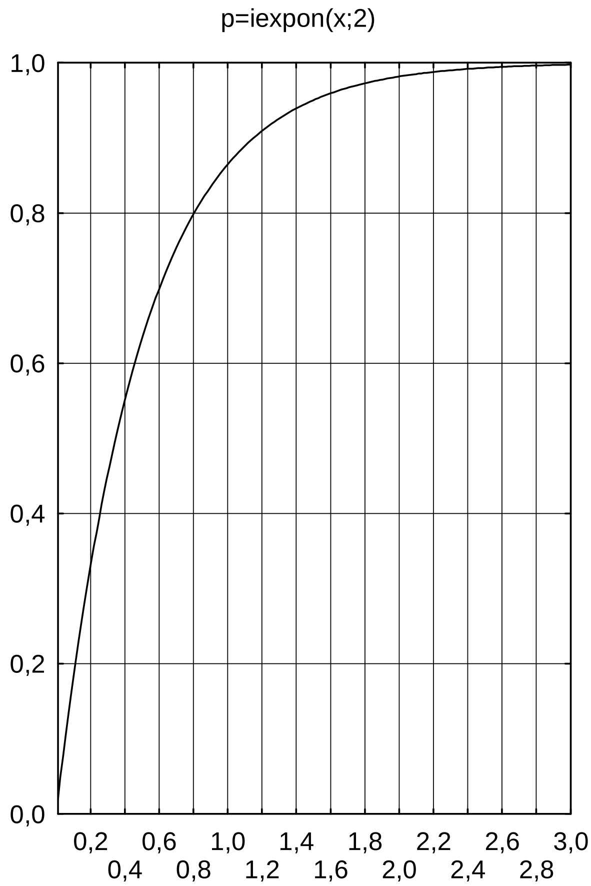
Příklad 3.: Dlouhodobým pozorováním v určité prodejně bylo zjištěno, že 40 % zákazníků je obslouženo do 3 minut. Lze předpokládat, že doba čekání se řídí exponenciálním rozložením. Jaké procento zákazníků bude na obsluhu čekat déle než 6 minut?

Výsledek: $X \sim \mathrm{Ex}(1/3),\ P(X > 6) = 0{,}36$. Znamená to, že 36 % zákazníků bude čekat na obsluhu déle než 6 minut.

Kreslení grafů funkcí $\varphi(x)$ a $\Phi(x)$ rozložení Ex(2) v systému STATISTICA pomocí Pravděpodobnostního kalkulátoru

Statistiky – Pravděpodobnostní kalkulátor – Rozdělení – Exponenciální. Vyplníme lambda: 2, zaškrtneme Vytv. graf – Výpočet.


y=expon(x;2)
2,0
1,8
1,6
1,4
1,2
1,0
0,8
0,6
0,4
0,2
0,0
0,2
0,4
0,6
0,8
1,0
1,2
1,4
1,6
1,8
2,0
2,2
2,4
2,6
2,8
3,0


p=iexpon(x;2)
1,0
0,8
0,6
0,4
0,2
0,0
0,2
0,4
0,6
0,8
1,0
1,2
1,4
1,6
1,8
2,0
2,2
2,4
2,6
2,8
3,0