Rozdělení NV

Úvod

Velké skupiny náhodných pokusů vykazují stejné pravděpodobnostní chování
Mince – panna/orel
Výška mužů/žen
NV mohou být spojeny s určitým pravděpodobnostním rozdělení
(již známe jeho hustotu pravděpodobnosti/pravděpodobnostní funkci)

Rozdělení slouží:
K přesnému popisu pravděpodobnostního chování NV
Střední hodnota, rozptyl, korelace atd.

Rozdělení diskrétního typu

Alternativní rozdělení A[p]

Možné pouze 2 výsledky
Jev A nebo jeho doplněk A´ - jev A nastane a nebo nenastane A´
p - pravděpodobnost výskytu jevu A – zavedeme X=1
(1-p) – pravděpodobnost výskytu jevu A´ - zavedeme X=0
Úspěch vs. Neúspěch

Pravděpodobnostní funkce s parametrem p pak:

$$\mathbf{p(x)=p^{x}.(1-p)^{1-x}\ \ pro\ x=0\ nebo1}$$

P(X=1)=p
P(X=0)= 1-p

E(X)=p
D(X)=p.(1-p)

Nastoupení jevu A v každém pokusu nezávisí na výsledcích předchozích pokusů
Pokusy jsou nezávislé

Binomické rozdělení Bi[n;p]

Náhodný pokus z A[p] opakujeme za stejných podmínek *n*-krát po sobě
Náhodné pokusy jsou nezávislé (s vracením)
NV X – kolikrát nastal jev A s pravděpodobností p(panna) v *n* opakování
A=(0,1,2,3…,k,…,n)
Jev A nenastane s pravděpodobností (1-p) (orel)
Jaká je pravděpodobnost, že jev A nastane k-krát při *n* opakování
X=5 – jev nastane 5-krát při 10 pokusech (panna 5x, při 10 hodech)

$$\mathbf{p^{k}.(1\text{-}p)^{n\text{-}k}}$$
$$\mathbf{p^{5}.(1\text{-}p)^{10\text{-}5}}$$

Možnost výskytu jevu A k-krát v *n* opakování? $\binom{n}{k}$ $\frac{n!}{k!.(n-k)!}$

Možnost výskytu panny 5-krát v 10 opakování? $\binom{10}{5}$

Pravděpodobnostní funkce

$$P(X=k)=\binom{n}{k}.p^{k}.(1-p)^{n-k}$$

Pravděpodobnost, že se jev A vyskytne k-krát v *n* opakování

$$P(X=k)=\binom{n}{k}.p^k.(1-p)^{n-k}$$

Bi[n;p]- dva parametry
n-počet pokusů (hodů mincí)
p- pravděpodobnost výskytu jevu A – úspěch, panna

E(X)=n.p
D(X)=n.p.(1-p)

Příklad

$$P(X = k) = \binom{n}{k} . p^k . (1-p)^{n-k}$$

Jaká je pravděpodobnost, že při 10 hodech kostkou padne

1) 2x šestka P(2)
2) Maximálně 3x šestka P(X≤3)

Pravděpodobnost, že padne šestka při 1 hodu je=1/6

V každém hodu je pravděpodobnost stejná

n=10

1) k=2

$$P(X = 2) = \binom{10}{2} . (\frac{1}{6})^2 . (1 - \frac{1}{6})^{10-2} \quad = 0{,}291$$

2) P(X≤3)=F(3)=P(0)+P(1)+P(2)+P(3)=

=0,161+0,323+0,291+0,155=0,93

Poissonovo rozdělení Po[λ]

Binomické rozdělení můžeme aproximovat Poissonovým rozdělením
Když n>30 a pravděpodobnost výskytu jevu A p je velmi malá $p \le 0{,}1$
Počet zmetků při výrobě, počet událostí v čase atd.
Parametr $\lambda = n.p$

Pravděpodobnostní funkce

Pro k=0,1,2.. A $\lambda > 0$

$$P(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

$E(X)=\lambda$
$D(X)=\lambda$

Co říká P(k)?
Provedeme *n* **nezávislých** pokusů a jev A se bude vyskytovat k-krát
Čím větší *n* a menší p, tím lepší je aproximace Bi Possonovým rozdělením

Kde nejdeme Poissonovo rozdělení ?

Rozdělením se často řídí:
Počet vadných výrobků ve výrobě (malá p vzniku zmetku)
Počet jevů(událostí) v čase (kolikrát nastane porucha za 1 den)

Poissonovské proudy
Jedná se o posloupnost NJ, které nastanou v náhodných časových okamžicích za určitý časový interval.
(kolikrát nastane porucha během směny za 1 den)

$$P_t(k) = \frac{(t.\lambda)^k}{k!}.e^{-t\lambda}$$

Parametr(λ) – počet výskytů daného jevu (porucha) za časovou jednotku (směnu)

k-kolikrát nastane daný náhodný jev (porucha)
t-počet časových okamžiků

Příklad:

Ve výrobě během směny vznikají náhodné poruchy
Proud poruch lze považovat za poissonovský
Z minulosti víme - během jedné směny dochází v průměru ke 3 poruchám
Časový interval je 24 hodin a pracuje se ve dvousměnném provozu
Jaká je pravděpodobnost, že během 24 hodin dojde právě k 1 poruše?

$\lambda=3$
$t=2$

$$P_t(k) = \frac{(t.\lambda)^k}{k!}.e^{-t\lambda}$$

$$P_t(1) = \frac{(3.2)^1}{1!}.e^{-2.3}$$

$$= 0{,}0148$$

Pozor 0!=1

Hypergeometrické rozdělení H[N;M;n]

Rozdíl oproti minulým – pokusy jsou závislé!!!
(bez vracení)
Výsledek nastoupení jevu A, závisí na předešlých pokusech

Pravděpodobnostní funkce

$$P(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

N-počet prvků v souboru (počet výrobků v krabici)
M-počet prvků s určitou vlastností (počet zmetků)
n-počet vybraných prvků ze souboru (z N)
x-konkrétní hodnota NV – P(x)-pravděpodobnost, že najdu x zmetků

$$E(X) = n.\frac{M}{N} \qquad D(X) = n.\frac{M}{N}.\left(1-\frac{M}{N}\right).\frac{N-n}{N-1}$$

Pro velké N a neměnné n, M/N, se H[N;M;n] blíží Bi[n;p]
Použití – test trvanlivosti materiálu(dochází k destrukci)

Příklad:

U každé výrobní sady (1000ks - N) se zjišťuje trvanlivost materiálu
Vždy je náhodně vybráno 10 vzorků (n)
Předpokládá se, že na 1000ks vznikne 50 zmetků (M)
Výrobky se pustí do oběhu když mezi 10 vybranými kusy nebude zmetek x=0

$$P(\mathbf{0}) = \frac{\binom{50}{0}\binom{1000-50}{10-0}}{\binom{1000}{10}} \qquad P(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$= 0{,}5973$$

Spojitá pravděpodobnostní rozdělení

Náhodná veličina má spojité rozdělení
Když jej známe, jsme schopni určit jeho E(X), D(X), pravděpodobnosti výskytu

Rovnoměrné rozdělení

$$f(x) = \frac{1}{b-a} \quad pro\ a \leq x \leq b$$

$$E(X) = \frac{a+b}{2} \qquad D(X) = \frac{1}{12}.(b-a)^2$$

Normální rozdělení N[μ;σ²]

Pro nás nejdůležitější
Velké množství NV se řídí tímto spojitým rozdělením
Výšky, váhy, objem hrudníku v cm, velikost …, **chyby v měření**
Často budeme jiná spojitá i nespojitá rozdělení aproximovat N[μ;σ]
Parametry - μ;σ
Hustota pravděpodobnosti f(x) normálně rozdělené NV:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{(x-\mu)^2}{2\sigma^2}} \; pro -\infty \leq x \leq \infty, \sigma > 0, \mu \in R$$

Po výpočtu!!!

$$E(X) = \mu \qquad D(X) = \sigma^2$$

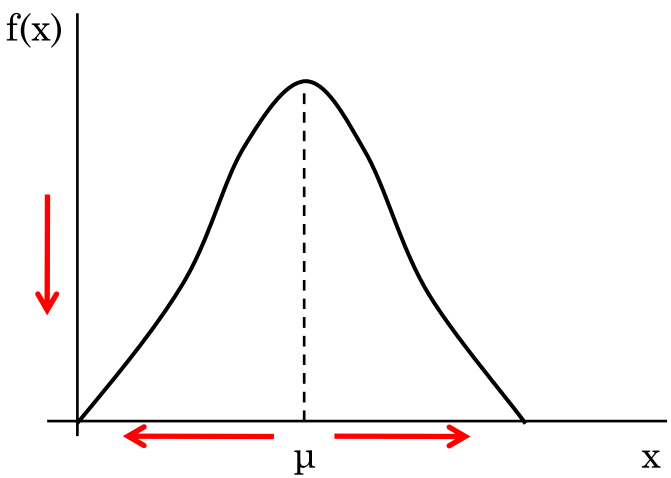
Graf f(x) normálního rozdělení je Gaussova křivka

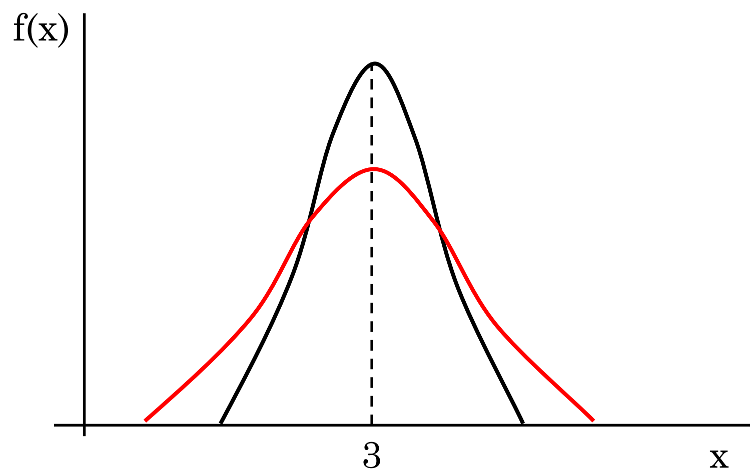
Maximum v bodě μ
Pro x blížícímu se k +/- nekonečnu – přibližování se k ose x
N[3;1] vs. N[3,2]

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \longrightarrow$$

Nepočítáme
Hledáme v tabulkách, nebo SW

Funkce je symetrická kolem x=μ





Normovaná veličina (U)

Pro zjednodušení transformujeme NV X (s N[μ;σ²])
na normovanou veličinu (U) - N[0;1] - tabelovaná

$$U = \frac{X - \mu}{\sigma}$$

$$f(u) = \frac{1}{\sqrt{2\pi}} . e^{-\frac{u^2}{2}} \qquad F(u) = \frac{1}{\sqrt{2\pi}} . \int_{-\infty}^{u} e^{-\frac{t^2}{2}} dt$$

μ=E(X)=0

σ²=D(X)=1

$$F(x) = P(X < x) = P(\frac{X - \mu}{\sigma} < \frac{x - \mu}{\sigma}) = P(U < u) = F(u)$$

Pravděpodobnost, že NV X bude menší než x
Po transformaci, NV U je menší než u
Netřeba se trápit F(u) zjistíme v tabulkách

$$u = \frac{x - \mu}{\sigma}$$

f(x)=f(-x)

Pak platí, že:

F(-x)=1-F(x)

Pozor za F se někdy dosazuje

$P(X > x) = 1 - P(X < x)$

Nezaměňovat !!

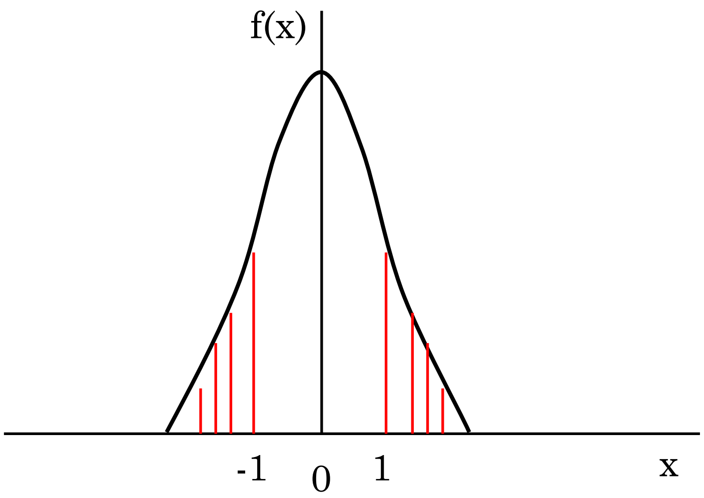
Nemusíme znát všechny hodnoty f(x) a F(X), stačí nám znát pouze pro x>0

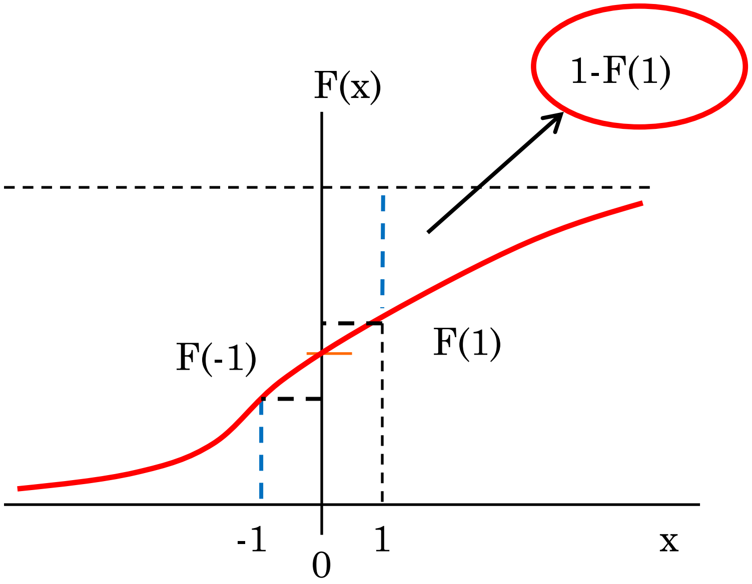
P(X<-1)=1-P(X<1)

P(X<-1)=P(X>1)

F(-1)=1-F(1)

Velikost ploch je stejná – plocha udává pravděpodobnost





Příklad:
NV X N[10;16]

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

1) P(X<3)
2) P(X>11)
3) P(9<X<11)

$$F(3) = \frac{1}{4.\sqrt{2\pi}} \cdot \int_{-\infty}^{3} e^{-\frac{(t-10)^2}{2.16}}\,dt$$

$$F(x) = P(X < x) = P\left(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right) = P(U < u) = F(u)$$

Řešení:
1) Složité – převedeme NV X na NV U, která má normované normální rozdělení

$$U = \frac{X-\mu}{\sigma}$$

$$P(X < 3) = P\left(\frac{X-10}{4} < \frac{3-10}{4}\right) = P(U < -1{,}75)$$

F(-x)=1-F(x)

$$= 1 - F(1{,}75)$$

Najdu hodnotu 1,75 ve statistických tabulkách
Pro N[0;1]
=1-0,95994=0,04006

2) P(X>11)

$$F(x) = P(X < x) = P\left(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right) = P(U < u) = F(u)$$

$$P(X > 11) = 1 - P(X < 11) = 1 - F(11)$$

Pozor normální rozdělení
ne normované normální!!!
Musíme transformovat

$$1 - P(X < 11) = 1 - P\left(\frac{X-10}{4} < \frac{11-10}{4}\right)$$

$$= 1 - P\left(U < \frac{1}{4}\right)$$

Najdu v tabulkách u=0,25
A pro ní F(0,25)

$$= 1 - 0{,}59871$$

Pravděpodobnost, že náhodná veličina X, bude mít vyšší hodnotu než 11 je 40,219%

$$P(a < X < b) = F(b) - F(a) = \int_a^b f(x)dx$$

3) P(9<X<11)

$$F\left(\frac{1}{4}\right) - F\left(-\frac{1}{4}\right)$$

$$P(9 < X < 11) = F(11) - F(9) = \int_9^{11} \frac{1}{4.\sqrt{2\pi}}.e^{-\frac{(t-10)^2}{2.16}}dt$$

Fuj…

$$P(9 < X < 11) = P\left(\frac{9-10}{4} < \frac{X-10}{4} < \frac{11-10}{4}\right) = P\left(-\frac{1}{4} < U < \frac{1}{4}\right)$$

Najdu v tabulkách u=0,25
A pro ní F(0,25); 1-F(0,25)

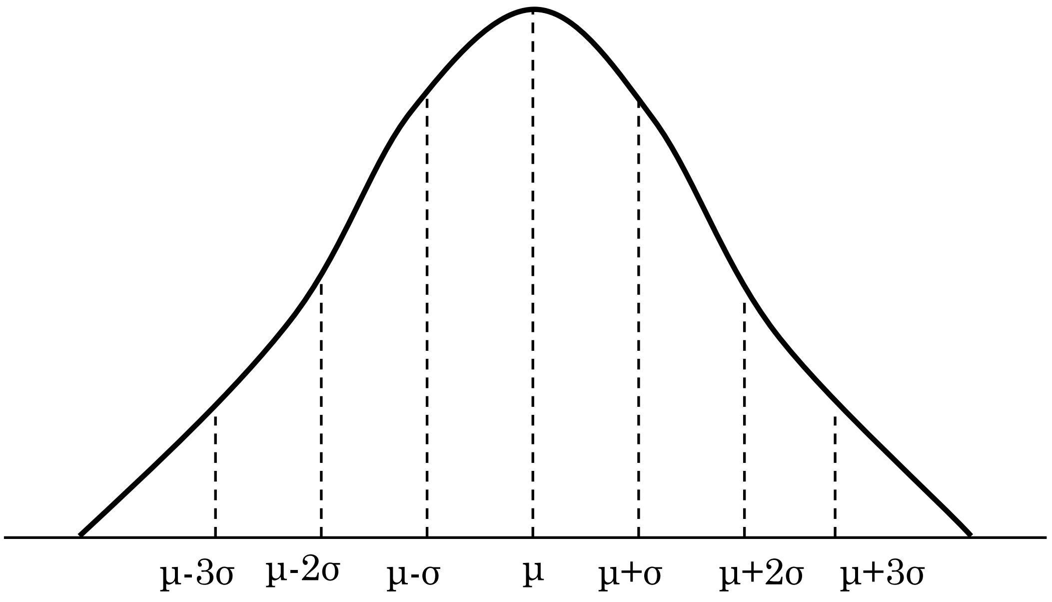
$$= 0{,}59871 - 0{,}40219 = 0{,}19652$$

Pravděpodobnost, že NV X spadne do intervalu (9;11) je 19,652%

Pravidlo šesti sigma

Pomůže při kreslení Gaussovi křivky a při výpočtech
Rozdělení křivky do 6 intervalů
Pravděpodobnost, že NV X padne do intervalu μ+/-3σ je 0,9973
Pravděpodobnost, že NV X padne do intervalu μ+/-2σ je 0,9545
Pravděpodobnost, že NV X padne do intervalu μ+/-1σ je 0,6827

Nezáleží na střední hodnotě a rozptylu – pokud má X normální rozdělení



Logaritmicko-normální rozdělení LN[μ;σ²]

Náhodná veličina Y=lnX má normální rozdělení N[μ;σ²]
Potom náhodná veličina X má logaritmicko-normální rozdělení LN[μ;σ²]
Hustota pravděpodobnosti:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} . e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \quad pro\ x > 0$$

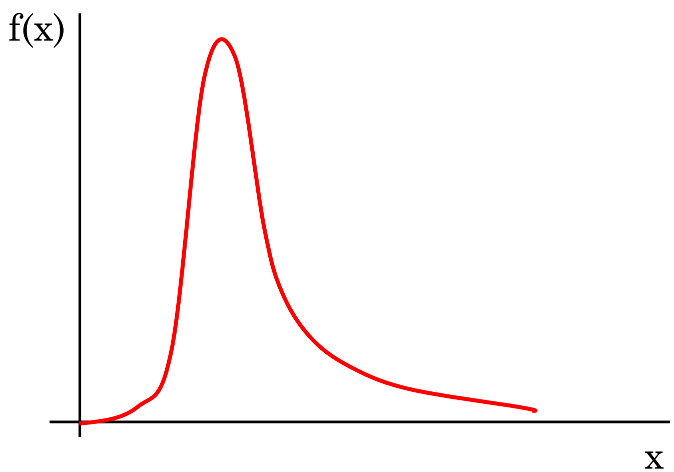
Postup stále stejný, jen jiná f(x)
NV převedeme na Y=lnX
A potom stejně jako u N[μ;σ²]

$$\mu = E(Y) = E(\ln X)$$

$$\sigma^2 = D(Y) = D(\ln X)$$

$$E(X) = e^{\mu+\frac{\sigma^2}{2}}$$

$$D(X) = e^{2\mu+\sigma^2}.(e^{\sigma^2} - 1)$$



Příklad: Jen (X)=lnX a pak jako u normovaného normálního rozdělení

LN[2;4]

1) F(6) – P(X<6)
2) Medián
3) 95% kvantil

$$P(X < 6) = P\left(\frac{lnX - 2}{2} < \frac{ln6 - 2}{2}\right) = F(-0{,}104) = 1 - F(0{,}104) = 1 - 0{,}534 = 0{,}466$$

Kvantily

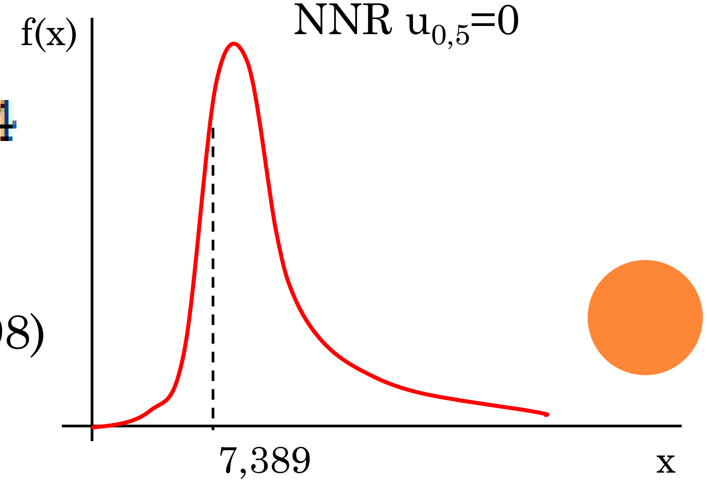
$$x_p = e^{(\mu + \sigma u_p)}$$

u_p – p% kvantil normovaného normálního rozdělení – hledáme v tabulce
Znám pravděpodobnost, ale neznám konkrétní hodnotu x

$$x_{0,5} = e^{(2+2.0)} = 7{,}389$$

50% kvantil
NNR $u_{0,5}$=0

$$x_{0,95} = e^{(2+2.1{,}645)} = 198{,}34$$



Pro dané rozdělení platí, že:
Pravděpodobnost, že X padne do intervalu (0;198)
Je 95%

Exponenciální rozdělení E[A;δ]

Životnost výrobků a zařízení
(A – minimální doba životnosti, nebo doba během níž nemůže daný jev nastat – A se často volí A=0)

Hustota pravděpodobnosti

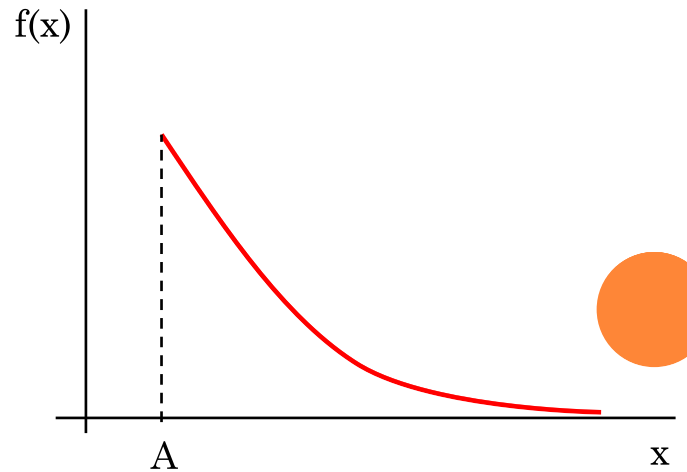
$$f(x) = \frac{1}{\delta} e^{-\frac{(x-A)}{\delta}} \; pro \; x > A \qquad E(X) = A + \delta \qquad D(X) = \delta^2$$

Distribuční funkce:

$$F(x) = 1 - e^{-\frac{(x-A)}{\delta}}$$

Kvantil

$$x_p = A - \delta \ln(1 - p)$$



Příklad:

Životnost notebooku je NV s E(X) = 10000 hodin
Jaká je pravděpodobnost, že životnost bude větší jak 12000 hodin?
Zkušenosti z minulosti – životnost notebooku se řídí exponenciálním rozdělením E[0;10000]

Hledáme $P(X_{\text{životnost}} > 12000)$
Distribuční funkce $F(x)=P(X<x)$

$$P(X > 12000) = 1 - P(X < 12000) = 1 - \int_0^{12000} \frac{1}{10000} . e^{\frac{-x}{10000}} dx$$

$$1 - F(x) = 1 - \left(1 - e^{-\frac{12000}{10000}}\right)$$

$$P(X > 12000) = 0{,}30119$$

Pravděpodobnost, že náhodně vybraný notebook bude mít životnost vyšší než 12000 hodin je 30%.

$$F(x) = 1 - e^{-\frac{(x-A)}{\delta}} \qquad E(X) = A + \delta$$

Rozdělení χ^2 (chí kvadrát) χ^2[v]

v-počet stupňů volnosti

Máme $U_1, U_2, \ldots, U_v$ nezávislých náhodných veličin a každá má N[0;1]

Počet NV udává počet stupňů volnosti

$$\chi 2 = \sum_{i=1}^{v} U_i^2$$

$$E(\chi 2) = \mathbf{v} \quad D(\chi 2) = \mathbf{2v}$$

S růstem stupňů volnosti se χ^2[v] blíží normovanému normálnímu rozdělení v>30

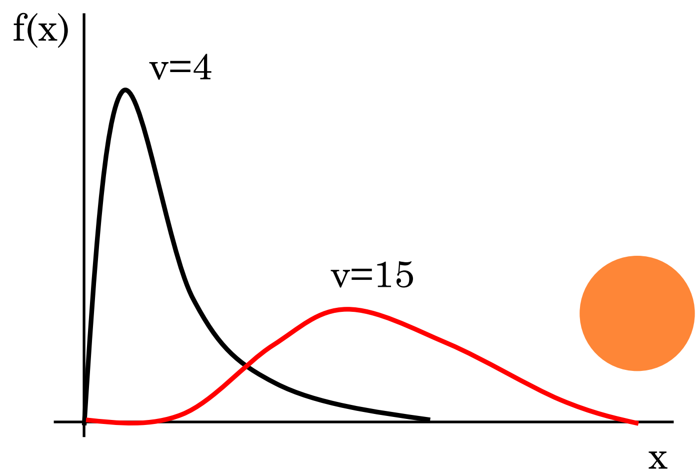
Kvantily

$$\chi 2_p[\mathrm{v}] \sim \frac{1}{2}(u_p + \sqrt{2\mathrm{v}-1})^2$$

$\chi^2_{0,90}$=? v=10 znám požadovou pravděpodobnost

Hledáme x, kdy bude platit, že P(X<x)=0,9

1) Najdu v tabulkách $\chi^2_{0,90}$**[10]= 16**



v=40

1) Najdu $u_p(u_{0,9})$-1,282

$$\chi 2_{0,9}[40] \sim \frac{1}{2}(1{,}282 + \sqrt{2.40-1})^2 = 51{,}716$$

t-rozdělení (Studentovo) t[v]

Pracujeme se 2 náhodnými veličinami
První U má normované normální rozdělení N[0;1]
Druhá NV má χ^2 rozdělení s (v) stupni volnosti
Náhodná veličina t má pak Studentovo rozdělení s (v) stupni volnosti

$$t = \frac{U}{\sqrt{\frac{\chi^2}{v}}}$$

Není hustota pravděpodobnosti, zjednodušení
Pouhý výpočet a hledání v tabulkách

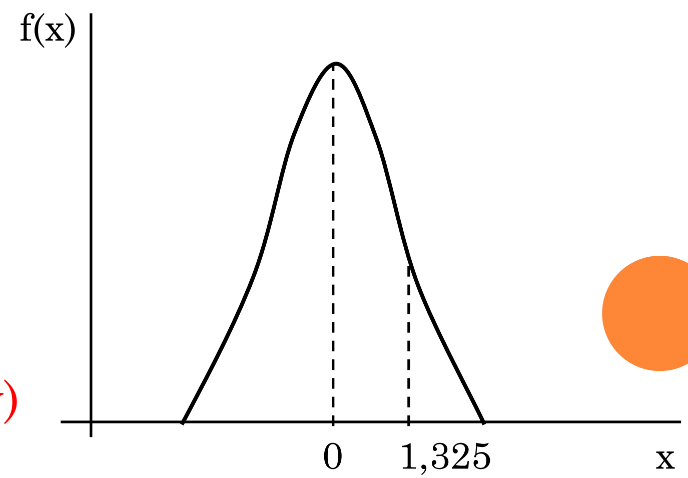
$$E(X) = 0 \quad D(X) = \frac{v}{v-2}$$

S růstem (v) se Studentovo rozdělení blíži normovanému normálnímu
v>30 – považujeme t-rozdělení za normální

Kvantily t_p

$t_p = -t_{1-p}$

v=20 p=0,9 – $t_{0,9}$ $P(t<t_{0,9})=0,9$
Kouknu do tabulky
$t_{0,9}=1,325$
Když bude v>30 – řeším už jako N[0;1] (tabulky)



Rozdělení F (Fisherovo-Snedecorovo) F[v_1;v_2] F[m;n]

Pracujeme se dvěma nezávislými náhodnými veličinami
První má $\chi^2{}_1$ rozdělení s (v_1) (m) stupni volnosti
Druhá má $\chi^2{}_2$ rozdělení s (v_2) (n) stupni volnosti

$$F = \frac{\frac{\chi^2{}_1}{v_1}}{\frac{\chi^2{}_2}{v_2}}$$

Pozor závisí na pořadí NV!!!

Kvantily

$$F_p = \frac{1}{F_{1-p}[v_2; v_1]}$$

Náhodná veličina F má rozdělení F
Známe v_1=5 a v_2=10
A potřebuju zjistit $F_{0,05}$
$F_{0,05}$[5;10]
Najdu v tabulkách $F_{0,95}$[10;5] =4,735

$$P(F < F_{0,05}) = 0,05$$

$F_{0,05}$[5;10] = 1/4,735=0,211

$$P(F < 0,211) = 0,05$$

Poslední 3 rozdělení budou důležitá při statistických hypotézách