

Programovanie v prostredí

III. Rozdelenia pravdepodobnosti

Aleš Kozubík

Katedra Matematických Metód a Operačnej Analýzy

23.11.2019

Aké funkcie sú k dispozícii

Pre väčšinu známych rozdelení sú k dispozícii tieto funkcie:

- **pravdepodobnostná funkcia** resp. **funkcia hustoty**, používa sa prefix **d**,
- **distribučná funkcia**, používa sa prefix **p**,
- **kvantilová funkcia**, používa sa prefix **q**,
- **generátor náhodných čísel**, používa sa prefix **r**.

Za príslušným prefixom nasleduje názov požadovaného rozdelenia s prípadnými argumentmi (parametrami rozdelenia).

Prehľad rozdelení – diskkrétne

Z diskrétnych rozdelení sú k dispozícii:

- **binomické** rozdelenie `binom` s parametrami `size` a `prob`,
- **geometrické** rozdelenie `geom` s parametrom `prob`,
- **hypergeometrické** rozdelenie `hyper` s parametrami `x`, `m`, `n`, `k`,
- **multinomické** rozdelenie `multinom` s parametrami `n`, `size`, `prob`,
- **negatívne binomické** rozdelenie `nbinom` s parametrami `size` a `prob`,
- **Poissonovo** rozdelenie `pois` s parametrom `lambda`.

Prehľad rozdelení – spojité

Z diskrétnych rozdelení sú implementované:

- **beta** rozdelenie beta s parametrami shape1, shape2 a ncp,
- **Cauchyho** rozdelenie cauchy s parametrami location a scale,
- **Chí-kvadrát** rozdelenie chisq s parametrom df,
- **exponenciálne** rozdelenie exp s parametrom rate
- **Fisherovo F** rozdelenie f s parametrami df1 a df2,
- **gama** rozdelenie gamma s parametrami shape a rate.

Spojité rozdelenia – pokračovanie

- **logistické** rozdelenie `logis` s parametrami `location` a `scale`,
- **logaritmicko-normálne** rozdelenie `lnorm` s parametrami `meanlog` a `sdlog`,
- **normálne** rozdelenie `norm` s parametrami `mean` a `sd`,
- **Studentovo t-rozdelenie** `t` s parametrom `df`,
- **rovnomerné** rozdelenie `unif` s parametrami `min` a `max`
- **Wiebullovo** rozdelenie `weibull` s parametrami `shape` a `scale`.

Náhodný výber hodnôt – základy

Základné situácie môžeme simulovať pomocou funkcie `sample()`.

Príklad

Chceme vybrať náhodných 5 celočíselných hodnôt v rozsahu od 1 do 40.

Riešenie:

```
> sample(1:40, 5)
[1] 18  2 38 13 32
>
```

Charakteristika funkcie `sample()`

Prvým argumentom (`x`) funkcie `sample` je vektor, z ktorého sa výber realizuje.

Druhým argumentom `size` je rozsah náhodného výberu.

Funkcia `sample` sa defaultne správa tak, že sa hodnoty vo výbere neopakujú. Ak to chceme zmeniť, je potrebné zadať hodnotu parametra `replace=TRUE`.

Simulácia hodu kockou

Príklad

Chceme simulovať výsledky 50 pokusov hodu kockou. V takom prípade sa budú výsledky hodov opakovať.

```
> sample(6,50,replace=TRUE)
 [1] 2 4 6 5 4 5 4 4 2 2 2 1 ...
[39] 5 4 6 4 5 6 6 3 6 5 3 1
>
```


Simulácia hodu mincou (falošnou)

Pre náhodný výber nemusia byť všetky alternatívy rovnako pravdepodobné.

Príklad

Chceme simulovať hod falošnou mincou, kde pravdepodobnosť padnutia „Panna“ je napr. 85 %.

Musíme vo funkcii `sample()` zadať ako vektor parameter `prob`.

```
> sample(c("Panna", "Orol"), 20, replace=TRUE,
+ prob=c(0.85, 0.15))
 [1] "Panna" "Panna" "Orol"  "Panna"
     "Panna" "Orol"  "Panna" "Panna" "Panna"
 [10] "Panna" "Panna" "Panna" "Panna" "Panna"
     "Panna" "Panna" "Panna" "Panna"
 [19] "Panna" "Panna"
```

Generovanie náhodných hodnôt z rozdelenia

Binomické rozdelenie

Pre generovanie hodnôt používame funkciu `rbinom()`.

Funkcia má tri argumenty:

- počet pozorovaní, ktoré chceme vybrať,
- počet pokusov, ktoré chceme v každom pozorovaní vykonať,
- pravdepodobnosť úspechu v každom pokuse.

Výsledkom je vektor, udávajúci počet úspešných pokusov v jednotlivých pozorovaniach.

Generovanie náhodných hodnôt z rozdelenia

Binomické rozdelenie

Príklad

Vykonávame výstupnú kontrolu v továrni, ktorá produkuje 150 výrobkov denne. Poruchovosť jednotlivých výrobkov je 3%, pričom každý poruchový výrobok sa musí opraviť. Simulujeme počet výrobkov, ktoré je potrebné opraviť každý deň počas týždňa.

Riešenie:

```
> rbinom(7, 150, 0.03)
[1] 4 1 6 5 5 1 4
>
```

Generovanie náhodných hodnôt z rozdelenia

Binomické rozdelenie, Bernoulliovské pokusy

Jednotlivý Bernoulliovský pokus môžeme modelovať nastavením počtu pozorovaní na hodnotu 1. (Např. 10 hodov mincou.)

```
> rbinom(10,1,0.5)
[1] 1 1 1 0 1 0 0 0 1 0
>
```

Alebo počty pozitívnych výsledkov pri 100 opakovaníach:

```
> rbinom(10,100,0.5)
[1] 50 54 44 58 48 48 53 48 45 49
>
```

Generovanie náhodných hodnôt z rozdelenia

Rovnomerné rozdelenie

Pre generovanie náhodných hodnôt z rovnomerného rozdelenia používame funkciu `runif()` s tromi parametrami:

- `n` rozsah náhodného výberu,
- `min` počiatočná hodnota pre rozdelenie,
- `max` koncová hodnota pre rozdelenie.

Generovanie náhodných hodnôt z rozdelenia

Rovnomerné rozdelenie – ukážka

Náhodný výber 7 hodnôt z rovnomerného rozdelenia na intervale $\langle 5, 10 \rangle$.

```
> runif(7, 5, 10)
[1] 6.231284 8.341515 8.832025 5.090891
     7.225523 5.094918 7.455470
>
```

Generovanie náhodných hodnôt z rozdelenia

Normálne rozdelenie

Najbežnejšie sa na charakterizovanie náhodnej zložky dejov používa normálne rozdelenie.

Pre generovanie náhodných hodnôt z normálneho rozdelenia používame funkciu `rnorm()` s tromi parametrami:

- `n` rozsah náhodného výberu,
- `mean` stredná hodnota rozdelenia,
- `sd` smerodajná odchýlka rozdelenia.

Generovanie náhodných hodnôt z rozdelenia

Normálne rozdelenie – ukážka

Generovanie 5 hodnôt zo štandardizovaného normálneho rozdelenia:

```
> rnorm(5,0,1)
[1] -1.8354235 -0.4797944 0.9080089 -1.3624083 1.0512592
>
```

Generovanie 5 hodnôt zo všeobecného normálneho rozdelenia:

```
> rnorm(5,10,2)
[1] 11.526437 8.728263 12.484347 11.155544 10.226890
>
```


Pravdepodobnostná funkcia

Binomické rozdelenie

Príklad

Vykonáme 10 hodov kockou. Aká je pravdepodobnosť, že šesťka padne práve 2-krát?

Riešenie:

Využijeme funkciu `dbinom()` s parametrami `x=2`, `size=10` a `prob=1/6`. Tak dostávame

```
> dbinom(2,10,1/6)
[1] 0.29071
>
```

Pravdepodobnostná funkcia

Binomické rozdelenie – graf

Príklad

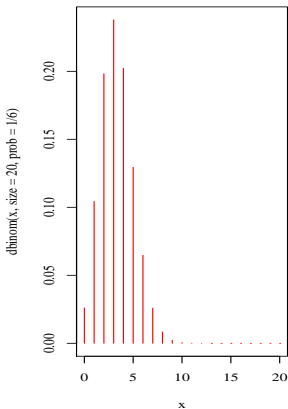
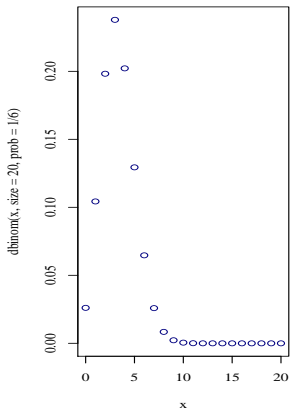
Zobrazme pravdepodobnostnú funkciu rozdelenia počtu padnutých šesťiek pri dvadsiatich hodoch kockou.

Riešenie:

```
x <- 1:20
> dev.new(width=14,height=8)
> par(mfrow=c(1,2),family="Times")
> plot(x,dbinom(x,size=20,prob=1/6),type="p",col="navy")
> plot(x,dbinom(x,size=20,prob=1/6),type="h",col="red")
```

Pravdepodobnostná funkcia

Binomické rozdelenie – graf, výsledok



Pravdepodobnostná funkcia

Poissonovo rozdelenie

Príklad

Cez most prejde za minútu v priemere 12 áut. Aká je pravdepodobnosť, že za určitú minútu prejde cez most od 10 do 15 áut?

Riešenie:

Využijeme funkciu `dpois()` s parametrami

$x=10, 11, 12, 13, 14, 15$, a $\lambda=12$. Tak dostávame

```
> dpois(10,12)+dpois(11,12)+dpois(12,12)+dpois(13,12)+  
+ dpois(14,12)+dpois(15,12)  
[1] 0.6020235  
>
```

Pravdepodobnostná funkcia

Poissonovo rozdelenie – graf

Príklad

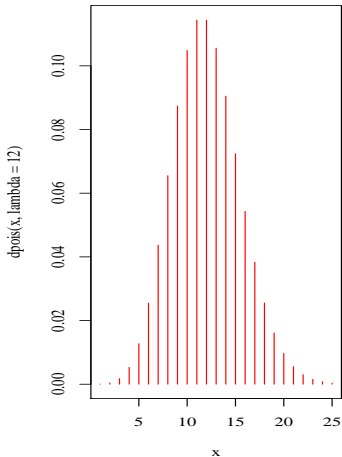
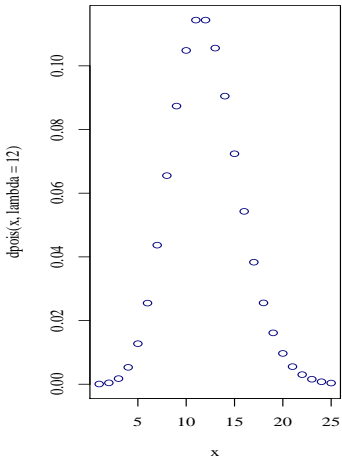
Zobrazme pravdepodobnostnú funkciu rozdelenia počtu áut, ktoré v danej minúte prejdú cez most z predchádzajúceho príkladu.

Riešenie:

```
x<-1:25  
> dev.new(width=14,height=8)  
> par(mfrow=c(1,2),family="Times")  
> plot(x,dpois(x,lambda=12),type="p",col="navy")  
> plot(x,dpois(x,lambda=12),type="h",col="red")
```

Pravdepodobnostná funkcia

Poissonovo rozdelenie – graf, výsledok



Hustota pravdepodobnosti

Normálne rozdelenie

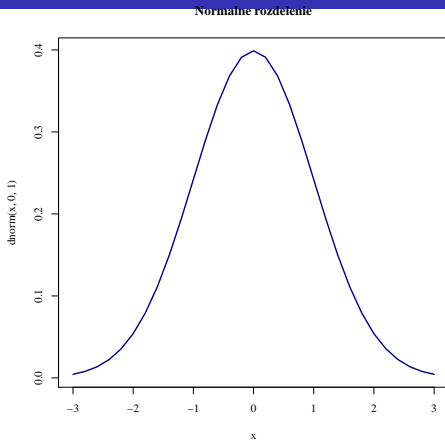
Normálne rozdelenie s parametrami $\text{mean}=0$ a $\text{sd}=1$ nazývame štandardizované normálne rozdelenie.

Jeho hustotu si môžeme zobrazit pomocou funkcie `dnorm()` takto:

```
> x<-seq(-3,3,0.2)
> par(family="Times")
> plot(x,dnorm(x,0,1),type="l",col="navy",
+ lwd=2,main="Normalne rozdelenie")
>
```

Hustota pravdepodobnosti

Normálne rozdelenie – graf



Hustota pravdepodobnosti

Exponenciálne rozdelenie

Popisuje dobu čakania na výskyt nezávislých náhodných udalostí.

Ak je priemerná doba čakania na výskyt udalosti μ , tak jeho parametrom je rate, čo je hodnota $\frac{1}{\mu}$.

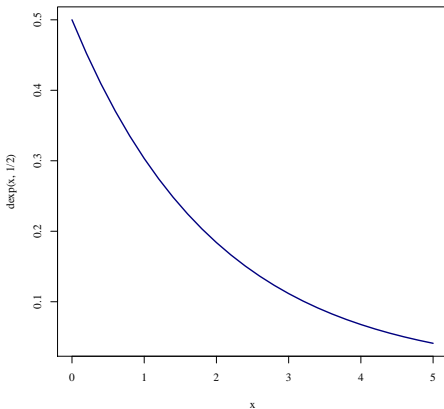
hustotu exponenciálneho rozdelenia s parametrom $\frac{1}{2}$ si môžeme zobraziť pomocou funkcie `dexp()`.

```
> x<-seq(0,5,0.2)
> par(family="Times")
> plot(x,dexp(x,1/2),type="l",col="navy",
+ lwd=2,main="Exponencialne rozdelenie")
```

Hustota pravdepodobnosti

Exponenciálne rozdelenie – graf hustoty

Exponenciálne rozdelenie



Hustota pravdepodobnosti

χ^2 rozdelenie

Rozdelenie súčtu $X_1^2 + X_2^2 + \dots + X_n^2$, kde jednotlivé náhodné premenné majú štandardizované normálne rozdelenia.

Hodnota n sa nazýva počet stupňov voľnosti.

Grafy pre rôzny počet stupňov voľnosti získame pomocou funkcie `dchisq()`. Jej argumentmi sú vektor nezáporných hodnôt x a počet stupňov voľnosti $df = \langle \text{počet} \rangle$.

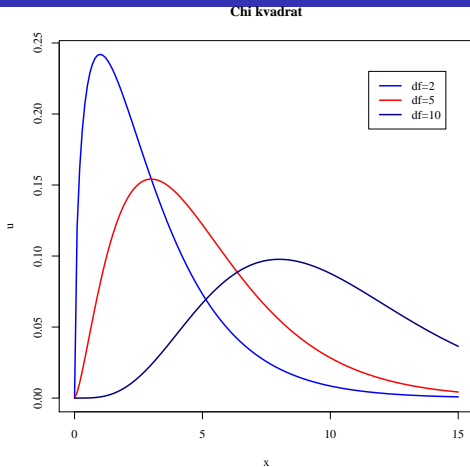
Hustota pravdepodobnosti

χ^2 rozdelenie – zdroják

```
> x<-seq(0,15,0.1)
> u<-dchisq(x,df=2)
> z<-dchisq(x,df=5)
> y<-dchisq(x,df=10)
> plot(x,u,col="blue",type="l",lwd=2,main="Chi_kvadrat")
> lines(x,z,col="red",lwd=2)
> lines(x,y,col="navy",lwd=2)
> legend(11.5,0.23,legend=c("df=2","df=5","df=10"),
+ col=c("blue","red","navy"),lty=c(1,1,1),
+ lwd=c(2,2,2))
>
```

Hustota pravdepodobnosti

χ^2 rozdelenie – graf



Distribučná funkcia

Binomické rozdelenie

Príklad

Test obsahuje 12 otázok, na každú z nich ponúka 5 alternatív odpovedí, z ktorých je vždy len jedna správna. Určte pravdepodobnosť uhádnutia nanajvýš štyroch správnych odpovedí, ak študent vyberá odpovede náhodne.

Jednou možnosťou je sčítať pravdepodobnosti uhádnutia žiadnej, práve jednej, atď. až práve štyroch správnych odpovedí. využijeme hustotu binomického rozdelenia.

```
> dbinom(0, size=12, prob=0.2)+dbinom(1, size=12, prob=0.2)+  
+ dbinom(2, size=12, prob=0.2)+dbinom(3, size=12, prob=0.2)+  
+ dbinom(4, size=12, prob=0.2)  
[1] 0.9274445
```

Distribučná funkcia

Binomické rozdelenie

Príklad

Test obsahuje 12 otázok, na každú z nich ponúka 5 alternatív odpovedí, z ktorých je vždy len jedna správna. Určte pravdepodobnosť uhádnutia nanajvýš štyroch správnych odpovedí, ak študent vyberá odpovede náhodne.

Druhou možnosťou je využitie distribučnej funkcie binomického rozdelenia `pbinom()`.

```
> pbinom(4, size=12, prob=0.2)
[1] 0.9274445
>
```

Distribučná funkcia

Poissonovo rozdelenie

Príklad

Cez most prejde za minútu v priemere 12 áut. Aká je pravdepodobnosť, že za určitú minútu prejde cez most od 10 do 15 áut?

Úlohu sme už riešili pomocou hustoty `dpois()`.

Pomocou distribučnej funkcie určíme pravdepodobnosť $\mathbb{P}[10 \leq X \leq 15]$ ako rozdiel $\mathbb{P}[X \leq 15] - \mathbb{P}[X \leq 10]$. Teda

```
> ppois(15,12) - ppois(9,12)
[1] 0.6020235
>
```


Distribučná funkcia

Poissonovo rozdelenie

Príklad

Cez most prejde za minútu v priemere 12 áut. Aká je pravdepodobnosť, že za určitú minútu prejde cez 17 a viac áut?

Pomocou distribučnej funkcie určíme pravdepodobnosť $\mathbb{P}[X \geq 15]$ ako rozdiel $1 - \mathbb{P}[X \leq 16]$. Teda

```
> 1-ppois(16,12)
[1]
> 0.101291
```

Tiež môžeme použiť premennú `lower=FALSE`:

```
> ppois(16,12,lower=FALSE)
[1]
```

Distribučná funkcia

Exponenciálne rozdelenie

Príklad

Predpokladajme, že priemerná doba obsluhy pri pokladni supermarketu je 3 minúty. Aká je pravdepodobnosť, že obsluha zákazníka neprekročí dobu 2 minút?

Premenná λ je prevrátenou hodnotou priemernej doby obsluhy, teda $\frac{1}{3}$. Použijeme distribučnú funkciu exponenciálneho rozdelenia `pexp()`.

```
> pexp(2, 1/3)
[1] 0.4865829
>
```

Distribučná funkcia

Normálne rozdelenie

Príklad

Detské IQ má normálne rozdelenie so strednou hodnotou $\mu = 100$ a smerodajnou odchýlkou $\sigma = 15$. Aký podiel detí má IQ v rozpätí 80 až 120?

Najskôr si pripravíme hodnoty pre výpočet a zobrazenie výsledku

```
> mean=100; sd=15
> dolnah=80; hornah=120
> x <- seq(-4,4,length=100)*sd + mean
> y <- dnorm(x,mean,sd)
> plot(x, y, type="n", xlab="Hodnoty IQ", ylab="",
      main="Normalne rozdelenie", axes=FALSE)
# bez zobrazenia krivky, iba nastavenie oblasti
```

Distribučná funkcia

Normálne rozdelenie

Pomocou funkcie `polygon()` vyfarbíme plochu zodpovedajúcu hľadanej pravdepodobnosti. Argumentmi funkcie sú vektory `x`, `y` udávajúce súradnice vrcholov polygónu.

```
# zobrazenie grafu hustoty  
> lines(x, y)  
# indikátor intervalu  
> i <- x >= dolnah & x <= hornah  
> polygon(c(dolnah, x[i], hornah), c(0, y[i], 0), col="red")
```

Distribučná funkcia

Normálne rozdelenie

Dopočítame hľadanú pravdepodobnosť (veľkosť vyfarbenej plochy) a zobrazíme ju.

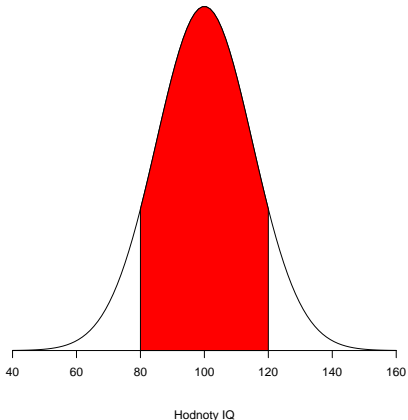
```
> area <- pnorm(hornah, mean, sd) - pnorm(dolnah, mean, sd)
> result <- paste("P(",hornah,"< IQ <",dolnah,") = ",
+   signif(area, digits=3))
> mtext(result,3)
# dokreslenie osi x
> axis(1, at=seq(40, 160, 20), pos=0)
```

Distribučná funkcia

Normálne rozdelenie – výsledok

Normálne rozdelenie

$$P(80 < IQ < 120) = 0.818$$



Kvantily normálneho rozdelenia

Príklad

Výrobca špeciálnych rukavíc jednotnej univerzálnej veľkosti chce navrhnúť rukavicu tak, aby vyhovovala aspoň 99 percentám populácie. Je známe, že veľkosť ruky (meraná ako rozpätie od palca k malíčku) má normálne rozdelenie so strednou hodnotou 195 mm a smerodajnou odchýlkou 17 mm. Na aký rozsah rozpätia musí byť rukavica schopná sa prispôbiť?

Najskôr určíme hodnotu rozpätia prstov takú, že len 0.5 percenta populácie má rozpätie menšie

```
> qnorm(0.005, 195, 17)
[1] 151.2109
```

Kvantily normálneho rozdelenia

Podobne určíme hodnotu rozpätia, ktorú presahuje len 0.5 percenta populácie

```
> qnorm(0.005,195,17,lower.tail=F)
[1] 238.7891
# funguje aj qnorm(0.995,195,17)
```

Rukavica teda musí vyhovovať rozpätiu prsov v rozsahu 151 až 239 mm.

Kvantily Poissonovho rozdelenia

Príklad

Počet homárov objednaných v reštaurácii počas jedného dňa sa riadi Poissonovým rozdelením so strednou hodnotou 20. Ak chce manažér reštaurácie uspokojiť aspoň 80 % objednávok, koľko homárov má objednať každý deň?

Príslušný počet homárov určíme pomocou kvantilovej funkcie Poissonovho rozdelenia

```
> qpois(0.8, 20)
[1] 24
```

Každý deň je teda potrebné objednať 24 homárov.

Funkcia hazardu a prežitia

Funkcia prežitia vyjadruje pravdepodobnosť, že sledovaný objekt prežije časový okamih t , tj.

$$S(t) = \mathbb{P}(T > t), \quad 0 < t < \infty.$$

Funkcia prežitia býva definovaná aj prostredníctvom **funkcie hazardu**, ktorá je formálne vyjadrená ako

$$h(t) = \lim_{\delta \rightarrow 0} \frac{\mathbb{P}(t < T < t + \delta | T > t)}{\delta}.$$

Táto funkcia sa tiež nazýva intenzitou úmrtnosti.

Vyjadrenie funkcie hazardu a prežitia

Funkcia prežitia je doplnkovou pravdepodobnosťou ku distribučnej funkcii $F(t) = \mathbb{P}(T \leq t)$. Pre hustotu pravdepodobnosti potom platí

$$f(t) = -\frac{d}{dt}S(t) = \frac{d}{dt}F(t).$$

Funkcia hazardu je potom s hustotou pravdepodobnosti spojená vzťahom

$$h(t) = \frac{f(t)}{S(t)}.$$

Uživateľom definované funkcie

Uživateľ definuje vlastné funkcie pomocou konštrukcie:

```
mojaFcia <- function(argumenty){  
  telo funkcie  
  return(objekt)  
}
```

Funkciu potom voláme pomocou konštrukcie

```
mojaFcia(hodnoty argumentov).
```

Funkcia hazardu pre Weibullovo rozdelenie

V analýze prežitia sa často používa Weibullovo rozdelenie, ktoré ponúka väčšiu flexibilitu než rozdelenie exponenciálne.

Príslušnú funkciu hazardu definujeme takto:

```
> hweibull <- function(x, shape, scale)
+ dweibull(x, shape=shape, scale=scale)
+ /pweibull(x, shape=shape, scale=scale,
+ lower.tail=FALSE)
```

Pretože funkcia je definovaná jediným príkazom, nemusíme používať zátvorky {}.

Zobrazenie funkcie hazardu

Pre zobrazenie funkcie pre rôzne hodnoty parametra shape použijeme funkciu `curve()`.

```
> curve(hweibull(x, shape=1.5, scale=1/0.03),  
+ from=0, to=80, ylab="Hazard", xlab="Time", col="red")  
> curve(hweibull(x, shape=1, scale=1/0.03),  
+ from=0, to=80, ylab="Hazard", xlab="Time", col="black",  
+ add=TRUE)  
> curve(hweibull(x, shape=0.75, scale=1/0.03),  
+ from=0, to=80, ylab="Hazard", xlab="Time",  
+ col="navy", add=TRUE)
```

Parameter `add=TRUE` zabezpečí pridanie príslušnej krivky do existujúceho grafu.

Zobrazenie funkcie hazardu – výsledok

