

Statistika a spolehlivost v lékařství

1. cvičení, 1. dubna 2014

- **Příklad 1** Tahová síla papíru používaného pro výrobu potravinových sáčků je důležitá charakteristika kvality. Je známo, že síla (označme x) podléhá normálnímu rozložení se střední hodnotou $\mu = 40\text{lb/in}^2$ a standardní odchylkou $\sigma = 2\text{lb/in}^2$, tj. $x \approx N(40, 2^2)$. Odběratel sáčků vyžaduje, aby sáčky měly sílu alespoň $\mu = 35\text{lb/in}^2$. Jaká je pravděpodobnost, že sáček vyhoví požadavkům odběratele?

Řešení: Pravděpodobnost při normálním rozložení je dána vztahem:

$$P(x \leq a) = F(a) = \int_{-\infty}^a \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Tento integrál nemůže být vyčíslen analyticky, nicméně můžeme provést substituci $z = \frac{x-\mu}{\sigma}$, čímž dostaneme funkci, která není závislá na μ a σ^2 :

$$P(x \leq a) = P\left(z \leq \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right),$$

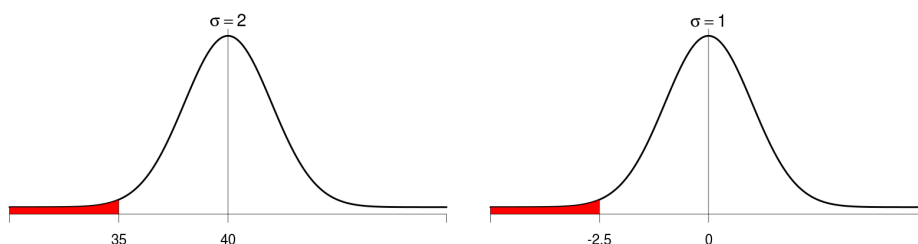
kde $\Phi(\cdot)$ standardnímu normálnímu rozložení ($N(0, 1)$). My chceme vyjádřit $P(x \geq 35) = 1 - P(x \leq 35)$.

$$P(x \leq 35) = P\left(z \leq \frac{35-40}{2}\right) = P(z \leq -2.5) = \Phi(-2.5) = 1 - \Phi(2.5) = 0.0062.$$

Požadovaná pravděpodobnost je tedy

$$P(x \geq 35) = 1 - P(x \leq 35) = 1 - 0.0032 = 0.9938.$$

Pozn: vyčíslení funkce Φ bývá součástí statistických tabulek (viz např. D. Montgomery: *Introduction to Statistical Quality Control*).



Obrázek 1: Standardizace normálního rozdělení.

1 Charakteristiky spolehlivosti

Mějme spojitou proměnnou (typicky čas $t \geq 0$) a uvažujme pravděpodobnost, že porucha, resp. čas poruchy, ξ (spojitá nezávislá veličina) nastane v čase t . **Pravděpodobnost poruchy** $Q(t)$ definujeme jako distribuční funkci spojitě nezávislé veličiny ξ (čas poruchy):

$$Q(t) = F(t) = P(\xi \leq t).$$

Pravděpodobnost bezporuchového provozu $R(t)$ je definována jako doplněk k pravděpodobnosti poruchy:

$$R(t) = 1 - Q(t),$$

což je pravděpodobnost, že porucha nenastane dříve než v čase t :

$$R(t) = P(\xi > t).$$

Hustota pravděpodobnosti $f(t)$ je derivace distribuční funkce a tedy:

$$f(t) = \frac{\partial F(t)}{\partial t} = \frac{\partial(1 - R(t))}{\partial t} = -\frac{\partial R(t)}{\partial t}.$$

Další veličinou popisující spolehlivost je **intenzita poruch** $\lambda(t)$, kterou definujeme jako podíl hustoty a pravděpodobnosti bezporuchového provozu:

$$\lambda(t) = \frac{f(t)}{R(t)}.$$

Střední doba bezporuchového provozu:

$$T_s = \int_0^{\infty} t f(t) dt.$$

Rozptyl náhodné doby poruchy:

$$D = E(\xi^2) - E^2(\xi) = \int_0^{\infty} t^2 f(t) dt - T_s^2.$$

- **Příklad 2** Nalezněte vztah pro výpočet pravděpodobnosti bezporuchového provozu z intenzity, tj. funkční předpis

$$R(t) = fce(\lambda(t)).$$

Řešení:

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\frac{\partial Q(t)}{\partial t}}{R(t)} = -\frac{\frac{\partial R(t)}{\partial t}}{R(t)}$$

Z toho po vynásobení ∂t dostáváme

$$-\lambda(t)\partial t = \frac{\partial R(t)}{R(t)}.$$

A po integraci¹ (uvědomte si, že integrál z podílu derivace funkce a této funkce je logaritmus dané funkce):

$$-\int_0^T \lambda(t) dt = [\ln R(t)]_0^T = \ln R(T) - \ln R(0).$$

Pravděpodobnost bezporuchového provozu v čase nula je $R(0) = 1$, tedy $\ln R(0) = 0$ a proto druhý člen z levého výrazu vypadne. Nakonec po úpravě získáváme výsledný vztah:

$$R(t) = e^{-\int_0^t \lambda(\tau) d\tau}.$$

- **Příklad 3** Nalezněte vztah pro výpočet střední doby bezporuchového provozu T_s z pravděpodobnosti bezporuchového provozu $R(t)$.

Řešení:

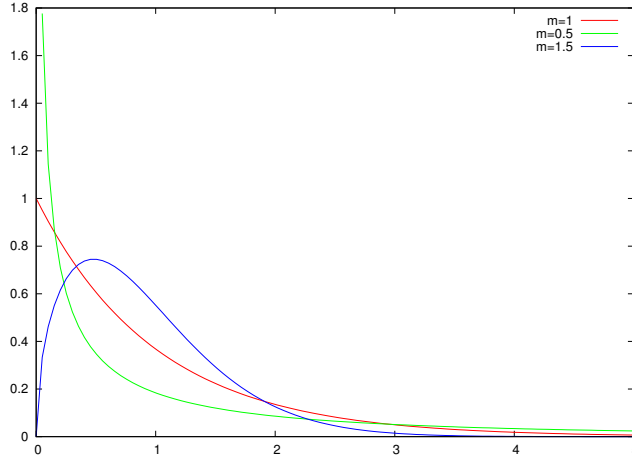
$$T_s = \int_0^\infty t f(t) dt = \int_0^\infty t \left(-\frac{dR(t)}{dt}\right) dt = \left| \begin{array}{l} u' = R' \\ v = t \end{array} \right| \begin{array}{l} u = R \\ v' = 1 \end{array} = -[Rt]_0^\infty + \int_0^\infty R dt = \int_0^\infty R(t) dt$$

První závorka $[Rt]_0^\infty$ je rovná nule, protože pro $t \rightarrow \infty$ platí typicky $R(\infty) \rightarrow 0$.

¹Určitým integrálem jako funkci horní meze, neboť nám jde také o určení (vyčíslení) konstatny, tu bychom jinak ztratily.

2 Exponenciální rozdělení

$$\begin{aligned}\lambda(t) &= \lambda && \lambda > 0, t \geq 0 \\ R(t) &= e^{-\int \lambda dt} = e^{-\lambda t} \\ Q(t) &= 1 - e^{-\lambda t} \\ f(t) &= \lambda e^{-\lambda t}\end{aligned}$$



Obrázek 2: Distribuční funkce exponenciálního rozdělení pro dvě různé hodnoty λ .

- **Příklad 4** Nalezněte vztah pro výpočet střední doby bezporuchového provozu T_s .

Řešení:

$$\begin{aligned}T_s &= \int_0^\infty t f(t) dt = \int_0^\infty t \lambda e^{-\lambda t} dt = \left| \begin{array}{l} u' = \lambda e^{-\lambda t} \\ v = t \end{array} \right| \begin{array}{l} u = -e^{-\lambda t} \\ v' = 1 \end{array} \Big|_0^\infty = \left[\left(-\frac{1}{\lambda} - t\right) e^{-\lambda t} \right]_0^\infty = \\ &= \lim_{t \rightarrow \infty} \left(-\frac{1}{\lambda} - t\right) e^{-\lambda t} + \frac{1}{\lambda} = \frac{1}{\lambda}.\end{aligned}$$

- **Příklad 5** Nalezněte vztah pro výpočet rozptylu náhodné doby poruchy D .

Řešení:

$$\begin{aligned}D &= \int_0^\infty t^2 \lambda e^{-\lambda t} dt - T_s^2 = \left| \begin{array}{l} u' = \lambda e^{-\lambda t} \\ v = t^2 \end{array} \right| \begin{array}{l} u = -e^{-\lambda t} \\ v' = 2t \end{array} \Big|_0^\infty = \left[-t^2 e^{-\lambda t} + 2 \int t e^{-\lambda t} dt \right]_0^\infty - T_s^2 = \\ &= \left| \begin{array}{l} u' = e^{-\lambda t} \\ v' = t \end{array} \right| \begin{array}{l} u = -\frac{1}{\lambda} e^{-\lambda t} \\ v = 1 \end{array} \Big|_0^\infty = \left[-t^2 e^{-\lambda t} - \frac{2t}{\lambda} e^{-\lambda t} + \frac{2}{\lambda} \int e^{-\lambda t} dt \right]_0^\infty - T_s^2 = \\ &= \lim_{t \rightarrow \infty} (-t^2 - 2t) e^{-\lambda t} + \frac{2}{\lambda^2} e^{-\lambda t} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.\end{aligned}$$

Pozn.: Výměna neporušeného prvku s exponenciálním rozdělením poruch nepřinese zlepšení bezporuchovosti.

- **Příklad 6** Předpokládejme nyní stroj, který se porouhá čtyřikrát za týden. Jaká je doba T_β zaručující, že pravděpodobnost bezporuchového provozu je $\beta = R(T_\beta) = 0.99$?

Řešení: V našem případě je intenzita poruch konstantní a rovna $\lambda = 4\text{týden}^{-1} = \frac{4}{7 \cdot 24 \cdot 60} \text{min}^{-1} \doteq 4 \cdot 10^{-4} \text{min}^{-1}$. Nyní hledáme t takové, že $R(T_\beta) = \beta$. Po dosazení

$$R(t) = e^{-\int_0^T \lambda(t)dt} = e^{-\lambda T_\beta} = \beta. \quad (1)$$

dostáváme

$$-\lambda T_\beta = \ln \beta \quad (2)$$

a tedy

$$T_\beta = -\frac{1}{\lambda} \ln \beta = -\frac{1}{4 \cdot 10^{-4}} \ln 0.99 = 25 \text{min}.$$

Je dobré si uvědomit, že doba bezporuchového provozu T_β je velmi citlivá na volbu β . To je způsobeno logaritmem ve vzorci 2. Už při volbě $\beta = 0.90$ (tedy změně o 0.09) vzroste T_β zhruba desetkrát na $T_\beta = 263$ minut.

- **Příklad 7** Předpokládejme součástku, jejíž spolehlivost má exponenciální rozdělení. Jaká je pravděpodobnost, že se součástka porouchá před očekávanou životností?

Řešení: Intenzita poruch je λ a očekávaná životnost (střední doba bezporuchového provozu) je $\frac{1}{\lambda}$. Chceme tedy vyčíslit:

$$P(x \leq \frac{1}{\lambda}) = \int_0^{\frac{1}{\lambda}} \lambda e^{-\lambda t} dt = 1 - e^{-1} = 0.63212$$

- **Příklad 8** Výrobce kalkulaček nabízí jednorocní záruku. Pokud je kalkulačka v záruce vadná, je zákazníkovi vyměněna za novou. Spolehlivost kalkulačky je popsána následující hustotou pravděpodobnosti:

$$f(x) = 0.125e^{-0.125x}, x > 0$$

- (a) Kolik procent kalkulaček se rozbije během záruční doby?
 (b) Výrobní cena kalkulačky je \$50 a zisk z jedné kalkulačky je \$25. Jaký je efekt výměny kalkulačky na čistý zisk?
 (c) Jaká musí být prodejní cena kalkulačky, abychom při výrobní ceně kalkulačky \$50 měli čistý zisk (po započítání ztráty z výměny vadné kalkulačky) \$25.

Řešení: (a) Dle hustoty pravděpodobnosti podléhá spolehlivost kalkulačky exponenciálnímu rozdělení s $\lambda = 0.125$. Pravděpodobnost poruchy v čase $t = 1$ rok je tedy dána takto:

$$F(1) = 1 - R(t) = 1 - e^{-\lambda t} = 1 - e^{-0.125} \approx 1 - 0.8825 = 0.1175,$$

tj. rozbije se cca 11.75% kalkulaček.

(b) Za každou rozbitou kalkulačku je nutno k nákladům přičíst \$50, což jsou výrobní náklady na výměnu kalkulačky. To sníží profit z jedné kalkulačky o $0.1175 \times 50 = 5.875$ \$.

(c) Výsledná cena je součtem výrobní ceny, ceny za výměnu kalkulačky a požadovaného zisku: $\text{cena} = 50 + 5.875 + 25 = \80.875

- **Příklad 9** Vypočtete hodnotu mediánu pro exponenciální rozdělení s parametrem $\lambda = 0.125$.

Řešení: Pro medián t_m platí:

$$\int_{-\infty}^{t_m} \lambda e^{-\lambda t} dt = 0.5$$

Po integraci

$$| -e^{-\lambda t} |_0^{t_m} = 0.5$$

Dosadíme meze:

$$-e^{-\lambda t_m} + 1 = 0.5$$

a vypočteme medián t_m :

$$t_m = \frac{-\ln 0.5}{\lambda} = \frac{\ln 2}{\lambda}.$$

Po dosazení:

$$t_m = \frac{\ln 2}{0.125} = 5.54.$$

- **Příklad 10** Výrobce HDD uvádí hodnotu $MTBF=10^6$ hodin. Kolik procent disků se porouchá v prvním roce?

Řešení:

MTFB (Mean Time Between Failures) je střední doba mezi poruchami, tj. $T_s = MTBF$, tedy $\lambda = \frac{1}{T_s} = \frac{1}{10^6} = 10^{-6}$. Pravděpodobnost poruchy v 1. roce je:

$$Q(t) = 1 - R(t) = 1 - e^{-\lambda t} = 1 - e^{-10^{-6} \cdot 8760} = 1 - 0.9913 = 0.0087.$$

Pozn: počet hodin v roce je $24 \cdot 365 = 8760$.

U HDD se často jako parametr spolehlivosti uvádí AFR (Annualized failure rate), který je definován jako

$$AFR = 1 - e^{-8760/MTBF},$$

kde $MTBF$ je střední doba mezi poruchami v hodinách. AFR tedy není nic jiného než nám známá pravděpodobnost poruchy $ARF = Q(1rok)$.

3 Rayleighovo rozdělení

$$\begin{aligned}\lambda(t) &= kt & k \geq 0, t \geq 0 \\ R(t) &= e^{\int kt dt} = e^{-\frac{k}{2}t^2} \\ Q(t) &= 1 - e^{-\frac{k}{2}t^2} \\ f(t) &= kte^{-\frac{k}{2}t^2} \\ D &= (2 - \frac{\pi}{2})\frac{1}{k}\end{aligned}$$

► **Příklad 11** Nalezněte vztah pro výpočet střední doby bezporuchového provozu T_s .

Řešení: Postup, při kterém bychom se snažili aplikovat *per partes* není shůdný, neboť mocnina t se nesníží, proto výraz upravíme na vztah $\int e^{-\frac{kt^2}{2}}$.

$$\begin{aligned}T_s &= \int_0^\infty t f(t) dt = \int_0^\infty t \frac{d(1 - R(t))}{dt} dt = - \int_0^\infty t \frac{dR}{dt} dt = \left| \begin{array}{l} u' = R' \\ v = t \end{array} \right| \begin{array}{l} u = R \\ v' = 1 \end{array} \Big|_0^\infty \\ &= - [tR(t)]_0^\infty + \int_0^\infty R(t) dt = - \left[te^{-\frac{kt^2}{2}} \right]_0^\infty + \int_0^\infty e^{-\frac{kt^2}{2}} dt\end{aligned}$$

První závorka je nulová pro oba limitní případy ($t \rightarrow 0, t \rightarrow \infty$). K výpočtu integrálu použijeme následující vztah:

$$\begin{aligned}N(0, 1) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \\ \int_{-\infty}^\infty N(0, 1) dx &= 1 \Rightarrow \int_0^\infty e^{-\frac{x^2}{2}} = \frac{1}{2} \sqrt{2\pi} = \sqrt{\frac{\pi}{2}}\end{aligned}$$

Abychom mohli tento výsledek použít, musíme provést následující substituci: $\left. \begin{array}{l} \sqrt{kt} = a \\ t = \frac{1}{\sqrt{k}} a \\ dt = \frac{1}{\sqrt{k}} da \end{array} \right\}$.

Potom

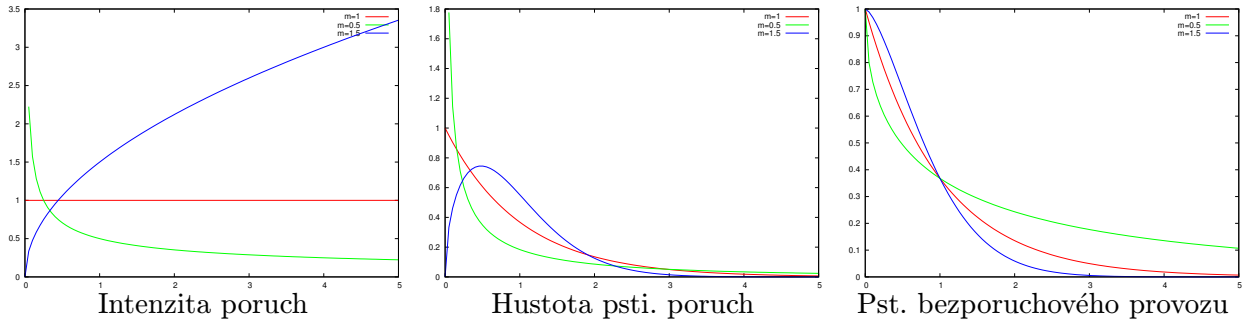
$$T_s = \frac{1}{\sqrt{k}} \int_0^\infty e^{-\frac{a^2}{2}} da = \frac{1}{\sqrt{k}} \sqrt{\frac{\pi}{2}} = \sqrt{\frac{\pi}{2k}}$$

4 Weibullovo rozdělení

$$\lambda(t) = \frac{m}{t_0} t^{m-1} \quad m > 0, t_0 > 0, t \geq 0,$$

kde t_0 je tzv. *scale*, normalizační konstanta (časová) a m je bezrozměrný parametr, reprezentující tvar charakteristiky, tzv. *shape* (viz. Obr. 3).

$$\begin{aligned}R(t) &= e^{-\frac{t^m}{t_0^m}}, \\ Q(t) &= 1 - e^{-\frac{t^m}{t_0^m}}, \\ f(t) &= \frac{m}{t_0} t^{m-1} e^{-\frac{t^m}{t_0^m}}, \\ T_s &= t_0^m \Gamma\left(\frac{1}{m} + 1\right), \\ D &= t_0^{\frac{2}{m}} \left[\Gamma\left(\frac{2}{m} + 1\right) - \Gamma^2\left(\frac{1}{m} + 1\right) \right],\end{aligned}$$



Obrázek 3: Příklady Weibullova rozdělení pro různé hodnoty m

kde

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt.$$

Pro $x \in \mathbb{N}$ platí:

$$\Gamma(x) = (x - 1)!$$

Pozn.:

Pro $m = 1$ → exponenciální rozdělení.

Pro $m = 2$ → Rayleighovo rozdělení,

Pro $m < 1$ → průběh $\lambda(t)$ klesající průběh, což je vhodné pro popis počátečních fází provozu.

- **Příklad 12** Intenzity poruch jednotlivých rozdělení můžeme kombinovat. To se hodí pro modelování složitějších (reálných) průběhů intenzit, např. pro popis tzv. vanové charakteristiky (viz 2. cvičení). Uvažujme intenzitu poruch, která je popsána kombinací dvou Weibullových rozdělení:

$$\lambda(t) = \frac{m_1}{t_1} t^{m_1-1} + \frac{m_2}{t_2} t^{m_2-1}$$

Uvažujme, že $t_1 = 1$, $t_2 = 10^6$ a shape konstanty $m_1 = 0.1$ a $m_2 = 5$. Odpovídající intenzita poruch je zobrazena na Obr. 4 Pro tuto intenzitu odvoďte $R(t)$ a $f(t)$.

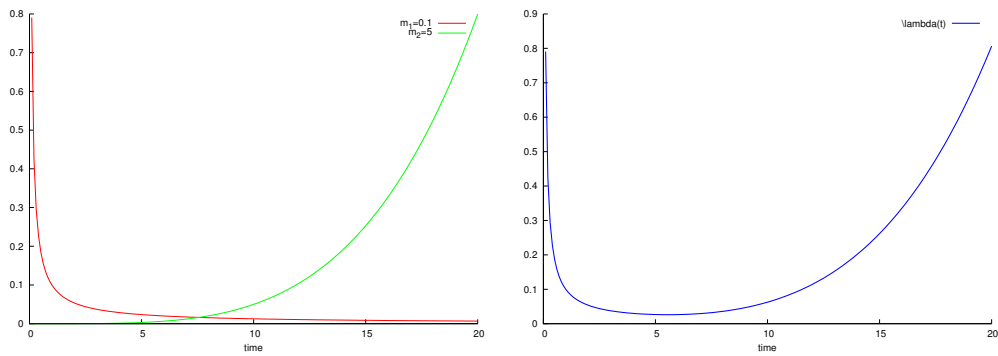
Řešení:

$$R(t) = e^{-\int_0^t \lambda(\tau) d\tau} = e^{-\left[m_1 \frac{t^{m_1}}{m_1} + 10^{-6} m_2 \frac{t^{m_2}}{m_2}\right]_0^t} = e^{-\left[t^{m_1} + 10^{-6} t^{m_2}\right]_0^t} = e^{-(t^{m_1} + 10^{-6} t^{m_2})}$$

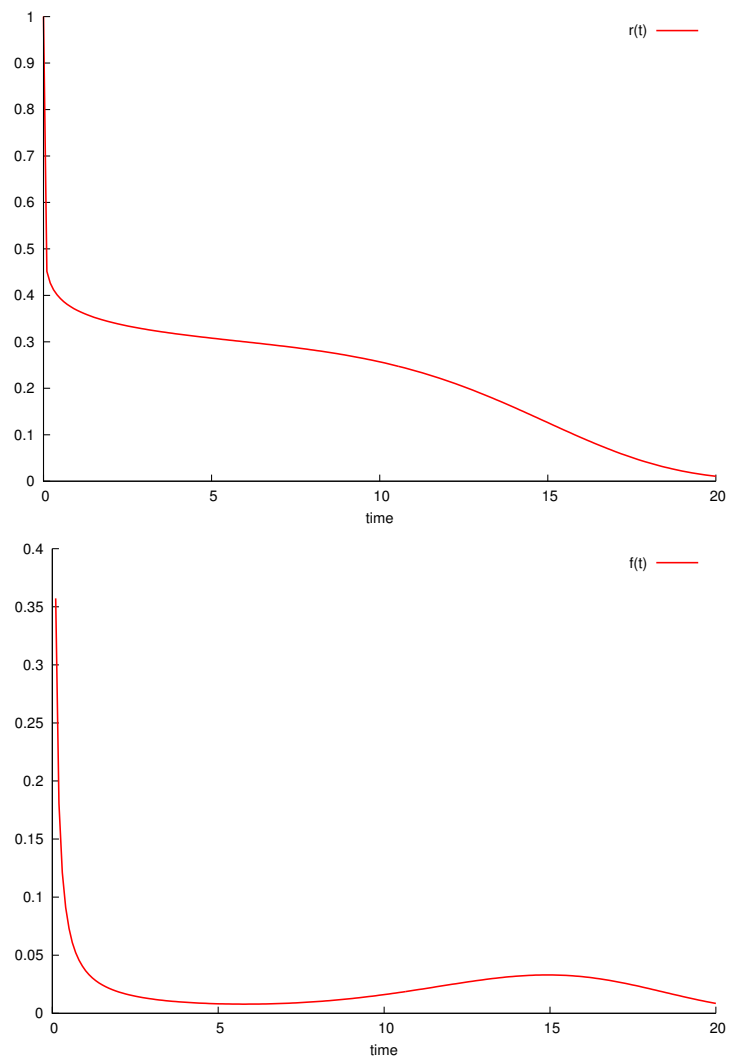
Obdobně lze odvodit hustotu pravděpodobnosti poruch:

$$f(t) = \frac{-\partial R(t)}{\partial t} = (m_1 t^{m_1-1} + 10^{-6} m_2 t^{m_2-1}) e^{-(t^{m_1} + 10^{-6} t^{m_2})}$$

Průběh $R(t)$ a $f(t)$ je zobrazen na Obr.5.



Obrázek 4: Intenzita poruch složená z exp. a Weibulova rozdělení



Obrázek 5: $R(t)$ a $f(t)$ pro složení dvou Weibullových rozdělení

- **Příklad 13** Máme systém s intenzitou poruch danou Weibullovým rozdělením s $m = \frac{1}{4}$. Intenzita poruch v čase $t = 1$ je $\lambda(1) = \frac{1}{5}$. Určete střední dobu poruch tohoto systému.

Řešení: Nejprve je nutné určit parametr t_0 Weibullova rozdělení:

$$\lambda(1) = \frac{1}{4t_0} 1^{\frac{1}{4}-1} = \frac{1}{5} \Rightarrow t_0 = \frac{5}{4}.$$

Střední doba bezporuchového provozu je potom dána

$$T_s = t_0^{1/m} \Gamma\left(\frac{1}{m} + 1\right) = \left(\frac{5}{4}\right)^4 \Gamma(5) = \frac{625}{256} 4! = 58.6.$$

- **Příklad 14** Intenzita poruch je popsána funkcí $\lambda(t) = at^2 + \left(\frac{t}{b}\right)^c$ kde $a, b, c > 0$ jsou parametry a t je čas (v hodinách).

1. Odvoďte $R(t)$ a $f(t)$.
2. Vypočítejte, jaká je pravděpodobnost poruchy v období od $t = 10$ hod do $t = 11$ hod, hodnotu vyčíslete pro $a = 10^{-6}$, $b = 1000$ a $c = 3$.

Řešení:

$$R(t) = e^{-\int_0^t \lambda(\tau) d\tau} = e^{-\int_0^t a\tau^2 + \frac{\tau}{b}^c d\tau} = e^{-\left(\frac{at^3}{3} + \frac{1}{b^c} \frac{t^{c+1}}{c+1}\right)}$$

$f(t)$ spočítáme jako:

$$f(t) = \frac{-\partial R(t)}{\partial t} = e^{-\left(\frac{at^3}{3} + \frac{1}{b^c} \frac{t^{c+1}}{c+1}\right)} \left(\frac{1}{at^2 + \left(\frac{t}{b}\right)^c} \right)$$

Pravděpodobnost poruchy v daném časovém intervalu lze spočítat jako

$$Q(11) - Q(10) = (1 - R(11)) - (1 - R(10)) = R(10) - R(11)$$

nebo

$$Q(11) - Q(10) = \int_{10}^{11} f(t) dt$$

Po dosazení:

$$R(10) - R(11) = 0.99966 - 0.99955 = 0.011\%$$