

Statistika a spolehlivost v lékařství

Charakteristiky spolehlivosti prvků I

- **Příklad 1** Tahová síla papíru používaného pro výrobu potravinových sáčků je důležitá charakteristika kvality. Je známo, že síla (označme x) podléhá normálnímu rozdělení 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 rozdělení 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 nelze vypočítat analyticky, nicméně můžeme provést substituci $z = \frac{x-\mu}{\sigma}$, čímž dostaneme funkci, která není závislá na μ a σ^2 :

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

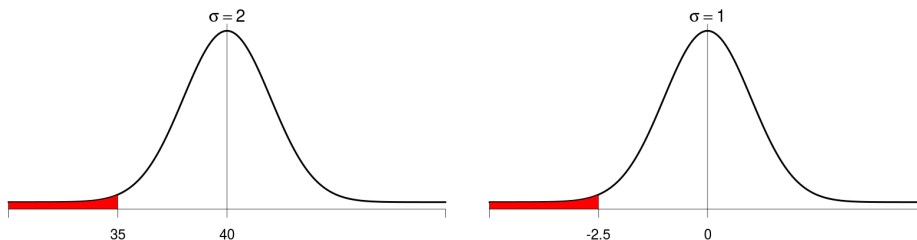
kde $\Phi(\cdot)$ je normální rozdělení $N(0, 1)$. My chceme vyjádřit $P(x \geq 35) = 1 - P(x \leq 35)$.

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

Požadovaná pravděpodobnost je tedy

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

Pozn: vyčíslení funkce $\Phi(\cdot)$ bývá součástí statistických tabulek.



Obrázek 1: Standardizace normálního rozdělení.

1 Charakteristiky spolehlivosti

Předpokládejme, že poruchy nastávají náhodně v čase. Čas poruchy je $\xi \geq 0$. Pravděpodobnost, že porucha, resp. čas poruchy, ξ (spojitá nezávislá veličina) nastane v čase t označme $Q(t)$.

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 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:

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

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

Víme, že integrál z podílu derivace funkce a této funkce je logaritmus dané funkce. Tedy:

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

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:

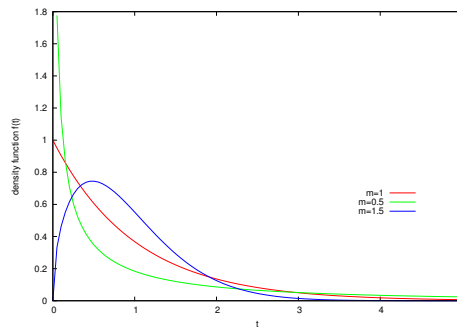
$$\begin{aligned} T_s &= \int_0^\infty t f(t) dt = \int_0^\infty t \left(-\frac{\partial R(t)}{\partial t}\right) dt = \int_0^\infty t(-\dot{R}(t)) dt = \\ &= \left| \begin{array}{l} u' = \dot{R}(t) \quad u = R(t) \\ v = t \quad v' = 1 \end{array} \right| = -[R(t)t]_0^\infty + \int_0^\infty R(t) dt =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

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

2 Exponenciální rozdělení

Základní charakteristiky:

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



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

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

Řešení:

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

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

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

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

- **Příklad 6** Předpokládejme nyní stroj, který se porouchá 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

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

dostáváme

$$-\lambda T_\beta = \ln \beta \tag{1}$$

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 (1). 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 jednoroční záruku. Pokud je kalkulačka v záruce vadná, je zákazníkovi vyměněna za novou. Spolehlivost kalkulačky je popsána následující hustotou pravděpodobnosti:

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

kde $x \geq 0$ je čas poruchy (roky).

- (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) Hustota pravděpodobnosti má tvar $f(t) = \lambda e^{-\lambda t}$. 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 $AFR = Q(1 \text{ rok})$.

- **Příklad 11** Máme naměřená data (časy poruch t_i). Odhadněte parametr λ pro exponenciální rozdělení metodou maximální věrohodnosti.

Řešení:

Nechť t_1, \dots, t_n je n měření časů, ve kterých došlo k poruše prvku. Předpokládáme, že časy poruch jsou náhodné, nezávislé a pocházejí ze stejného rozdělení hustoty pravděpodobnosti. Tuto hustotu pravděpodobnosti chceme modelovat exponenciálním rozdělením s hustotou pravděpodobnosti

$$f(t) = \lambda e^{-\lambda t}.$$

Věrohodnostní funkce (obecně) je

$$L(\theta|X_1, X_2, \dots, X_n) = f(t_1)f(t_2) \dots f(t_n),$$

kde θ jsou parametry hustoty pravěpodobnosti a X_i jsou měření. Exponenciálního rozdělení má jeden parametr (λ) a hodnoty $X_i = t_i$ jsou naše časy poruch.

$$L(\lambda|t_1, \dots, t_n) = \lambda e^{\lambda t_1} \dots \lambda e^{-\lambda t_n} = \lambda^n e^{-\lambda \sum_{i=1}^n t_i}.$$

Hledáme takové λ , které maximalizuje věrohodnostní funkci L , tj.

$$\hat{\lambda} = \arg \min_{\lambda \in \mathbb{R}} L(\lambda|t_1, \dots, t_n).$$

Někdy je pro pro jednodušší výpočet možné hledat maximum logaritmu věrohodnostní funkce, tj.

$$\log L = \log (f(t_1)f(t_2) \dots f(t_n)) = \sum_{i=1}^n \log(f(t_i)).$$

Pro případ exponenciálního rozdělení:

$$\log L = \log \left(\lambda^n e^{-\lambda \sum_{i=1}^n t_i} \right) = n \log \lambda - \lambda \sum_{i=1}^n t_i.$$

Hledáme takové $\hat{\lambda}$, které maximalizuje $\log L$

$$\frac{\partial \log L}{\partial \lambda} = 0$$

$$\frac{\partial \log L}{\partial \lambda} = n \frac{1}{\lambda} - \sum_{i=1}^n t_i \stackrel{!}{=} 0$$

$$\frac{n}{\lambda} = \sum_{i=1}^n t_i$$

a odtud

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n T_i}$$

Střední hodnota exponenciálního rozdělení je $T_s = \frac{1}{\lambda}$, tedy

$$T_s = \frac{\sum_{i=1}^n t_i}{n},$$

což lze interpretovat jako střední hodnota naměřených časů poruch. Jednoduchý výpočet (odhad) parametru λ z naměřených dat je jedna z výhod exponenciálního rozdělení.

- **Příklad 12** Pravděpodobnost bezporuchového provozu je dána $R(t) = 2e^{-\lambda t} - e^{-2\lambda t}$. Určete MTBF a hustotu pravděpodobnosti. Použijte metodu maximální věrohodnosti pro určení parametru λ .

Řešení:

$$\begin{aligned} \text{MTBF} = T_s &= \int_0^{\infty} t f(t) dt = \int_0^{\infty} R(t) dt = \int_0^{\infty} (2e^{-\lambda t} - e^{-2\lambda t}) dt = \\ &= \int_0^{\infty} 2e^{-\lambda t} dt - \int_0^{\infty} e^{-2\lambda t} dt = \frac{2}{\lambda} - \frac{1}{2\lambda} = \frac{3}{2\lambda}. \\ f(t) &= -\frac{\partial R(t)}{\partial t} = 2\lambda e^{-\lambda t} - 2\lambda e^{-2\lambda t} \end{aligned}$$

