ASYMPTOTICKÉ ROZVOJE RIZIKOVÝCH FUNKCÍ ODHADŮ SPOLEHLIVOSTI

Aleš Linka
KMA VŠST Liberec

1 Úvod

Uvažujme neporouchaný prvek, který začne pracovat v okamžiku $t = 0$ za určitých podmínek, o nichž budeme zatím předpokládat, že se v průběhu času nemění. V okamžiku $t = X$ se prvek porouchá. Doba X, po kterou prvek pracoval bez poruchy se nazývá doba do poruchy. Dále budeme předpokládat, že doba do poruchy X je s pravděpodobností 1 nezáporná náhodná veličina s distribuční funkcí

$$F(x) = P(X < x) \tag{1}$$

Ve většině praktických případů je $F(x)$ absolutně spojitá. Je-li tomu tak, nazývá se odpovídající hustota

$$f(x) = \frac{dF(x)}{dx} \tag{2}$$

hustotou poruch. S distribuční funkcí doby do poruchy je spojena funkce

$$R(x) = P(X \geq x) = 1 - F(x), \tag{3}$$

která se nazývá funkcí spolehlivosti, krátce spolehlivost.

Uvažujme třídu exponeciálních rozdělení s hustotou

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}\exp(\frac{x}{\theta}), & \text{pro } x > 0, \\ 0 & \text{jinak,} \end{cases} \tag{4}$$

kde θ neznámý parametr. Nechť c je kladné číslo. Jestliže X je náhodná veličina s hustotou (1), pak odpovídající funkce spolehlivosti (3) má tvar

$$R(c) = P(X > c) = \exp(\frac{x}{\theta}) \tag{5}$$

Bez újmy na obecnosti můžeme položit $c = 1$. Tedy (5) přejde do tvaru

$$R = R(1) = \exp(\frac{1}{\theta}). \tag{6}$$

Označme $L(\theta, d)$ ztrátovou funkci. Rizikovou funkci $R^*(\theta, \hat{R})$ příslušnou odhadu $\hat{R}$, je-li skutečná hodnota parametru θ, definujeme vztahem

$$R^*(\theta, \hat{R}) = E(L(\theta, \hat{R}) \mid \theta) \tag{7}$$

Při bayesovském přístupu předpokládáme, že paramatr θ je náhodná veličina s hustotou $g(\theta)$ a definujeme bayesovskou rizikovou funkci odhadu $\hat{R}$ za apriorní hustoty $g(\theta)$ parametru θ jako

$$\varrho(\hat{R}, g) = E_g R^*(\theta, \hat{R}) \tag{8}$$

V dalším textu budeme předpokládat, že ztrátová funkce má tvar

$$L(\theta, d)(\theta - d)^2. \tag{9}$$

V okamžiku $t = 0$ začneme za účelem zjištění charakteristik spolehlivosti pracovat n prvků stejného typu, a že jejich doba do poruchy je náhodná veličina s hustotou (4). Klasická situace nastane, jestliže pozorování budeme provádět, tak dlouho, dokud se všechny neporouchají. Výsledkem experimentu je úplný výběr $X_1, \ldots, X_n$ dob do poruchy.

Na základě toho výběru provedeme odhad spolehlivosti (6). Budeme studovat čtyři různé odhady. Nejlepší nestranný odhad můžeme psát ve tvaru

$$R_1 = \begin{cases} (1 - \frac{1}{nT_n})^{n-1}, & \text{jestliže } T_n > \frac{1}{n}, \\ 0 & \text{jinak,} \end{cases} \tag{10}$$

kde $T_n = \sum_{i=1}^n X_i$ (viz Pugh [6]). Užijeme-li dobře známého faktu, že T_n je maximálně věrohodný odhad θ. Pak pro maximálně věrohodný odhad (6) dostáváme

$$R_2 = \exp(\frac{1}{T_n}) \tag{11}$$

V bayesovském přístupu k odhadu předpokládáme, že $\delta = \theta^{-1}$ náhodná veličina s hustotou

$$g(\delta) = \begin{cases} \frac{\alpha^p \delta^{p-1}}{\Gamma(p)} \exp(-\alpha\delta) & \text{jestliže } \delta > 0, \\ 0 & \text{jinak.} \end{cases} \tag{12}$$

Bayesovský odhad, pak dostaneme jako střední hodnotu $\exp(-\delta)$ vzhledem k aposteriornímu rozdělení. Tedy

$$R_3 = (\frac{nT_n + \alpha}{nT_n + \alpha + 1})^{n+p} \tag{13}$$

Konečně uvažujme tzv. naivní odhad, který je určen četností výskytu jevu $\{X_i > 1\}$. Definujme $Z_1, \ldots, Z_n$ tak, že

$$Z_i = \begin{cases} 1 & \text{jestliže } X_i > 1, \\ 0 & \text{jinak.} \end{cases} \tag{14}$$

Pak naivní odhad můžeme psát ve tvaru

$$R_4 = \sum_{j=1}^{n} Z_i \tag{15}$$

2 Asymptotické vlastnosti odhadů

Nyní uvedeme některé vlastnosti uvedných odhadů spolehlivosti (viz např. [1]). Především odhady (10),(11),(12) a (13) jsou asymptoticky normální. Položme $\beta = \theta^{-1}$.

Věta 2.1 *Odhady R_1 a R_3 splňují tyto asymptotické rozvoje:*

$$R_1 = \exp(-\frac{1}{\bar{y}} + \frac{1}{n\bar{y}} - \frac{1}{2n\bar{y}^2})(1 + O(n^{-2}), \tag{16}$$

$$R_3 = \exp(-\frac{1}{\bar{y}} - \frac{p}{n\bar{y}} + \frac{(2\alpha+1)}{2n\bar{y}^2})(1 + O(n^{-2}). \tag{17}$$

Pro R_1, R_2 a R_3 platí

$$\sqrt{n}[R_i - \exp(-\beta)] \xrightarrow{D} N(0, \beta^2 \exp(-2\beta)). \tag{18}$$

Pro R_4 platí

$$\sqrt{n}[R_4 - \exp(-\beta)] \xrightarrow{D} N(0, \exp(-\beta)(1 - \exp(-\beta)). \tag{19}$$

Z věty plyne, že odhady R_1, R_2 a R_3 jsou asymptoticky ekvivalentní. Dále z věty 2.1 vyplývá, že R_2, R_3 jsou slabě asymptoticky eficientní, tzn. rozptyl jejich asymptotického rozdělení je stejný jako rozptyl asymptotického rozdělení nejlepšího nestranného odhadu R_1. Odhad R_4 není slabě asymptoticky eficientní, protože

$$\frac{\beta^2 \exp(-2\beta)}{\exp(-\beta)(1 - \exp(-\beta))} < 1, \tag{20}$$

když $\beta > 0$. Dále uvedeme pro odhady R_i asymptotické rozvoje pro ER_i, $varR_i$ a $R^*(R, R_i)$.

3 Riziková a bayesovská riziková funkce

Problematice asymptotických rozvojů číselných charakteristik odhadů spolehlivosti v exponenciálním rozdělení věnována práce [3], ale především práce [4] . Zde je uvedena i následující věta.

Věta 3.1 *Pro R_1, R_2 a R_3 platí*

$$ER_1 = \exp(-\beta), \tag{21}$$

$$ER_2 = \exp(-\beta)[1 + \frac{\beta}{2n}(\beta - 2)] + O(n^{-2}), \tag{22}$$

$$ER_3 = \exp(-\beta)[1 + \frac{\beta}{n}(\beta + \beta\alpha - p - 1)] + O(n^{-2}), \tag{23}$$

$$varR_1 = \frac{\beta^2}{n}\exp(-2\beta)[1 + \frac{1}{2n}(\beta - 2)^2] + O(n^{-\frac{5}{2}}), \tag{24}$$

$$varR_2 = \frac{\beta^2}{n}\exp(-2\beta)\{1 + \frac{1}{2n}[3(\beta - 2)^2 - 4]\} + O(n^{-\frac{5}{2}}), \tag{25}$$

$$varR_3 = \frac{\beta^2}{n}\exp(-2\beta)\{1 + \frac{1}{2n}[5(\beta-2)^2 - 4(p-3) - 4\beta(p-1) + 4\beta\alpha(\beta-2)]\} + O(n^{-\frac{5}{2}}), \tag{26}$$

$$R^*(R, R_1) = E(R_1 - \exp(-\beta))^2 = varR_1, \tag{27}$$

$$R^*(R, R_2) = E(R_2 - \exp(-\beta))^2 = \frac{\beta^2}{n}\exp(-2\beta)\{1 + \frac{1}{2n}[\frac{7}{2}(\beta - 2)^2 - 4]\} + O(n^{-\frac{5}{2}}), \tag{28}$$

$$R^*(R, R_3) = E(R_3 - \exp(-\beta))^2 = \frac{\beta^2}{n}\exp(-2\beta)\{1 + \frac{1}{2n}[5(\beta-2)^2 + 4(p-3) - 4\beta(p-1) + \tag{29}$$

$$+ 4\beta\alpha(\beta - 2) + 2(\beta + \beta\alpha - p - 1)^2]\} + O(n^{-\frac{5}{2}}). \tag{30}$$

Z daných rozvojů můžeme vidět, že všechny odhady jsou silně asymptoticky eficientní. Pro detailní porovnání R_i můžeme užít deficiency (viz [5]). Zhruba řečeno deficience znamená o kolik více (nebo méně) pozorování vyžaduje odhad B, má-li mít stejnou čtvercovou odchylku jako odhad A založený na výběru rozsahu n. V praxi se obvykle užívá asymptotická deficience pro $n \to \infty$. Jestliže označíme $V_n(A), V_n(B)$ střední čtvercové odchylky odhadů a platí

$$V_n(A) = \frac{a}{n^r} + \frac{b}{n^{(r+1)}} + o(n^{-(r+1)}), \tag{31}$$

$$V_n(B) = \frac{a}{n^r} + \frac{c}{n^{(r+1)}} + o(n^{-(r+1)}), \tag{32}$$

pak asymptotická deficience B vzhledem k A je rovna

$$D(B,A) = \frac{c-b}{ar} \tag{33}$$

Věta 3.2 *Pro deficienci odhadů R_1, R_2 a R_3 platí*

$$D(R_2,R_1) = \frac{5}{4}(\beta-2)^2 - 2, \tag{34}$$

$$D(R_3,R_1) = 4(\beta-2)^2 + 2(p-3) - 2\beta(p-1) + 2\beta\alpha(\beta-2) + (\beta+\alpha\beta-p-1)^2, \tag{35}$$

$$D(R_3,R_2) = \frac{4}{3}(\beta-2)^2 + 2(p-2) - 2\beta(p-1) + 2\beta\alpha(\beta-2) + (\beta+\alpha\beta-p-1)^2. \tag{36}$$

Z věty 3.2 plyne, že odhad R_2 je lepší než R_1 pouze pro hodnoty β blízké 2. Obdobné vysledky bychom chtěli získat na základě bayesovské rizikové funkce.

Dolní odhad pro bayesovské riziko je popsán v práci [2]. Nalézt asymptotické rozvoje pro $\varrho(R,R_i)$, však vyžaduje metodiku odlišnou od metodiky užité v případě porovnání odhadů na základě klasického přístupu, tj. na základě střední čtvercové odchylky. Použití numerických metod však ukázalo , že pro bayesovkou rizikovou funkci odhadů R_1, R_2 a R_3 platí

$$\varrho(R_i,g) = \frac{a(\alpha,p)}{n} + \frac{b_i(\alpha,p)}{n^2} + O(n^{-3}) \tag{37}$$

pro $n \to \infty$. Pomocí systému Mathematica byly vypočteny hodnoty $\varrho(R_i,g)(n)$ pro různé α, p a n= 1,...,35 . Získané výsledky byly použity pro odhad parametrů regresního modelu

$$\frac{1}{\varrho(R_i,g)(n)} = c_i + d_i n, \tag{38}$$

pro i 1,2,3 a $n = 2,\dots,35$. Tento model vedl ke stanovení asymptotických rozvojů

$$\varrho(R_i,g) = \frac{1}{d_i n} - \frac{c_i}{d_i^2 n} + O(n^{-3}) \tag{39}$$

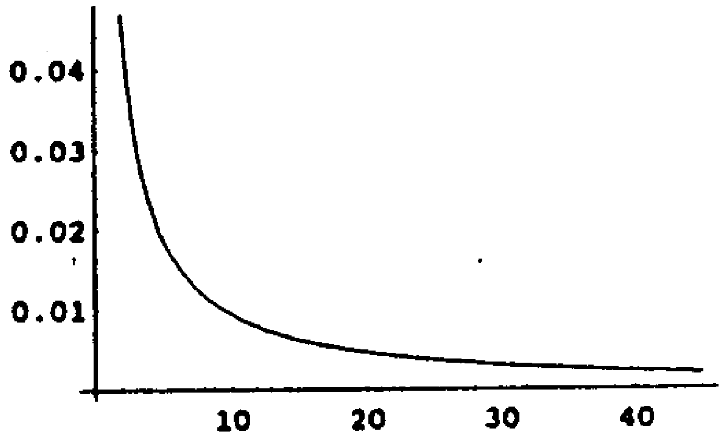
Výsledné rozvoje pro $\alpha = 2$ a $p = 3$ jsou uvedeny v příloze. V grafech jsou vyneseny též hodnoty $\varrho(R_i,g)(n)$.

Příloha.

Obrázek č.1

Asyptotický rozvoj pro bayesovskou rizikovou funkci maximálně věrohodného odhadu .

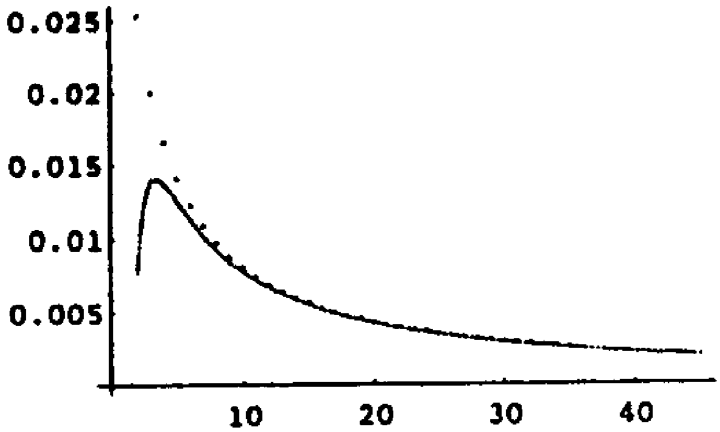
$$\frac{0.0942933}{n} - \frac{0.000525696}{n^{2}} + O\left[\frac{1}{n}\right]^{3}$$



Obrázek č.2

Asyptotický rozvoj pro bayesovskou rizikovou funkci bayesovského odhadu.

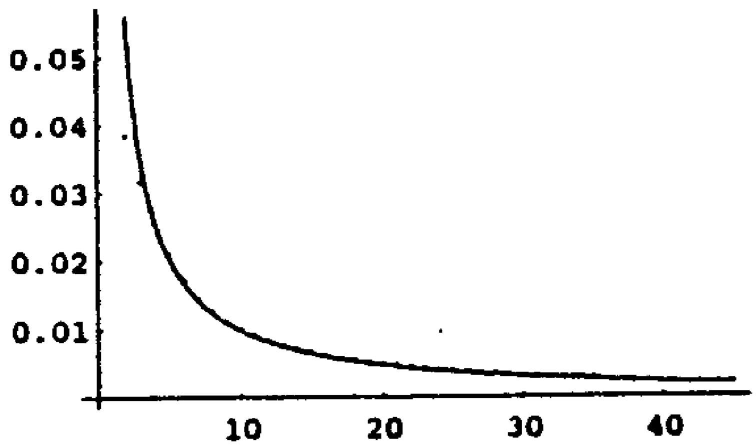
$$\frac{0.0938675}{n} - \frac{0.156559}{n^{2}} + O\left[\frac{1}{n}\right]^{3}$$



Obrázek č.3

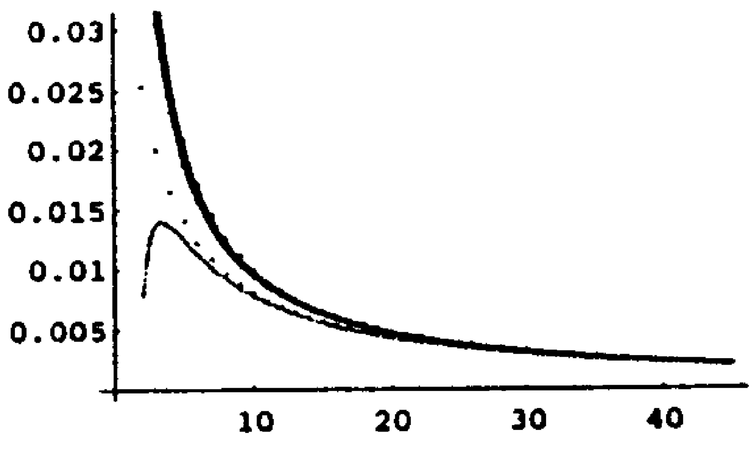
Asyptotický rozvoj pro bayesovskou rizikovou funkci nejlepšího nestranného odhadu.

$$\frac{0.0947147624624865 4}{n} + \frac{0.0348514}{n^2} + O\left[\frac{1}{n}\right]^3$$



Obrázek č.4

Porovnání asymptotických rozvojů.



Literatura

[1] *Harris,B.,Soms,A.*(1988). Sensitivity and asymtotic properties of Bayesian reliability estimates. Statistics Decisions, 33-47.

[2] *Brown,L.D.,Gajek,L.*(1990).Information inequalities for the bayes risk. Annals of Statit., 1578-1594 .

[3] *Chao,A.*(1981). Approximate Mean Squard Error of Estimators of Reliaability for k-out-of-m Systems in the Independent Exponencial Case. JASA 375, 720-724 .

[4] *Hurt,J.*(1976): On estimation of reliability in exponencial case. Aplikace matematiky 21, 263-272 .

[5] *Hodges,J.L.,Lehmannn,E.L.*(1970).Deficiency.Ann.Math.Statist.41,783-801.

[6] *Pugh,E.,L.*(1963). The best estimate of reliability in exponencial case. Oper.Res.11/17, 57-61.