12. METODY PRO KONSTRUKCI BODOVÝCH ODHADŮ NEZNÁMÝCH PARAMETRŮ ZNÁMÝCH ROZDĚLENÍ PRAVDĚPODOBNOSTI

Pro odhad hodnot parametrů pravděpodobnostních rozdělení se nejčastěji používá **metoda maximální věrohodnosti** (maximum likelihood), nebo **metoda momentů**.

12.1. Metoda momentů

Čas ke studiu: 25 minut

Cíl Po prostudování tohoto odstavce budete umět

- odhadovat parametry rozdělení pravděpodobnosti metodou momentů

Výklad

- **V čem spočívá princip metody momentů**

Metoda momentů je principielně jednoduchá metoda pro konstrukci bodových odhadů neznámých parametrů známých rozdělení, která spočívá v tom, že porovnáváme výběrové momenty získaných dat s odpovídajícími teoretickými momenty předpokládaného rozdělení s hustotou *f*(*t*).

Máme-li k dispozici zaznamenaná data (náhodný výběr) ($t_1, \dots, t_n$)´; pak *k*-tý výběrový obecný moment je dán vztahem $M_k = \frac{1}{n}\sum_{i=1}^{n} t_1^k$

(12.1)

Podobně *k* -tý výběrový centrální moment je dán

$$M_k = \frac{1}{n}\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^k \qquad (12.2)$$

kde $\bar{t}$ je výběrový průměr.

Odpovídající teoretické momenty jsou dány rovnicemi

$$\mu_k' = \int_0^\infty t^k f(t)dt$$

resp.

$$\mu_k = \int_0^\infty \left(t - \mu_1'\right)^k f(t)dt \qquad (12.3)$$

Jestliže rozdělení s hustotou $f(t)$ má r neznámých parametrů a jestliže soustava rovnic

$$M'_k = \mu_k', \; k = 1, \ldots, r$$

resp.

$$M_k = \mu_k, \; k = 1, \ldots, r$$

má jediné řešení, pak dává metoda momentů jednoznačně určené odhady r parametrů.

Řešený příklad 1

Uvažujme exponenciální rozdělení $f(t) = \lambda.e^{-\lambda t}$ *s neznámým parametrem* λ. *Máme najít odhad neznámého parametru* $\tilde{\lambda}$ *metodou momentů, je-li dán náhodný výběr* $(t_1, \ldots, t_n)'$.

$$\mu_1' = \int_0^\infty t\lambda.e^{-\lambda t}dt = \frac{1}{\lambda} \qquad a \qquad M_1' = \frac{\sum_{i=1}^n t_i}{n}.$$

Rovnice $\mu_1' = M_1'$ *přechází na rovnici*

$$\frac{1}{\lambda} = \frac{\sum_{i=1}^n t_i}{n} \qquad neboli \qquad \tilde{\lambda} = \frac{n}{\sum_{i=1}^n t_i} \qquad (12.4)$$

což je odhad neznámého parametru λ *získaný metodou momentů.*

Řešený příklad 2

Uvažujme nyní Weibullovo rozdělení $f(t) = \frac{\alpha}{\beta} t^{\alpha-1}.e^{-\frac{t^\alpha}{\beta}}$ *se dvěma neznámými parametry* α, β. *Pak porovnáním centrálních momentů dostáváme soustavu dvou rovnic pro neznámé* α, β:

$$\begin{aligned} \beta^{\frac{1}{2}}\Gamma\left(\frac{1+\alpha}{\alpha}\right) &= \frac{1}{n}\sum_{i=1}^n t_i \\ \beta^{\frac{2}{\alpha}}\Gamma\left(\frac{2+\alpha}{1+\alpha}\right) &= \frac{1}{n}\sum_{i=1}^n t_i^2 \end{aligned} \qquad (12.5)$$

kde $\Gamma(x)$ *je gama funkce.*

Soustava (12.5) je ovšem řešitelná pouze numericky volbou vhodného iteračního procesu.

Shrnutí pojmů

Metoda momentů je principielně jednoduchá metoda pro konstrukci odhadů neznámých parametrů známých rozdělení, která spočívá v tom, že porovnáváme výběrové momenty získaných dat s odpovídajícími teoretickými momenty předpokládaného rozdělení s hustotou $f(t)$. Metoda vede na řešení soustavy takového počtu rovnic, kolik je neznámých parametrů.

Úlohy k řešení 12.1.

1. Nechť turbína elektrárny podléhá náhodným šokům, které splňují předpoklady Poissonových pokusů. Nechť při každém pátém šoku dojde k závažné poruše turbíny. Během dlouhodobého sledování byly zaznamenány následující doby do poruch turbíny (v hodinách): (1020, 1100, 960, 1500, 1450, 1320, 1255, 1165, 1385, 1410). Určete pravděpodobnostní rozdělení pro dobu do poruchy turbíny. Určete dále

- odhad neznámého parametru zjištěného rozdělení metodou momentů
- hazardní funkci turbíny
- ve které fázi svého životního cyklu se turbína nachází

12.2. Metoda maximální věrohodnosti

Čas ke studiu: 25 minut

Cíl Po prostudování tohoto odstavce budete umět

- odhadovat parametry rozdělení pravděpodobnosti metodou maximální věrohodnosti

Výklad

- **Na čem je založena metoda maximální věrohodnosti**

Odhady získané touto metodou se všeobecně vyznačují dobrými statistickými vlastnostmi.

Nechť $t_1, t_2, \ldots, t_n$ je náhodný výběr z rozdělení s hustotou $f(t;\Theta)$, kde Θ je neznámý parametr. Naším problémem bude nalézt funkci (zvanou funkce věrohodnosti) danou

$$L(t_1,...,t_n;\Theta) = f(t_1;\Theta).f(t_2;\Theta)...f(t_n;\Theta) \qquad (12.6)$$

a z ní pak získat $\hat{\Theta}$ tak, aby $\hat{\Theta} = (t_1,\dots,t_n)$ bylo co nejlepším odhadem pro Θ. Pravá strana rovnice (12.6) je sdružená hustota pravděpodobnosti n-nezávislých proměnných ($t_1, \dots, t_n$) se stejným rozdělením.

Ve skutečnosti $L(t_1,\dots,t_n;\Theta)$ může být uvažována jako apriorní pravděpodobnost pro získání hodnot $t_1, \dots, t_n$.
Jelikož L je jednoduše funkcí neznámého parametru Θ, který je odhadován, metoda maximální věrohodnosti je založena na získání takové hodnoty Θ, která maximalizuje L. Při praktických výpočtech bývá však výhodnější maximalizovat spíše funkci ln L namísto L, jelikož obě tyto operace jsou ekvivalentní a dávají stejné výsledky. Podmínkou optimality je tedy rovnice

$$\frac{\partial \ln L(t_1,...,t_n;\Theta)}{\partial \Theta} = 0 \qquad (12.7)$$

a hodnota z této podmínky získaná je funkcí náhodného výběru, $\hat{\Theta} = \hat{\Theta}(t_1,\dots,t_n)$.

Řešený příklad 1

Mějme exponenciální rozdělení s hustotou $f(t) = \lambda.e^{-\lambda t}$. *Funkce věrohodnosti pak bude dána výrazem*

$$L(t_1,\dots,t_n;\lambda) = (\lambda.e^{-\lambda t_1}).(\lambda.e^{-\lambda t_2})...(\lambda.e^{-\lambda t_n}) = \lambda^n \exp\left(-\lambda\sum_{i=1}^{n} t_i\right) \qquad (12.8)$$

a logaritmováním získáme

$$\ln L(t_1,...,t_n;\lambda) = n.\ln\lambda - \lambda.\sum_{i=1}^{n} t_i \qquad (12.9)$$

Rovnice (12.7) pak dává vztah pro maximálně věrohodný odhad parametru $\hat{\lambda}$:

$$\frac{\partial \ln L(t_1,...,t_n;\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} t_i = 0$$

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i} \qquad (12.10)$$

Řešený příklad 2

Uvažujme dvouparametrické Weibullovo rozdělení s hustotou $f(t) = \frac{\alpha}{\beta} t^{\alpha-1} \exp\left(-\frac{t^\alpha}{\beta}\right)$

Funkce věrohodnosti L je dána

$$L(t_1,\dots,t_n;\alpha,\beta)=\frac{\alpha}{\beta}t_1^{\alpha-1}\exp\left(-\frac{t_1^{\alpha}}{\beta}\right)\dots\frac{\alpha}{\beta}t_n^{\alpha-1}\exp\left(-\frac{t_n^{\alpha}}{\beta}\right)=\left(\frac{\alpha}{\beta}\right)^n\prod_{j=1}^{n}t_j^{\alpha-1}\exp\left(-\frac{1}{\beta}\sum_{j=1}^{n}t_j^{\alpha}\right) \tag{2.11}$$

A logaritmováním získáme

$$\ln L=n\ln\alpha-n\ln\beta+(\alpha-1)\sum_{j=1}^{n}\ln t_j-\frac{1}{\beta}\sum_{j=1}^{n}t_j^{\alpha} \tag{12.12}$$

Optimalizaci však provádíme s ohledem na oba neznámé parametry α,β*, takže rovnice (12.7) přechází v tomto případě na dvě následující rovnice:*

$$\frac{\partial\ln L}{\partial\alpha}=\frac{n}{\alpha}+\sum_{j=1}^{n}\ln t_j-\frac{1}{\beta}\sum_{j=1}^{n}t_j^{\alpha}\ln t_j=0$$

$$\frac{\partial\ln L}{\partial\beta}=-\frac{n}{\beta}+\frac{1}{\beta^2}\sum_{j=1}^{n}t_j^{\alpha}=0 \tag{12.13}$$

Z druhé rovnice můžeme snadno získat

$$\beta=\frac{\sum_{j=1}^{n}t_j^{\alpha}}{n} \tag{12.14}$$

zatímco z první rovnice dostaneme

$$\beta=\frac{\sum_{j=1}^{n}t_j^{\alpha}\ln t_j}{\frac{n}{\alpha}+\sum_{j=1}^{n}\ln t_j} \tag{12.15}$$

Porovnáním pravých stran posledních dvou rovnic získáme jednu rovnici pro jednu neznámou α*. Řešení je nutno provést opět numericky volbou vhodného iteračního procesu.*

Σ Shrnutí pojmů

Metoda maximální věrohodnosti je principielně jednoduchá metoda pro konstrukci odhadů neznámých parametrů známých rozdělení pravděpodobnosti, která je založena na podmínce maximalizace **věrohodnostní funkce,** což je sdružená hustota pravděpodobnosti daného náhodného výběru, brána ovšem jako funkce neznámých parametrů**.**

Úlohy k řešení 12.2.

1. Doba do poruchy dieselgenerátoru se řídí exponenciálním rozdělením pravděpodobnosti. Během dlouhodobého sledování byly zaznamenány následující poruchové doby v hodinách: (150, 190, 165, 177, 203, 178, 162, 181, 194, 168). Odhadněte parametr λ metodou maximální věrohodnosti. Charakterizujte hazardní funkci dieselgenerátoru, odhadněte funkci bezporuchovosti v čase t=100 hodin. Určete 90% -tní život dieselgenerátoru.

Klíč k řešení úloh 12.1. a 12.2.:

Řešení úloh 12.1.

1. Nechť turbína elektrárny podléhá náhodným šokům, které splňují předpoklady Poissonových pokusů. Nechť při každém pátém šoku dojde k závažné poruše turbíny. Během dlouhodobého sledování byly zaznamenány následující doby do poruch turbíny (v hodinách): (1020, 1100, 960, 1500, 1450, 1320, 1255, 1165, 1385, 1410). Určete pravděpodobnostní rozdělení pro dobu do poruchy turbíny. Určete dále
   - odhad neznámého parametru zjištěného rozdělení metodou momentů
   - hazardní funkci turbíny
   - ve které fázi svého životního cyklu se turbína nachází

Řešení:

Doba do poruchy se řídí Gamma rozdělením s hustotou f(t) = $\dfrac{\lambda^5}{\Gamma(5)}.t^4.e^{-\lambda t}$, s neznámým parametrem λ, který odhadneme metodou momentů:

Rovnice $\mu_1' = M_1'$ přechází na rovnici

$$\frac{5}{\lambda} = \frac{\sum_{i=1}^{10} t_i}{10} \qquad \text{neboli} \qquad \tilde{\lambda} = \frac{50}{\sum_{i=1}^{10} t_i} \cong 0{,}004$$

což je odhad neznámého parametru λ získaný metodou momentů.

Hazardní funkce je: $$h(t) = \frac{0{,}004}{24\sum_{j=0}^{4}\frac{1}{(4-j)!\left(0{,}004t\right)^{j}}}$$

Turbína se nachází ve třetí fázi svého životního cyklu, tj. v období poruch v důsledku stárnutí a opotřebení.

Řešení úloh 12.2.

1. Doba do poruchy dieselgenerátoru se řídí exponenciálním rozdělením pravděpodobnosti. Během dlouhodobého sledování byly zaznamenány následující poruchové doby v hodinách: (150, 190, 165, 177, 203, 178, 162, 181, 194, 168). Odhadněte parametr λ metodou maximální věrohodnosti. Charakterizujte hazardní funkci dieselgenerátoru, odhadněte funkci bezporuchovosti v čase t=100 hodin. Určete 90% -tní život dieselgenerátoru.

Řešení:

Věrohodnostní funkce: $L\left(t_1,\ldots,t_n;\lambda\right) = \left(\lambda.e^{-\lambda t_1}\right).\left(\lambda.e^{-\lambda t_2}\right)\ldots\left(\lambda.e^{-\lambda t_n}\right) = \lambda^{n}\exp\left(-\lambda\sum_{i=1}^{n}t_i\right)$

dává odhad parametru λ následovně: $\hat{\lambda} = \dfrac{n}{\sum_{i=1}^{n}t_i}$, což po dosazení zadaných hodnot

je $\hat{\lambda} = \dfrac{10}{1768} \cong 0{,}005656$.
Hazardní funkce je konstantní h(t) = 0,005656, dieselgenerátor je tedy v období stabilního života.
Funkce bezporuchovosti v čase 100 hodin je:
$R(t) = 1 - F(t) = e^{-\lambda t} = e^{-0{,}5656} \cong 0{,}568$