Bodové odhady – některé další metody konstrukce odhadů

Metoda momentů

Konzistentní odhad[1]

Volná interpretace této vlastnosti říká, že se jedná o takový (konzistentní) odhad, u kterého dochází s počtem pozorování „k zpřesňování“. Tj. přidáním dalších pozorování odhad „příliš nekazíme“. Taková interpretce může demonstrovat jen smysl a důvod pro jasně určený pojem, proto:

Statistika $\hat{g}(x_1, x_2, ..., x_n) \cong \hat{g}_n$ je konzistentním odhadem parametru g (parametrické funkce) právě když platí: $\forall \varepsilon > 0;\ \lim_{n \to +\infty} P\left(\left|\hat{g}_n - g\right| < \varepsilon\right) = 1$ pro každou možnou hodnotu odhadovaného parametru $g \in \Omega$, kde Ω je množina (prostor) možných hodnot odhadovaného parametru.

Tvrzení:
K tomu aby byl odhad konzistentní (statistika byla konzistentní) stačí splnění obou následujících podmínek:

1. $\lim_{n\to+\infty} E\left\{\hat{g}_n\right\} = g \equiv$ asymptotická nestrannost
2. $\lim_{n\to+\infty} \sigma^2\left(\hat{g}_n\right) = 0$

Důkaz:

Podle Čebyševovy nerovnosti platí: $\forall n = 1,...;\ P\left(\left|\hat{g}_n - E\left\{\hat{g}_n\right\}\right| < \varepsilon\right) \ge 1 - \frac{\sigma^2\left(\hat{g}_n\right)}{\varepsilon^2}$. Dále přepsáním předpokladu $\lim_{n\to+\infty} E\left\{\hat{g}_n\right\} = g$, dostaneme $\forall\varepsilon > 0; \exists n_0; \forall n > n_0 \left|E\left\{\hat{g}_n\right\} - g\right| < \varepsilon$.

Potom: $P\left(\left|\hat{g}_n - g\right| < \varepsilon\right) = P\left(-\varepsilon < \hat{g}_n - g < \varepsilon\right) = P\left(-\varepsilon - E\left\{\hat{g}_n\right\} < \hat{g}_n - g - E\left\{\hat{g}_n\right\} < \varepsilon - E\left\{\hat{g}_n\right\}\right) =$

$$= P\left(g - \varepsilon - E\left\{\hat{g}_n\right\} < \hat{g}_n - E\left\{\hat{g}_n\right\} < g + \varepsilon - E\left\{\hat{g}_n\right\}\right) = P\left(-\varepsilon + g - E\left\{\hat{g}_n\right\} < \hat{g}_n - E\left\{\hat{g}_n\right\} < \varepsilon + g - E\left\{\hat{g}_n\right\}\right) \ge$$

$\ge P\left(-\varepsilon - \varepsilon < \hat{g}_n - E\left\{\hat{g}_n\right\} < \varepsilon + \varepsilon\right)$. To platí při využití takových n, která splňují podmínku $n > n_0$ z definice limity $\forall\varepsilon > 0; \exists n_0; \forall n > n_0 \left|E\left\{\hat{g}_n\right\} - g\right| < \varepsilon$. Pro pravděpodobnost $P\left(-2\varepsilon < \hat{g}_n - E\left\{\hat{g}_n\right\} < 2\varepsilon\right) = P\left(\left|\hat{g}_n - E\left\{\hat{g}_n\right\}\right| < 2\varepsilon\right)$ platí omezení dané výše uvedenou

---

[1] Porovnejte s pojmem konzistence z předcházející části.

Čebyševovou nerovností, tedy: $P\left(\left|\hat{g}_n - E\left\{\hat{g}_n\right\}\right| < 2\varepsilon\right) \geq 1 - \frac{\sigma^2\left(\hat{g}_n\right)}{4\varepsilon^2}$. Shrnuto: $\forall\varepsilon > 0; \exists n_0; \forall n > n_0$ je $P\left(\left|\hat{g}_n - g\right| < \varepsilon\right) \geq 1 - \frac{\sigma^2\left(\hat{g}_n\right)}{4\varepsilon^2}$. Protože platí: $\lim_{n\to+\infty} \sigma^2\left(\hat{g}_n\right) = 0$, platí $\lim_{n\to+\infty} P\left(\left|\hat{g}_n - g\right| < \varepsilon\right) \geq 1 \Rightarrow \lim_{n\to+\infty} P\left(\left|\hat{g}_n - g\right| < \varepsilon\right) = 1$ neboť pravděpodobnost nemůže být větší než jedna.

Vlastní metoda momentů

Označíme statistiku $m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$, která je nazývávána *k-tým* výběrovým obecným momentem.

Je evidentní, že: $E\{m_k\} = \frac{1}{n}\sum_{i=1}^{n} E\{x_i^k\} = \frac{1}{n}\sum_{i=1}^{n} \mu_k = \mu_k$. Tedy $m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$ je **nestranným** odhadem μ_k, kde μ_k je *k-tým* obecným momentem náhodné proměnné od které jsou pozorované hodnoty (náhodný výběr, iid) $x_i ; i = 1,...,n$.

Dále platí: $\sigma^2\left(x_i^k\right) = E\left\{\left(x_i^k\right)^2\right\} - \left(E\left\{x_i^k\right\}\right)^2 = E\left\{x_i^{2k}\right\} - \left(\mu_k\right)^2 = \mu_{2k} - \left(\mu_k\right)^2$. Dále platí $\sigma^2\left(\sum_{i=1}^{n} x_i^k\right) = \sum_{i=1}^{n}\sigma^2\left(x_i^k\right) = n\left(\mu_{2k} - \left(\mu_k\right)^2\right)$, proto $\sigma^2\left(\frac{1}{n}\sum_{i=1}^{n} x_i^k\right) = \frac{1}{n}\left(\mu_{2k} - \left(\mu_k\right)^2\right) = \sigma^2\left(m_k\right)$. Odtud za pomoci Čebyševovy nerovnosti $P\left(\left|m_k - \mu_k\right| < \varepsilon\right) \geq 1 - \frac{\sigma^2\left(m_k\right)}{\varepsilon^2} = 1 - \frac{\mu_{2k} - \left(\mu_k\right)^2}{n\varepsilon^2}$ dostaneme: $\lim_{n\to+\infty} P\left(\left|m_k - \mu_k\right| < \varepsilon\right) = 1$. Proto $m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$ je také **konzistentním** odhadem μ_k.

Pro náhodnou proměnnou, jejíž hustota závisí (funkčně) na parametrech $g_1,...,g_m$, platí $\mu_k = \mu_k\left(g_1,...,g_m\right)$, pro $k = 1,...,m$ pokud takové momenty existují.

Budeme-li řešit soustavu rovnic $m_k = \mu_k\left(g_1,...,g_m\right)$, vůči neznámým $g_1,...,g_m$ můžeme v některých případech využít vlastností nestrannosti a konzistence odhadů $m_k = \mu_k\left(g_1,...,g_m\right)$ k získání odhadů $\hat{g}_1,...,\hat{g}_m$ parametrů $g_1,...,g_m$.

Příklad: Pro náhodný výběr z normálního rozdělení s parametry μ, σ^2 platí: $\mu_1 = \mu; \mu_2 = \sigma^2 + \mu^2$. Statistické momentové rovnice pak budou vypadat:

$m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \mu; m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \sigma^2 + \mu^2$. Jejich řešením dostaneme: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ a $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$.

Námět: Srovnejte se statistikami $\bar{x}$, S_0^2, S_1^2, S^2 a ze srovnání určete vlastnosti odhadů $\hat{\mu}$ a $\hat{\sigma}^2$.

Příklad: Pro náhodný výběr z rozdělení $f_\xi(x) = \frac{1}{\tau} e^{-\frac{x}{\tau}}$ platí $\mu_1 = \tau$ a odtud $\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Námět: Určete, zda se jedná o nestranný a konzistentní odhad.

Metoda maximální věrohodnosti

Mějme (iid) náhodný výběr $\{x_1, x_2, ..., x_n\}$ rozsahu *n* náhodné proměnné ξ s distribuční funkcí $F(x;g)$ a hustotou $f(x;g)$, kde g je hledaný parametr(y); může se jednat i o vektor parametrů).

Označíme $V(x_1, x_2, ..., x_n; g) = \prod_{i=1}^{n} f(x_i; g)$. Za maximálně věrohodný odhad budeme považovat statistiku $\hat{g}$ pro kterou platí $V(x_1, x_2, ..., x_n; g) \le V\left(x_1, x_2, ..., x_n; \hat{g}\right)$, $\forall$ možná g.

Tedy statistika $\hat{g}$ maximalizuje věrohodnostní funkci $V(x_1, x_2, ..., x_n; g) = \prod_{i=1}^{n} f(x_i; g)$.

Z praktických důvodů se velice často používá logaritmická věrohodnostní funkce $L(x_1, x_2, ..., x_n; g) = lgV(x_1, x_2, ..., x_n; g) = \sum_{i=1}^{n} lg\, f(x_i; g)$. Vzhledem k tomu, že funkce *lg* je rostoucí je evidentní, že: $\left[V(x_1, x_2, ..., x_n; g) \le V\left(x_1, x_2, ..., x_n; \hat{g}\right)\right] \Leftrightarrow \left[L(x_1, x_2, ..., x_n; g) \le L\left(x_1, x_2, ..., x_n; \hat{g}\right)\right]$.

Filosofie tvorby takového odhadu $\hat{g}$ je založena na předpokladu, že to co se stalo, muselo být maximálně věrohodné (pravděpodobné; jedná se intuitivní předpoklad, nikoliv o matematický, proto musí být obecně vlastnosti odhadů následně ověřovány, stejně jako u předchozích metodik).

Příklad: Maximálně věrohodné odhady parametrů normálního rozdělení $f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$lg\left(f_\xi(x)\right) = -0{,}5 lg(2\pi) - lg\sigma - \frac{(x-\mu)^2}{2\sigma^2}$ a

$L(x_1, x_2, ..., x_n; \mu; \sigma^2) = -n\, lg(2\pi) - \frac{n}{2} lg\,\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$.

$$\frac{\partial}{\partial\mu}L(x_1,x_2,...,x_n;\mu;\sigma^2)=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0\Rightarrow\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i\,,$$

$$\frac{\partial}{\partial\sigma^2}L(x_1,x_2,...,x_n;\mu;\sigma^2)=-\frac{n}{2}\frac{1}{\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2=0\Rightarrow\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$, po dosazení

z první rovnice dostaneme $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\hat{\mu}\right)^2$. Matice druhých parciálních derivací:

$$\begin{pmatrix}\frac{\partial^2}{\partial\mu^2}L & \frac{\partial^2}{\partial\mu\partial\sigma^2}L\\ \frac{\partial^2}{\partial\mu\partial\sigma^2}L & \frac{\partial^2}{\partial(\sigma^2)^2}L\end{pmatrix}=\begin{pmatrix}\frac{-1}{\sigma^2} & \frac{-1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)\\ \frac{-1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu) & \frac{n}{2}\frac{1}{(\sigma^2)^2}-\frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}(x_i-\mu)^2\end{pmatrix}$$

Námět: Ověřte, že je tato matice negativně definitní.
Odtud plyne, že řešení normálních rovnic realizuje maximum logaritmické věrohodnostní funkce. Proto: $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i$ a $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\hat{\mu}\right)^2$ jsou maximálně věrohodnými odhady parametrů μ,σ^2.

Příklad: Maximálně věrohodný odhad parametru rovnoměrného rozdělení na intervalu $\langle 0,h\rangle$ (příklad toho, kdy je vhodnější nepoužít logaritmické věrohodnostní funkce).
$f_\xi(x)=\frac{1}{h}\Leftrightarrow x\in\langle 0,h\rangle;\, jinak\ f_\xi(x)=0$. Věrohodnostní funkce (vztažená k odhadovanému parametru h) je pak $V(x_1,x_2,...,x_n;h)=\prod_{i=1}^{n}f(x_i;h)=\frac{1}{h^n}$ a ta nabývá svého maxima pro statistiku pro $\hat{h}=x_{(n)}$, protože $h\geq x_i\forall i=1,...,n\Rightarrow\frac{1}{h}\leq\frac{1}{x_i}\forall i=1,...,n\Rightarrow\frac{1}{h^n}\leq\frac{1}{(x_i)^n}\forall i=1,...,n$.
Nerovnost $h\geq x_i$ musí být splněna i pro odhad h, proto $\hat{h}=x_{(n)}$.

Příklad: Maximálně věrohodný odhad parametru Poissonova rozdělení $p_i=e^{-\lambda}\frac{\lambda^i}{i!}$.
$lg\ p_i=-\lambda+i\,lg(\lambda)-lg(i!)$, $\{i_1,i_2,...,i_n\}$ bude náhodným výběrem rozsahu n z tohoto rozdělení. Logaritmická věrohodnostní funkce pak bude: $L(i_1,i_2,...i_n;\lambda)=\sum_{i=1}^{n}lg\,p_i=-n\lambda+lg(\lambda)\sum_{j=1}^{n}i_j-\sum_{j=1}^{n}lg(i_j!)$ a

$\frac{\partial}{\partial\lambda}L(i_1,i_2,...i_n;\lambda)=-n+\frac{1}{\lambda}\sum_{j=1}^{n}i_j;\ \frac{\partial^2}{\partial\lambda^2}L(i_1,i_2,...i_n;\lambda)=-\frac{1}{\lambda^2}\sum_{j=1}^{n}i_j$. Odtud plyne, že maximálně věrohodný odhad parametru λ dostaneme řešením rovnice $-n+\frac{1}{\lambda}\sum_{j=1}^{n}i_j=0\Leftrightarrow\hat{\lambda}=\frac{1}{n}\sum_{j=1}^{n}i_j$.

Vlastnosti maximálně věrohodných odhadů

Ve skriptech Machek J.: Teorie odhadu, SPN Praha 1974, str. 98-110, jsou dokázány následující vlastnosti maximálně věrohodných odhadů (nejen) pro určité skupiny rozdělení:

1. Existuje-li vydatný odhad, je jediným maximálně věrohodným odhadem.
2. Maximálně věrohodný odhad je konzistentním odhadem.
3. Existuje-li postačující statistika pro parametr g, pak je maximálně věrohodný odhad parametru g funkcí této postačující statistiky

Zavedení postačující statistiky (sufficient)

Definice: Statistika $\hat{g}(x_1, x_2, ..., x_n) \cong \hat{g}_n$ se nazývá postačující statistikou pro parametr(y) g rozdělení $f_\xi(x;g)$ jestliže podmíněné rozdělení $f_\xi(x/\hat{g}_n)$ nezávisí (funkčně) na parametru(ech) g. Pro diskrétní náhodné proměnné platí totéž pro pravděpodobnosti.

Filosofie: Postačující statistika tedy „obsahuje všechnu postačující „informaci“ obsaženou v g“ pro pravděpodobnostní popis náhodné veličiny ξ. Svým způsobem znalost postačující statistiky je náhradou za neznámý parametr g pro daná pozorování.

Příklad:
Pro náhodný výběr z alternativního rozdělení $P(x) = p^x(1-p)^{1-x}; x \in \{0,1\}$ je postačující statistikou pro odhad parametru p statistika $t(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} x_i$. Statistika $t(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} x_i$ má binomické rozdělení pravděpodobnosti $P(t=i) = \binom{n}{i} p^i (1-p)^{n-i}$.
Potom:

$$P(x_1, x_2, ..., x_n / t) = \frac{P(x_1, x_2, ..., x_n, t)}{P(t)} = \frac{P(x_1, x_2, ..., x_n)}{P(t)} = \frac{\prod_{j=1}^{n} p^{x_j}(1-p)^{1-x_j}}{\binom{n}{t} p^t (1-p)^{n-t}} = \frac{p^{\sum x_j}(1-p)^{n-\sum x_j}}{\binom{n}{t} p^t (1-p)^{n-t}} =$$

$$= \frac{p^t(1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} = \frac{1}{\binom{n}{t}}$$. Shrnuto $P(x_1, x_2, ..., x_n / t) = \frac{1}{\binom{n}{t}} = \frac{1}{\binom{n}{\sum_{i=1}^{n} x_i}}$ a je vidět, že takové podmíněné rozdělení nezávisí na parametru p.

Příklad:

Pro náhodný výběr z rovnoměrného rozdělení na intervalu $\langle 0,h \rangle$, $f_\xi(x)=\frac{1}{h} \Leftrightarrow x\in\langle 0,h\rangle; jinak\ \ f_\xi(x)=0$ je postačující statistikou $t(x_1,x_2,...,x_n)=x_{(n)}=max\{x_1,x_2,...,x_n\}$.

Statistika $x_{(n)}$ má, v takovém případě, rozdělení $f_{(n)}(x)=nf(x)F^{n-1}(x)=n\frac{1}{h}\left(\frac{x}{h}\right)^{n-1}$.

A podmíněná hustota $f\left(x_1,x_2,...,x_n/t\right)=\dfrac{f\left(x_1,x_2,...,x_n,t\right)}{f(t)}=\dfrac{f\left(x_1,x_2,...,x_n\right)}{n\frac{1}{h}\left(\frac{t}{h}\right)^{n-1}}=\dfrac{\frac{1}{h^n}}{n\frac{1}{h}\left(\frac{t}{h}\right)^{n-1}}=\dfrac{1}{nt^{n-1}}$.

Shrnuto: $f\left(x_1,x_2,...,x_n/t=x_{(n)}\right)=\dfrac{1}{nt^{n-1}}=\dfrac{1}{n\left(x_{(n)}\right)^{n-1}}$.

Poznámky

Z předchozí klasifikace bodových odhadů plynou některé další poznatky.

1. Nestranné odhady nemusí být vydatné, vydatné odhady nemusí být nestranné.
2. Kritérium minimalizaci rozptylu odhadu (vydatnost) nemusí být jedinou mírou kvality odhadu.
3. Postačující statistiky slouží ke generování odhadů, které mají společné dvě vlastnosti: nestrannost a vydatnost.
4. V matematické statistice dochází k dalším klasifikacím (např. úplné statistiky, „mediánová“ nestrannost, …) umožňujícím popisovat vlastnosti jednotlivých odhadů.
5. Pro generování odhadů jsou vhodné metody momentů a metoda maximální věrohodnosti. Existují i jiné metody. Převážná část odhadů má za sebou část intuitivní procesy a následné exaktní zkoumání vlastností takových odhadů.

Doporučená a zdrojová literatura:

Jiří Reif	Metody matematické statistiky, ZČU v Plzni 2004
Jaroslav Hátle, Jiří Likeš	Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974.
C. Radhakrishna Rao	Lineární metody statistické indukce a jejich aplikace, ACADEMIA, Praha 1978
Machek J.	Teorie odhadu, SPN Praha 1974, skripta MFFUK