Bodové odhady

Exponenciální skupina (typ, family) rozdělení

Rozdělení s hustotou $f(x)$ budeme nazývat exponenciálního typu vůči parametru(ům) g, pokud $f(x,g) = e^{K(x)Q(g)+R(g)+S(x)}$, kde $K(x)$, $S(x)$ jsou funkce nezávislé na parametru g a $Q(g)$, $R(g)$ jsou nezávislé na proměnné x. Zde je slovo nezávislost použito ve smyslu funkcionální nezávislosti.

Příklady:

1. Alternativní rozdělení $A(p) \cong f(x) = p^x(1-p)^{1-x} = e^{x\,lg(\frac{p}{1-p})+lg(1-p)}, x = 0{,}1$ je exponenciálního typu protože:
   $$K(x) = x; Q(p) = lg\left(\frac{p}{1-p}\right); R(p) = lg(1-p); S(x) = 0$$

2. Posunuté exponenciální rozdělení je exponenciálního typu „vzhledem k δ“, protože:
   $$K(x) = x - A; Q(\delta) = -\frac{1}{\delta}; R(\delta) = -lg(\delta); S(x) = 0 .$$

Cramér – Raova nerovnost

Věrohodnostní funkce za předpokladu iid: $f(x_1, x_{21}, ..., x_n, g) = \prod_{i=1}^{n} f(x_i, g)$, pak

$\log f(x_1, x_2, ..., x_n, g) = \sum_{i=1}^{n} \log f(x_i, g)$, je logaritmická věrohodnostní funkce vztažená k pozorováním $x_1, x_2, ..., x_n$ a parametru g.

Vzniká otázka jak přesně statistika $T(x_1, x_{21}, ..., x_n)$ odhaduje zjišťovanou hodnotu g. Další otázkou je jak přesně lze vůbec takovou zjišťovanou hodnotu odhadnout. To řeší následující metodika:

Nazveme V skóre náhodné proměnné X vůči zjišťovanému parametru g:

$$V = \frac{d}{dg} log\, f(x,g) = \frac{\frac{d}{dg} f(x,g)}{f(x,g)}$$ a

$$E\{V\} = \int_{Df} \frac{\frac{d}{dg} f(x,g)}{f(x,g)} f(x,g)dx = \frac{d}{dg} \int_{Df} f(x,g)dx = \frac{d}{dg} 1 = 0 \quad \textbf{(1)}$$

Proto: $E\{V^2\} = \sigma^2(V)$

Odtud: **Fischerova informace** $J(g) = E_g\left\{\left(\frac{d}{dg} log\, f(x,g)\right)^2\right\} = E_g\{V^2(x)\}$**.** Střední hodnota se počítá přes náhodnou proměnnou pozorování.

Skóre pro pozorování $V(x_1, x_2, ..., x_n, g) = \sum_{i=1}^{n} \frac{d}{dg} \log f(x_i, g) = \sum_{i=1}^{n} V(x_i, g)$. Jednotlivá pozorování jsou nezávislá, proto jsou i nezávislá jim odpovídající skóre:
$E_g\{V^2(x_1, x_2, ..., x_n, g)\} = \sum_{i=1}^{n} E_g\{V^2(x_i, g)\} = nJ(g)$. Nechť je dále statistika $T(x_1, x_2, ..., x_n)$ nestranným odhadem parametru g, tedy $E_{(x_1, x_2, ..., x_n)}\{T(x_1, x_2, ..., x_n)\} = g$. Potom platí:

$\sigma^2\{T(x_1, x_2, ..., x_n)\} \geq \frac{1}{nJ(g)}$ $\equiv$ Cramér – Raova nerovnost.

Fisherova míra informace tedy vypovídá o tom, jak přesně **lze odhadnout** zjišťovanou hodnotou z pozorování. Převrácená hodnota této míry je tedy dolní mezí pro rozptyl odhadu daného parametru, je ideální dolní mezí - nezávislou na tvaru statistiky = odhadovací funkce = estimátoru.

Důkaz: Nechť je $V(x)$ skóre, $T(x)$ estimátor parametru g (statistika). Potom:
$[E\{(V - E(V))(T - E(T))\}]^2 \leq [E\{(V - E(V))^2\} E\{(T - E(T))^2\}]$ (to je Cauchy – Schwarzova nerovnost). Protože ale $E\{V\} = 0; E\{T\} = g$ je

$$[E\{(V - E(V))(T - E(T))\}]^2 \leq [E\{(V - E(V))^2\} E\{(T - E(T))^2\}] = J(g)\sigma^2(T).$$

Ale: $E\{(V - E(V))(T - E(T))\} = E\{V(T - g)\} = \int_{Df} \frac{\frac{d}{dg} f(x, g)}{f(x, g)} (T - g) f(x, g) dx = 1$,

protože:

$$\int_{Df} \frac{\frac{d}{dg} f(x, g)}{f(x, g)} (T - g) f(x, g) dx = \int_{Df} \frac{\frac{d}{dg} f(x, g)}{f(x, g)} T f(x, g) dx - \int_{Df} \frac{\frac{d}{dg} f(x, g)}{f(x, g)} g f(x, g) dx =$$

$$= \frac{d}{dg} \int_{Df} T f(x, g) dx - g \frac{d}{dg} \int_{Df} f(x, g) dx = \frac{d}{dg} E\{T\} - g \frac{d}{dg} 1 = \frac{d}{dg} g = 1$$

Dosazením tohoto výsledku do nerovnosti:
$[E\{(V - E(V))(T - E(T))\}]^2 \leq [E\{(V - E(V))^2\} E\{(T - E(T))^2\}] = J(g)\sigma^2(T)$, dostaneme:

$$[1]^2 \leq J(g)\sigma^2(T) \Leftrightarrow \sigma^2(T) \geq \frac{1}{J(g)}$$

Pro **exponenciální typ rozdělení** $f(x, g) = e^{K(x)Q(g) + R(g) + S(x)}$: je
$\log f(x, g) = K(x)Q(g) + R(g) + S(x)$ a $\frac{d}{dg} \log f(x, g) = K(x)\frac{dQ}{dg} + \frac{dR}{dg}$ a

$$J(g) = E\left\{\left(K(x)\frac{dQ}{dg} + \frac{dR}{dg}\right)^2\right\} = E\left\{K^2(x)\left(\frac{dQ}{dg}\right)^2 + \left(\frac{dR}{dg}\right)^2 + 2K(x)\frac{dQ}{dg}\frac{dR}{dg}\right\} =$$

$$= E\{K^2(x)\}\left(\frac{dQ}{dg}\right)^2 + \left(\frac{dR}{dg}\right)^2 + 2E\{K(x)\}\frac{dQ}{dg}\frac{dR}{dg} \quad \textbf{(2)}$$

Protože podle vztahu (1) je:

$0 = \mathrm{E}\left\{\frac{d}{dg} log\ f(x,g)\right\} = E\{K(x)\}\frac{dQ}{dg} + \frac{dR}{dg} \Rightarrow E\{K(x)\}\frac{dQ}{dg} = -\frac{dR}{dg}$ a to dosazením do vztahu (2) dá: $J(g) = E\{K^2(x)\}\left(\frac{dQ}{dg}\right)^2 - \left(\frac{dR}{dg}\right)^2$.

Příklad:

Pro alternativní rozdělení $A(p) \cong f(x) = p^x(1-p)^{1-x} = e^{x\,lg(\frac{p}{1-p}) + lg(1-p)}, x = 0{,}1$ je

$K(x) = x; Q(p) = lg\left(\frac{p}{1-p}\right); R(p) = lg(1-p); S(x) = 0$.

$\frac{dQ}{dp} = \frac{1}{p(1-p)}; \frac{dR}{dp} = \frac{-1}{1-p}; E\{K(x)\} = E\{x^2\} = p$, pak:

$$J(p) = E\{K^2(x)\}\left(\frac{dQ}{dp}\right)^2 - \left(\frac{dR}{dp}\right)^2 = p\frac{1}{p^2(1-p)^2} - \frac{1}{(1-p)^2} = \frac{1}{p(1-p)} \Rightarrow nJ(p) = \frac{n}{p(1-p)}$$

proto:

$$\sigma^2\left(\hat{p}\right) \geq \frac{p(1-p)}{n}; \text{pokud}: \hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i; x_i \in \{0{,}1\} \text{ je } \sigma^2\left(\hat{p}\right) = \frac{1}{n}p(1-p)$$

Příklad:

Pro posunuté exponenciální rozdělení je $f(x;\delta) = \frac{1}{\delta}e^{\frac{x-A}{\delta}}, x \geq A$

$$K(x) = x - A; Q(\delta) = -\frac{1}{\delta}; R(\delta) = -lg(\delta); S(x) = 0.$$

$\frac{dQ}{d\delta} = \delta^{-2}; \frac{dR}{dp} = -\delta^{-1}; E\{K(x)\} = E\{(x-A)^2\} = 2\delta^2$, pak:

$$J(\delta) = E\{K^2(x)\}\left(\frac{dQ}{d\delta}\right)^2 - \left(\frac{dR}{d\delta}\right)^2 = 2\delta^2\delta^{-4} - \delta^{-2} = \delta^{-2} \Rightarrow nJ(\delta) = \frac{n}{\delta^2}; \ \sigma^2\left(\hat{\delta}\right) \geq \frac{\delta^2}{n}$$

Příklad souvislosti teorie informace a statistického testování

Základním pojmem teorie informace je entropie pravděpodobnostního rozdělení:

$H(X)=H(f)=-E_f\{\log f(x)\}=-\int_{-\infty}^{+\infty} f(x)\log f(x)dx$, pro spojitou náhodnou proměnnou a

$H(X)=H(p)=-E_p\{\log p(x)\}=-\sum_{x\in X} p(x)\log p(x)$, pro diskrétní náhodnou proměnnou. Za základ logaritmu se obvykle užívá 2, často také e (z „odvozovacích“ důvodů). Entropie je jistou mírou neurčitosti náhodné proměnné modelované daným rozdělením pravděpodobnosti (hustotou, pravděpodobností).

Zavedené definice předpokládají úplnou znalost pravděpodobnostního popisu. Pro případ, že tento předpoklad není ověřitelný nebo přijatelný se používá „křížová entropie“:

$H(X,g)=-E_f\{\log g(x)\}=-\int_{-\infty}^{+\infty} f(x)\log g(x)dx$ ve spojitém případě a v diskrétním případě

$H(X,g)=-E_p\{\log g(x)\}=-\sum_{x\in X} p(x)\log g(x)$.

Zde: $f(x), p(x)$ je skutečná hustota nebo pravděpodobnost a
$g(x)$ je modelem (statistikem) předpokládaná hustota nebo pravděpodobnost.

Toto pojetí již odlišuje skutečnost (obvykle nedostupnou) a model (jako aproximaci, neúplný popis skutečnosti, předpoklad o skutečnosti, …),

Poznámka a námět: nadále bude pracováno se „spojitými“ variantami, nebude problém dané postupy a výpočty převést na formy pro diskrétní náhodné proměnné (při využití aparátu teorie míry se oba přístupy shodují).

Pravděpodobnostnímu popisu $H(X,g)=-E_p\{\log g(x)\}=-\sum_{x\in X} p(x)\log g(x)$ odpovídá jeho výběrová (statistická) analogie:

$H_n((x_1,...,x_n),g)=\frac{1}{n}\left(-\sum_{i=1}^{n}\log g(x_i)\right)$, tedy střední hodnotu odhadujeme průměrem z náhodného výběru $(x_1,...,x_n)$, při zvoleném modelu $g(x)$, odtud dle „zákona velkých čísel“

$H_n((x_1,...,x_n),g)\xrightarrow{P} H(X,g)$ (v pravděpodobnosti).

Intuitivně: Čím větší je (výběrová) neurčitost, tím menší váhu přikládáme získaným pozorováním. Proto má smysl (zatím intuitivní) hledat u výběrové křížové entropie takový model $g(x)$, který ji minimalizuje. Lze dokázat, že $H_n((x_1,...,x_n),g)\to\min$ pro $g(x)\to f(x)$.

Námět: Dokažte toto tvrzení pro diskrétní rozdělení pravděpodobnosti s konečnou množinou elementárních jevů.

Statistické využití

U pozorovaných dat, daných náhodným výběrem $(x_1,...,x_n)$ předpokládáme, že se řídí rozdělením z parametrické rodiny $g(x;\nu)$ kde ν je hledaný parametr. Pak hledáme takovou hodnotu ν^* parametru ν, pro kterou platí:

$$-\frac{1}{n}\sum_{i=1}^{n}\lg g(x_i, v) \geq -\frac{1}{n}\sum_{i=1}^{n}\lg g(x_i, v^*) \Leftrightarrow \sum_{i=1}^{n}\lg g(x_i, v) \leq \sum_{i=1}^{n}\lg g(x_i, v^*) .$$

To ale není nic jiného, než metoda maximální věrohodnosti. Předchozí vztahy nám pak dávají jiný pohled na tuto metodu. Tedy odhad parametru minimalizujícího „křížovou entropii“ lze získat metodou maximální věrohodnosti.

Bayesovský přístup k statistickým odhadům a testování

Bayesovský přístup ve statistice (a i pravděpodobnosti) je založen na faktu, že parametr hustoty (pravděpodobnosti) lze chápat jako novou náhodnou proměnnou: $f(x;p) \stackrel{def}{=} f(x/p)$. V tom případě samozřejmě potřebujeme další informaci o (náhodnosti) parametru, obvykle se jedná o jeho apriorní hustotu $f(p)$. Pak můžeme konstruovat sdruženou hustotu pravděpodobnosti $f(x,p) = f(x/p)f(p) = f(x;p)f(p)$. Oprávnění pro takovou konstrukci dává Bayesova věta. Obvykle náhodná (i vektorová) proměnná x reprezentuje pozorování. Potom je užitečné znát popis „vlastností" parametru za předpokladu známých pozorovaných hodnot:

$$f(p/x) = \frac{f(x,p)}{f(x)} = \frac{f(x,p)}{\int f(x,\pi)d\pi} = \frac{f(x,p)}{\int f(x/\pi)f(\pi)d\pi} = \frac{f(x/p)f(p)}{\int f(x/\pi)f(\pi)d\pi}.$$

Za optimální (bayesovský) odhad $\hat{p}$ parametru(ů) p pak bývá považován odhad minimalizující hodnotu funkce $Q(r) = \int_\Omega (r-p)^2 f(p/x)dp$, kde Ω je definiční obor parametru p. Dále:

$$Q(r) = \int_\Omega \left(r^2 + p^2 - 2rp\right) f(p/x)dp = r^2 \int_\Omega f(p/x)dp + \int_\Omega p^2 f(p/x)dp - 2r\int_\Omega pf(p/x)dp =$$

$$= r^2 - 2rE\{p/x\} + \int_\Omega p^2 f(p/x)dp.$$

Odtud $\frac{dQ}{dr} = 2r - 2E\{p/x\} = 2(r - E\{p/x\}); \frac{d^2Q}{dr^2} = 2$, proto je minimum $Q(r)$ dosaženo pro $r_{opt} = \hat{p} = E\{p/x\}$, pokud uvedená střední hodnota existuje. To znamená, že optimální (ve výše uvedeném smyslu) bayesovský odhad parametru p je jeho „střední hodnota podmíněná znalostí pozorování".

Námět: Pokuste se odvodit totéž pro odhad vektorového parametru.

Oprávnění (neoprávnění) pro bayesovský přístup

V literatuře se vede poměrně rozsáhlá diskuse o oprávněnosti či neoprávněnosti bayesovského přístupu. Tento text si nečiní ambice takový problém řešit. Některé praktické úlohy však naznačují jeho přípustnost (v konkrétním případě).

Např.: Každý člověk má svůj hlasový frekvenční rozsah, ten je dán fyziologickými vlastnostmi hlasivek a celkové konstituce těla, včetně předcházejícího hlasového tréninku. Vlastní, realizovatelný, hlasový rozsah je však podmíněn náhodnými časově omezenými vlivy. Rozsah je jiný po „probdělé" noci, při nachlazení, jiný hlasový rozsah má dobře odpočatý a rozezpívaný zpěvák než neodpočatý a po náročném alkoholickém večírku, …. Jedná se tu o dvojí náhodnost. Jedna je v testovaném individuu a jedna je dána dalšími vlivy na takové individuum.

Jiné oprávnění může být i pragmatické. Bayesovské odhady mohou mít některé vlastnosti, které více „odpovídají realitě" než klasické

Motivační příklad

Odhad parametru alternativního rozdělení – triviální pohled

$P(x;p) = P(x/p) = p^x(1-p)^{1-x};\, x \in \{0,1\}$

$f(p) = 1 \Leftrightarrow 0 \le p \le 1$

$f(p) = 0 \Leftrightarrow p \notin \langle 0,1 \rangle$

$$P(x,p) = P(x/p)f(p) = p^x(1-p)^{1-x};\, x \in \{0,1\};\, p \in \langle 0,1 \rangle$$
$$= 0;\, p \notin \langle 0,1 \rangle$$

Poznámka: $P(x,p)$ ***je vůči proměnné*** x ***pravděpodobností a vůči proměnné*** p ***hustotou.***

Odtud: $P(x) = \int_0^1 P(x/p)f(p)dp = \int_0^1 p^x(1-p)^{1-x}dp = \dfrac{\Gamma(x+1)\Gamma(2-x)}{\Gamma(3)} = \dfrac{x!(1-x)!}{2!};\, x \in \{0,1\}$,

tedy $P(0) = P(1) = 1/2$ a proto $f(p/x) = \dfrac{P(x,p)}{P(x)} = \dfrac{p^x(1-p)^{1-x}}{1/2} = 2p^x(1-p)^{1-x}$. Potom:

$$\hat{p} = E\{p/x\} = \int_0^1 pf(p/x)dp = \int_0^1 2p^{x+1}(1-p)^{1-x}dp = 2\int_0^1 p^{x+1}(1-p)^{1-x}dp = 2\frac{\Gamma(x+2)\Gamma(2-x)}{\Gamma(4)} =$$

$$= 2\frac{(x+1)!(1-x)!}{3!} = \frac{(x+1)!(1-x)!}{3} = \frac{(x+1)}{3} = \frac{x+1}{2+1} = \frac{\text{pocet priznivych} + 1}{\text{pocet pozorovani} + \text{pocet moznych jevu}}$$

$\hat{p}(x) = \begin{matrix} x = 1 \Rightarrow \hat{p} = \dfrac{2}{3} \\ x = 0 \Rightarrow \hat{p} = \dfrac{1}{3} \end{matrix}$, tj. pokud máme k dispozici jediné pozorování je optimálním bayesovským odhadem parametru $\hat{p} = 1/3$, pokud nebyl pozorován jev sledovaný alternativním rozdělením tj. ***x=0*** a $\hat{p} = 2/3$ pokud byl pozorován jev sledovaný alternativním rozdělením tj. ***x=1***.

Poznámka: uvedený odhad nevede na nulové pravděpodobnosti. **Námět**: jak by vypadal klasický ne-bayesovský odhad parametru ***p*** při jediném pozorování?

Odhad parametru alternativního rozdělení – dokonalejší pohled

Mějme ***n*** nezávislých pozorování alternativního jevu $n(A)$ je počet pozorování, v nichž se jev A realizoval a $n - n(A)$ počet pozorování v nichž se jev nerealizoval. Potom náhodná proměnná $n(A)$ má binomické rozdělení pravděpodobnosti $P(n(A) = k) = \binom{n}{k} p^k(1-p)^{n-k}$.

To opět můžeme chápat jako podmíněné rozdělení pravděpodobnosti $P(n(A) = k/p) = \binom{n}{k} p^k(1-p)^{n-k}$. O p budeme pak předpokládat jeho rovnoměrné apriorní rozdělení pravděpodobnosti, tj. $\begin{array}{ll} P(k,p)=\binom{n}{k}p^k(1-p)^{n-k} & \Leftrightarrow p\in\langle 0,1\rangle \\ =0 & \Leftrightarrow p\notin\langle 0,1\rangle \end{array}$, proto

$$P(k)=\int_0^1\binom{n}{k}p^k(1-p)^{n-k}dp=\binom{n}{k}\int_0^1 p^k(1-p)^{n-k}dp=\binom{n}{k}\frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)}=\frac{n!}{(n-k)!k!}\frac{k!(n-k)!}{(n+1)!}=$$

$$=\frac{1}{n+1}$$

Tedy apriorní rozdělení počtu příznivých pozorování danému jevu A, při apriorním rozdělení neznámého parametru p je rovnoměrné na množině $\{0,1,\ldots,n\}$.

Potom: $f(\,p/k\,)=\dfrac{P(\,k,p\,)}{P(\,k\,)}=\dfrac{\binom{n}{k}p^k(1-p)^{n-k}}{\frac{1}{n+1}}=(n+1)\binom{n}{k}p^k(1-p)^{n-k}$ a

$$\hat{p}=E\{p/k\}=\int_0^1 pf(\,p/x\,)dp=\int_0^1 p(n+1)\binom{n}{k}p^k(1-p)^{n-k}dp=(n+1)\binom{n}{k}\int_0^1 p^{k+1}(1-p)^{n-k}dp=$$

$$=(n+1)\binom{n}{k}\frac{\Gamma(\,k+2\,)\Gamma(\,n-k+1\,)}{\Gamma(\,n+3\,)}=(n+1)\frac{n!}{(\,n-k\,)!\,k!}\frac{(\,k+1\,)!\,(\,n-k\,)!}{(\,n+2\,)!}=\frac{k+1}{n+2}$$ a proto:

$$\hat{p}=E\{p/k\}=\frac{k+1}{n+2}=\frac{\text{pocet priznivych}+1}{\text{pocet pozorovani}+\text{pocet moznych jevu}}$$ a

$$\text{pocet pozorovatelnych moznosti}=\left|\{0,1,\ldots,n\}\right|.$$

Naproti tomu klasický (frekvenční odhad) $\tilde{p}=\dfrac{k}{n}=\dfrac{\text{pocet priznivych}}{\text{pocet pozorovani}}$. Zatímco klasický odhad dává při nulovém počtu příznivých pozorování $\tilde{p}=0$ (= nepoužitelný výsledek) bayesovský dává $\hat{p}=E\{p/k\}=\dfrac{1}{n+2}=\dfrac{1}{\text{pocet pozorovani}+\text{pocet moznych jevu}}$, tedy ne-nulu a takovým odhadem jde již dále pracovat.

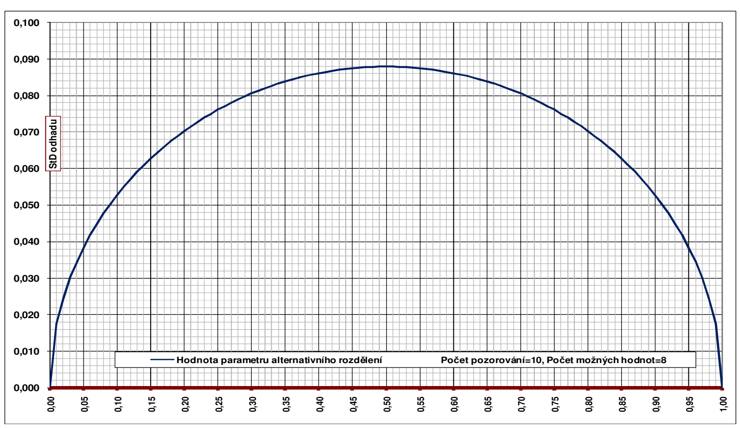
Výsledku $\tilde{p}=0$ se říká ***ZERO FREQUENCY PROBLEM.*** Komplikuje někdy statistická zkoumání např. v genetice. Zde často pracujeme se situací, kdy musíme předpokládat existenci dosud neobjeveného genu (NULL GEN), ten samozřejmě nebude při běžném statistickém zkoumání pozorován. Při užití klasického odhadu pak bychom pracovali s jeho nulovou pravděpodobností a to je spor s předpokladem, že mimo známé geny může být objeven další (a má(mají) nenulovou pravděpodobnost). Obdobná situace i v jednodušších příkladech (např. při odhadech pravděpodobnosti poruchy spolehlivého zařízení, málo pravděpodobný, leč možný typ poruchy nemusí být pozorován během statistického šetření).

Námět: Pro náhodnou (jevovou) proměnnou s hodnotami v abecedě (jevů) $A=\{a_1,\ldots,a_k\}$ s počtem hodnot $|A|=k<+\infty$ s multinomickým rozdělením pozorovaných počtů jednotlivých jevů odvoďte optimální bayesovský odhad $\hat{p}_i = \dfrac{n_i + 1}{n + k} = \dfrac{n_i + 1}{n + |A|}$ za předpokladu apriorního rovnoměrného rozdělení hodnot parametrů p_i na „intervalu“ $\langle 0,1 \rangle^k$.

Vlastnosti odhadu $\hat{p}_i = \dfrac{n_i + 1}{n + k} = \dfrac{n_i + 1}{n + |A|}$. Tento odhad bývá někdy také nazýván MAX-ENTROPY odhadem.

1. $\displaystyle\sum_{i=1}^{k} \hat{p}_i = \sum_{i=1}^{k} \frac{n_i + 1}{n + k} = \frac{\sum_{i=1}^{k} (n_i + 1)}{n + k} = \frac{n + k}{n + k} = 1$
2. $\hat{p}_i = \dfrac{n_i + 1}{n + k} = \dfrac{\frac{n_i}{n} + \frac{1}{n}}{1 + \frac{k}{n}} = \dfrac{\tilde{p}_i + \frac{1}{n}}{1 + \frac{k}{n}} = \tilde{p}_i \dfrac{n}{n + k} + \dfrac{1}{n + k}$
3. Protože n_i má při dané realizaci p_i binomické rozdělení pravděpodobnosti tedy $E\{n_i\} = np_i$ a proto $E_{n_i}\{\hat{p}_i\} = \dfrac{E\{n_i\} + 1}{n + k} = \dfrac{np_i + 1}{n + k} = p_i \dfrac{n}{n + k} + \dfrac{1}{n + k}$.
4. $\sigma^2\{n_i\} = np_i(1 - p_i)$, odtud $\sigma^2_{n_i}\{\hat{p}_i\} = \dfrac{\sigma^2\{n_i\}}{(n + k)^2} = \dfrac{np_i(1 - p_i)}{(n + k)^2} = p_i(1 - p_i)\dfrac{n}{(n + k)^2}$ a $\sigma_{n_i}\{\hat{p}_i\} = \sqrt{p_i(1 - p_i)}\dfrac{\sqrt{n}}{(n + k)}$.

   Závislost chyby odhadu (měřená jeho směrodatnou odchylkou) na hodnotě tohoto parametru:



Další příklad:

Poissonovo rozdělení

Pozorujeme v daném časovém intervalu nebo prostorovém objemu počet výskytů jevu A.

$$P\{n(A) = k; \lambda\} = e^{-\lambda} \frac{\lambda^k}{k!}$$

Apriorní rozložení parametru λ je rovnoměrné na intervalu $[\lambda_D, \lambda_H]$. Pak:

$$f(\lambda)=\frac{1}{\lambda_{\mathrm{H}}-\lambda_{\mathrm{D}}} \Leftrightarrow \lambda\in[\lambda_{\mathrm{D}},\lambda_{\mathrm{H}}]$$
jinak $=0$

$, P\{n(A)=k/\lambda\}=e^{-\lambda}\frac{\lambda^k}{k!},$

$$P\{n(A)=k,\lambda\}=\frac{1}{\lambda_{\mathrm{H}}-\lambda_{\mathrm{D}}}e^{-\lambda}\frac{\lambda^k}{k!}\chi_{[\lambda_{\mathrm{D}},\lambda_{\mathrm{H}}]}(\lambda).$$

Potom:

$P\{n(A)=k\}=\frac{1}{k!}\frac{1}{\lambda_{\mathrm{H}}-\lambda_{\mathrm{D}}}\int_{\lambda_{\mathrm{D}}}^{\lambda_{\mathrm{H}}}\lambda^k e^{-\lambda}d\lambda=\frac{1}{k!}\frac{1}{\lambda_{\mathrm{H}}-\lambda_{\mathrm{D}}}I_k;kde\ \ I_k=\int_{\lambda_{\mathrm{D}}}^{\lambda_{\mathrm{H}}}\lambda^k e^{-\lambda}d\lambda$, ale:

$I_k=\int_{\lambda_{\mathrm{D}}}^{\lambda_{\mathrm{H}}}\lambda^k e^{-\lambda}d\lambda=\left[-e^{-\lambda}\lambda^k\right]_{\lambda_{\mathrm{D}}}^{\lambda_{\mathrm{H}}}+k\int_{\lambda_{\mathrm{D}}}^{\lambda_{\mathrm{H}}}\lambda^{k-1}e^{-\lambda}d\lambda=\left[-e^{-\lambda}\lambda^k\right]_{\lambda_{\mathrm{D}}}^{\lambda_{\mathrm{H}}}+kI_{k-1}$ (per partes), shrnuto:

$I_k=e^{-\lambda_{\mathrm{D}}}(\lambda_{\mathrm{D}})^k-e^{-\lambda_{\mathrm{H}}}(\lambda_{\mathrm{H}})^k+kI_{k-1}$, kde: $I_0=\int_{\lambda_{\mathrm{D}}}^{\lambda_{\mathrm{H}}}\lambda^0 e^{-\lambda}d\lambda=e^{-\lambda_{\mathrm{D}}}-e^{-\lambda_{\mathrm{H}}}$.

A potom: $P\{n(A)=k\}=\frac{1}{k!}\frac{I_k}{\lambda_{\mathrm{H}}-\lambda_{\mathrm{D}}}$, $f(\lambda/n(A)=k)=\frac{P(n(A)=k,\lambda)}{P(n(A)=k)}=$

$=\dfrac{\frac{1}{\lambda_{\mathrm{H}}-\lambda_{\mathrm{D}}}e^{-\lambda}\frac{\lambda^k}{k!}\chi_{[\lambda_{\mathrm{D}},\lambda_{\mathrm{H}}]}(\lambda)}{\frac{1}{k!}\frac{1}{\lambda_{\mathrm{H}}-\lambda_{\mathrm{D}}}I_k}=\dfrac{e^{-\lambda}\lambda^k\chi_{[\lambda_{\mathrm{D}},\lambda_{\mathrm{H}}]}(\lambda)}{I_k}$. Pak pro bayesovský optimální odhad $\overline{\lambda}=E\{\lambda/n(A)=k\}=\int_{\lambda_D}^{\lambda_H}\lambda\frac{e^{-\lambda}\lambda^k}{I_k}d\lambda'=\frac{1}{I_k}\int_{\lambda_D}^{\lambda_H}e^{-\lambda}\lambda^{k+1}d\lambda'=\frac{I_{k+1}}{I_k}$. Celkové shrnutí:

Pozorujeme: $n(A)=k$ a dostáváme: $I_0=e^{-\lambda_{\mathrm{D}}}-e^{-\lambda_{\mathrm{H}}}$, $I_k=e^{-\lambda_{\mathrm{D}}}(\lambda_{\mathrm{D}})^k-e^{-\lambda_{\mathrm{H}}}(\lambda_{\mathrm{H}})^k+kI_{k-1}$, $\overline{\lambda}=\frac{I_{k+1}}{I_k}$.

Poznámka: $E\{n(A)=k\}=\lambda$ *a také* $\sigma^2\{n(A)=k\}=\lambda$.

Konjugovaný systém hustot (pravděpodobností)

Pro výpočetní efektivitu je vhodné, aby apriorní $f(r)$ a aposteriorní $f(r/x_1,\ldots,x_n)$ rozdělení byla „stejného typu“. O apriorních hustotách $f(r)\in Q$, řekneme, že jsou konjugovány s hustotami $f(x_1,\ldots,x_n/r)$, pokud platí $f(r/x_1,\ldots,x_n)=\frac{f(x_1,\ldots,x_n,r)}{f(x_1,\ldots,x_n)}\in Q$.

Detailněji rozepsáno: $f(r/x_1,\ldots,x_n)=\frac{f(x_1,\ldots,x_n/r)f(r)}{\int_{r\in R}f(x_1,\ldots,x_n/r)f(r)dr}\in Q$. Pokud jsou naše pozorování iid pak $f(r/x_1,...,x_n)=\dfrac{\prod_{i=1}^{n} f(x_i/r)f(r)}{\int_{r\in R}\prod_{i=1}^{n} f(x_i/r)f(r)dr}$. Je zřejmé, že takovéto definici vyhovuje soubor Q všech možných hustot. Proto je vhodné zavést užší systémy konjugovaných hustot. Podstatnou roli v takových systémech hrají postačující statistiky. Detaily lze nalézt v Marie Hušková: Bayesovské metody, skripta MFF-UK, Praha 1985, str. 13-26. Zde jen některé „konjugované“ systémy:

Hustota – pravděpodobnost	$f(x/r)$	Konjugovaná hustota (parametrický systém), apriorní rozdělení	Aposteriorní rozdělení při ***n*** pozorováních
Binomická	$f(x/r)=\binom{m}{x}r^x(1-r)^{m-x}$	$f(r)=\dfrac{1}{B(a,b)}r^{a-1}(1-r)^{b-1}$ = Beta rozdělení	$f(r/x_1,...,x_n)=\dfrac{1}{B(a,b)}r^{a+\sum x_i-1}(1-r)^{b+mn-\sum x_i-1}$
Poissonovo	$f(x/r)=e^{-r}\dfrac{r^x}{x!}$	$f(r)=\dfrac{a^p}{\Gamma(p)}r^{p-1}e^{-ar}$ = Gama rozdělení	$f(r/x_1,...,x_n)=\dfrac{(a+n)^{p+\sum x_i}}{\Gamma(p+\sum x_i)}r^{p+\sum x_i-1}e^{-(a+n)r}$
Rovnoměrné	$f(x/r)=\begin{matrix}1/r \Leftrightarrow 0\le x\le r\\ 0 \quad \text{jinak}\end{matrix}$	$f(r)=\begin{matrix}\dfrac{m}{t}\left(\dfrac{t}{r}\right)^{m+1} \Leftrightarrow r\ge t\\ 0 \quad \text{jinak}\end{matrix}$ = Paretovo rozdělení	$f(r/x_1,...,x_n)=\begin{matrix}\dfrac{m+n}{\max(t,x_1,...,x_n)}\left(\dfrac{\max(t,x_1,...,x_n)}{r}\right)^{m+n+1} \Leftrightarrow r\ge t\\ 0 \quad \text{jinak}\end{matrix}$
Normální vůči μ	$f(x/r)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-r)^2}{2\sigma^2}}$	$f(r)=\dfrac{1}{\sqrt{2\pi}b}e^{-\frac{(r-a)^2}{2b^2}}=N(a,b^2)$	$N(\mu_1,\sigma_1^2);\mu_1=\dfrac{b^2\sum x_i+a\sigma^2}{nb^2+\sigma^2}$ $\sigma_1^2=\dfrac{b^2\sigma^2}{nb^2+\sigma^2}$
……..			

Jiné způsoby volby apriorního rozdělení:

1. Histogram.
2. Volba hustoty podle principu neurčitosti – volí se „maximálně neurčité rozdělení“.
3. Nevlastní hustoty.
4. …..

Histogram

Volí se často u bayesovských přístupů vycházejících ze statistických šetření, zde $f(r/x)=\dfrac{f(x,r)}{f(x)}=\dfrac{f(x/r)f(r)}{f(x)}$.

Nechť je dáno kvantování intervalu pro „neznámý parametr“ r, $r_1,r_2,...,r_k$ s „relativními četnostmi $f_0 \Leftrightarrow r<r_1, f_1 \Leftrightarrow r_1\le r<r_2,...,f_{k-1} \Leftrightarrow r_{k-1}\le r<r_k, f_k \Leftrightarrow r\ge r_k$. Pak:

$$f(x,r) = f_0 f(x/r) \Leftrightarrow r < r_1, f(x,r) = f_1 f(x/r) \Leftrightarrow r_1 \le r < r_2, ...,$$
$$f(x,r) = f_{k-1} f(x/r) \Leftrightarrow r_{k-1} \le r < r_k, f(x,r) = f_k f(x/r) \Leftrightarrow r \ge r_k$$

Volba hustoty podle principu neurčitosti – volí se „maximálně neurčité rozdělení“.

Tato metoda již byla demonstrována na odhadu parametru(ů) alternativního rozdělení – za apriorní byla volena rovnoměrná. Někdy jsou voleny hustoty s maximální entropií.

Nevlastní hustoty.

Vzhledem ke tvaru aposteriorního rozdělení $f(r/x) = \dfrac{f(x/r)f(r)}{\int f(x;r)f(r)dr}$ není nezbytné, aby

apriorní hustota byla hustotou v přesném slova smyslu, např. nemusí platit $\int f(r)dr = 1$.

Příklad:
Mějme náhodný výběr $\{x_1, x_2, ..., x_n\}$ z exponenciálního (neposunutého) rozdělení $f(x/r) = re^{-rx}$. Potom $f(x_1, x_2, ..., x_n / r) = r^n e^{-r\sum x_i}$. Budeme předpokládat, že $f(r) = \begin{matrix} 1 \Leftrightarrow r > 0 \\ 0 \Leftrightarrow r \le 0 \end{matrix}$. Tj. nejedná se hustotu!!!!!! Její integrál na definičním oboru je dokonce nevlastní!!!!!!

Potom: $f(x_1, x_2, ..., x_n, r) = r^n e^{-r\sum x_i} \Leftrightarrow r > 0$ a je rovno 0 jinak. Označíme $S = \sum x_i$, pak:
$f(x_1, x_2, ..., x_n, r) = r^n e^{-rS} \Leftrightarrow r > 0$ a je rovno 0 jinak. Proto
$f(x_1, x_2, ..., x_n) = \int_0^{+\infty} r^n e^{-rS} dr = \dfrac{1}{S^{n+1}} \int_0^{+\infty} y^n e^{-y} dy = \dfrac{n!}{S^{n+1}}$ a $f(r / x_1, x_2, ..., x_n) = \dfrac{S^{n+1}}{n!} r^n e^{-rS}$. Dále

$$E\{r / x_1, x_2, ..., x_n\} = \frac{S^{n+1}}{n!} \int_0^{+\infty} r^{n+1} e^{-rS} dr = \frac{S^{n+1}}{n!} \frac{1}{S^{n+2}} \int_0^{+\infty} y^{n+1} e^{-y} dy = \frac{S^{n+1}}{n!} \frac{(n+1)!}{S^{n+2}} = \frac{n+1}{S} = \frac{n+1}{\sum x_i} = \overline{\lambda}.$$

Srovnejte s maximálně věrohodným odhadem $\hat{\lambda} = \dfrac{n}{\sum x_i}$.

Námět: Pokuste se řešit tutéž úlohu pro parametrizaci: $f(x/\tau) = \dfrac{1}{\tau} e^{-\frac{x}{\tau}}$

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.
Alfréd Rényi	Teorie pravděpodobnosti, ACADEMIA, Praha 1972
C. Radhakrishna Rao	Lineární metody statistické indukce a jejich aplikace, ACADEMIA, Praha 1978
Marie Hušková	Bayesovské metody, skripta MFF-UK, Praha 1985
Marie Hušková	Sekvenční analýza, skripta MFF-UK, Praha 1982
SHANNON, C. E.	A Mathematical Theory of Communication, The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.
Thomas M. Cover, Joy A. Thomas	Elements of Information Theory. 1991 John Wiley & Sons, Inc. ISBN 0-471-06259-6. Existuje i novější vydání.

Další z teorie informace:

INDER JEET TANEJA	GENERALIZED RELATIVE INFORMATION AND INFORMATION, Journal of Inequalities in Pure and Applied Mathematics, http://jipam.vu.edu.au/ Volume 5, Issue 1, Article 21, 2004
PRANESH KUMAR AND ANDREW JOHNSON	ON A SYMMETRIC DIVERGENCE MEASURE AND INFORMATION INEQUALITIES, Journal of Inequalities in Pure and Applied Mathematics http://jipam.vu.edu.au/ Volume 6, Issue 3, Article 65, 2005
	http://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence http://en.wikipedia.org/wiki/Jensen-Shannon_divergence
Vajda, I.:	Teória informácie a štatistického rozohodovania, Alfa Bratislava 1982.
Cziszár, I.:	Information Theoretic Methods in Probability and Statistics, Conference Fifty Years of Shannon Theory, 1998 and Shannon Lecture ISIT'97.
Cziszár, I., Körner J.:	Information Theory, Coding Theorems for Discrete Memoryless Systems. KIADÓ, BUDAPEST 1986.
Zhai, C.X.:	Essential Probability and Statistics, Department of Computer Science, University of Illinois, Urbana-Champaign, Lecture for CS397-CXZ Algorithms in Bioinformatics. Jan. 23, 2004.