Verze 2004-12-22 Ref: in Analýza dat 2004/II (K. Kupka, ed.), Trilobyte Statistical Software, Pardubice, 2005, pp. 21–33.
Michal Friesl

BAYESOVSKÉ ODHADY V NĚKTERÝCH MODELECH

MICHAL FRIESL

Katedra matematiky, Fakulta aplikovaných věd, Západočeská univerzita, Univerzitní 22, 306 14 Plzeň
E-mail: friesl@kma.zcu.cz*, URL:* http://home.zcu.cz/~friesl/

Abstrakt. Příspěvek se v několika příkladech zabývá bayesovskými odhady a některými jejich vlastnostmi. Popsány jsou odhady parametrů určitých základních rozdělení, odhady z oblasti regrese, nebo analýzy dat o přežití. Příklady zahrnují i empirické bayesovské odhady a neparametrické bayesovské odhady.

1. Bayesovský přístup

V obvyklém modelu matematické statistiky máme k dispozici výsledky náhodného pokusu v podobě pozorování náhodných veličin $X = (X_1, \dots, X_n)$ s hustotou $f(x) = f(x;\theta)$. Hustota f je známa jen částečně, je znám její tvar až na několik parametrů, θ. Např. víme, že X tvoří náhodný výběr z normálního rozdělení $\mathrm{N}(\mu, \sigma^2)$, jehož parametry, $\theta = (\mu, \sigma^2)$ ale neznáme. Úkolem je učinit určité závěry o rozdělení pozorovaných hodnot týkající se parametru θ nebo jeho funkcí. Řešíme takové úlohy, jako je hledání bodového nebo intervalového odhadu, testování hypotéz, apod.

Při klasickém přístupu počítáme s parametrem θ jako s neznámou, ale pevnou konstantou, k závěrům používáme tvar hustoty $f(x;\theta)$ a pozorování X. Různé typy odhadu zahrnují odhad metodou maximální věrohodnosti, momentové odhady, atd.

Při bayesovském přístupu naproti tomu považujeme parametr θ za náhodnou veličinu, jejíž hodnotu sice nepozorujeme, ale jejíž rozdělení známe. Hustotu veličiny θ značme $\pi(\theta)$. Tato hustota vyjadřuje apriorní informaci o možných hodnotách parametru θ, informaci, kterou máme ještě před pokusem, tedy získanou nezávisle na pozorováních X. Technicky, hustotu pozorování $f(x;\theta)$ chápeme při tomto přístupu jako podmíněnou hustotu $f(x\,|\,\theta)$ veličiny X při daných hodnotách veličiny θ. K závěrům pak oproti klasickému přístupu navíc použijeme apriorní hustotu $\pi(\theta)$.

Apriorní hustota může být zvolena zcela objektivně, např. na základě zkušenosti s pozorováními z minulosti, nebo na základě vnějších informací, např. z fyzikální podstaty problému. Možná je ale také subjektivní volba, vyjadřující individuální názor na pravděpodobnosti výskytu jednotlivých hodnot parametru. Vyskytuje se i volba z nouze tak, aby „to šlo spočítat“. Pokud nám jde ale jen o konkrétní hodnoty odhadů parametrů a nepotřebujeme odhady studovat třeba teoreticky, není problém zvolit apriorní hustotu, která věrně odráží apriorní informace, a odhady hledat pro konkrétní data s pomocí počítače.

Veškeré závěry založené na datech se odvíjejí od aposteriorního rozdělení parametru. Vztah mezi apriorní a aposteriorní hustotou zachycuje Bayesova věta:

Tato práce vznikla s podporou výzkumného záměru MSM 4977751301.

Tvrzení 1.1. *Má-li vektor (X, Y) sdruženou hustotu $f(x, y)$, pak podmíněná hustota složky Y za podmínky, že $X = x$, je*

$$f(y \mid x) = \frac{f(x \mid y) f(y)}{f(x)},$$

kde $f(x \mid y)$ značí podmíněnou hustotu X při daných hodnotách složky Y, $f(x)$ a $f(y)$ jsou marginální hustoty složek.

Přepíšeme-li tvrzení v našem označení, dostáváme pro aposteriorní hustotu parametru θ při daných hodnotách pozorování $X = x$ vztah

$$\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta). \tag{1.1}$$

Aposteriorní hustota je až na nějakou normovací konstantu rovna součinu věrohodnostní funkce $L(\theta) = f(x; \theta) = f(x \mid \theta)$ pro parametr θ na základě pozorování X a apriorní hustoty $\pi(\theta)$. Kombinuje tak apriorní informaci o parametru s informací obsaženou v pozorováních.

Jako apriorní hustoty se používají i tzv. nevlastní hustoty, tj. nezáporné funkce $\pi(\theta)$ s $\int \pi(\theta) = \infty$ (integrál není roven 1, π tedy není vlastní hustotou, ani ji nelze znormovat tak, aby se hustotou stala), pokud aposteriorní hustota vyjde vlastní. V případě práce s diskrétními rozdělení na místech hustot vystupují pravděpodobnostní funkce a na místech integrálů sumy.

Podobně jako u klasických odhadů, rozlišujeme i při bayesovském přístupu různé typy odhadů. Jako bodový odhad můžeme použít obdobu maximálně věrohodného odhadu, tj. tu hodnotu θ, kde aposteriorní hustota $\pi(\theta|x)$ nabývá nejvyšší hodnoty. Jiným typem odhadů jsou střední hodnota aposteriorního rozdělení $\widehat{\theta} = \mathrm{E}(\theta \mid X)$, nebo aposteriorní medián, odpovídající minimalizaci průměrné ztráty při ztrátové funkci kvadratické, $L(\widehat{\theta}, \theta) = (\widehat{\theta} - \theta)^2$, resp. s absolutní hodnotou, $L(\widehat{\theta}, \theta) = |\widehat{\theta} - \theta|$. Bayesovskou $(1-\alpha)$-konfidenční oblast (interval) definujeme jako takovou množinu I, pro niž $\mathrm{P}(\theta \in I|X) = 1-\alpha$. O hypotézách můžeme rozhodovat přímo porovnáním pravděpodobností, s jakými při aposteriorním rozdělení nastávají, nebo také pomocí ztrátových funkcí.

Podrobný výklad o bayesovských metodách, včetně odpovídajících partií z teorie rozhodování, lze nalézt např. v knize Berger (1985) nebo ve skriptech Hušková (1985). My se teď budeme věnovat slibovaným příkladům bayesovských odhadů.

2. Pravděpodobnosti úspěchů

V prvním příkladu se budeme zabývat Bernoulliovým schématem. Naším úkolem je odhadnout pravděpodobnost p výskytu určitého jevu (pravděpodobnost úspěchu) v náhodném pokusu. Proto pokus n-krát nezávisle zopakujeme a sledujeme počet výskytů jevu v těchto n opakováních. Náhodná veličina S udávající počet výskytů v n opakováních, tedy počet „úspěchů“ v bernoulliovské posloupnosti, má binomické rozdělení $S \sim \mathrm{Bi}(n, p)$, věrohodnostní funkce pro parametr p má tedy tvar

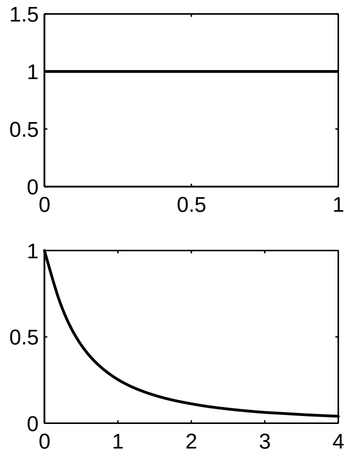
$$L(p) = \mathrm{P}(S = k \mid p) = \binom{n}{k} p^k (1-p)^{n-k}, \tag{2.1}$$

pozorovaný počet výskytů k může být $0, 1, \ldots, n$.

Nemáme-li žádnou apriorní informaci o pravděpodobnosti p, zvolíme pro ni neurčitou hustotu rovnoměrně rozloženou na oboru hodnot

$$\pi(p) = 1, \quad p \in (0, 1). \tag{2.2}$$

Je ovšem nutné poznamenat, že tomuto neurčitému apriornímu rozdělení pro pravděpodobnost p odpovídá u poměru šancí $\gamma = p/(1-p)$ hustota $\pi(\gamma) = (1+\gamma)^{-2}$, $\gamma > 0$, preferující nízké hodnoty parametru γ před většími (viz obr. 1). Neurčité



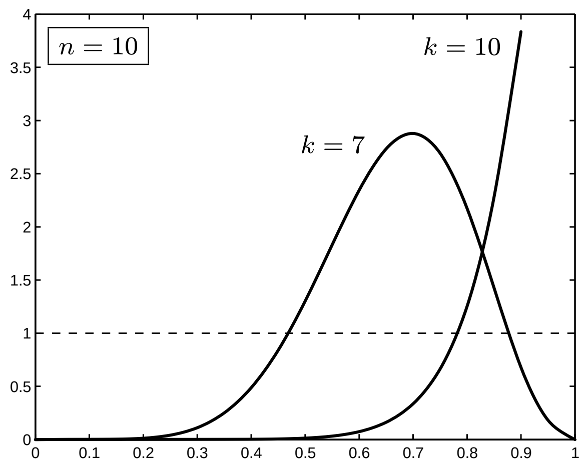
Obrázek 1. Odpovídající si hustoty parametru p (nahoře, rovnoměrná) a parametru γ (dole, rovnoměrná není).

informaci o pravděpodobnosti p tedy neodpovídá neurčitá, rovnoměrně rozložená informace o poměru šancí γ.

Výpočet aposteriorního rozdělení je jednoduchý, podle Bayesovy věty je úměrná součinu (2.1) a (2.2). Při výpočtu si nemusíme všímat násobící konstanty $\binom{n}{k}$ z (2.1), protože neobsahuje proměnnou p. Píšeme stručně

$$\pi(p \mid S = k) \propto p^k (1-p)^{n-k} \cdot 1, \quad p \in (0,1).$$

Až na normovací konstantu vidíme ve vyjádření aposteriorní hustoty hustotu beta rozdělení $\mathrm{B}(k+1, n-k+1)$, na obr. 2 je znázorněn její tvar v konkrétním případě (další znázornění hustot beta rozdělení viz obr. 3). Bayesovským odhadem hledané



Obrázek 2. Apriorní hustota (čárkovaně) a aposteriorní hustoty parametru p v případě $k = 7$, resp. $k = 10$ úspěchů při $n = 10$ opakováních.

pravděpodobnosti p je pomocí aposteriorní střední hodnoty odhad

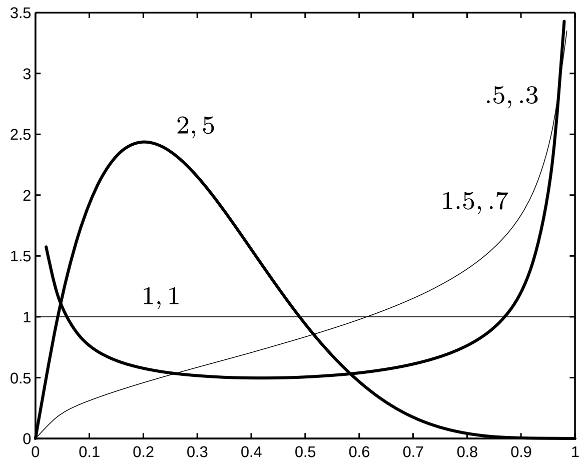
$$\widehat{p} = \mathrm{E}(p \mid S = k) = \frac{k+1}{n+1},$$

zároveň jde i o bayesovský odhad maximálně věrohodného typu.

Apriorním rozdělením, které umožňuje zachytit bohatší než neurčitou apriorní informaci, a přitom ještě nezkomplikuje výpočet, je beta rozdělení $\mathrm{B}(a, b)$, $a, b > 0$, s hustotou

$$\pi(p) = B(a,b)^{-1} p^{a-1}(1-p)^{b-1}, \quad p \in (0,1), \tag{2.3}$$

kde $B(a, b)$ označuje beta funkci. Rovnoměrnou hustotu (2.2) dostaneme při $a =$



Obrázek 3. Hustota beta rozdělení $\mathrm{B}(a, b)$ pro různé kombinace hodnot parametrů a a b.

$b = 1$. Aposteriorním rozdělením plynoucím z (2.3) je rozdělení s hustotou

$$\begin{aligned} \pi(p \mid S = k) &\propto p^k (1-p)^{n-k} \cdot p^{a-1}(1-p)^{b-1} \\ &= p^{a+k}(1-p)^{b+n-k} \sim \mathrm{B}(a+k, b+n-k), \end{aligned} \tag{2.4}$$

tedy znovu beta rozdělení. Přechod od apriorní k aposteriorní hustotě tak spočívá jen v úpravě parametrů rozdělení. Parametr p bychom bayesovsky odhadli jako

$$\widehat{p} = \frac{a+k}{a+b+n}, \tag{2.5}$$

což je střední hodnota aposteriorního rozdělení.

Někdy máme důvod uvažovat jako apriorní rozdělení určitý typ rozdělení, neznáme ale jeho parametry (zde parametry a a b). Je-li v takové situaci k dispozici alespoň informace o pravděpodobnosti výskytu různých jejich hodnot, můžeme i parametry apriorního rozdělení dále považovat za náhodné veličiny (mluví se o hyperparametrech), zavést další úroveň apriorních rozdělení a analogické vzorce pro aposteriorní rozdělení při daných pozorováních odvodit i v takto komplikovanějším modelu.

Jinou možností je nahradit neznámé parametry apriorního rozdělení vhodnými odhady. Při jejich hledání můžeme využít i případná další pozorování, jejichž rozdělení vychází ze stejného apriorního rozdělení. Bayesovské odhady, kde k odhadu parametrů apriorního rozdělení bylo použito dat, se označují jako empirické bayesovské. Takový typ odhadu použijeme v následujícím příkladu.

3. Nehody řidičů

Chceme pro jednotlivé řidiče odhadnout pravděpodobnost, s jakou budou mít v příštím roce nehodu (např. kvůli stanovení výše pojistného pro nastávající rok v pojištění vozidel). K dispozici máme informace o počtech „nehodových“ roků za uplynulých $n = 10$ let od $N = 20$ řidičů:

$$0,0,2,0,0,2,2,0,6,4,3,1,1,1,0,0,5,1,1,0. \tag{3.1}$$

Předpokládáme, že pravděpodobnosti nehod se u řidičů v pozorovaných letech neměnily, u j-tého řidiče ji značme p_j, $j = 1,\dots,N$. Zaveďme dále pomocné veličiny X_{ji} s hodnotami $X_{ji} = 1$, když j-tý řidič měl v i-tém roce nehodu, a $X_{ji} = 0$, když ji neměl. Roky považujeme za nezávislé, tedy náhodné veličiny $S_j = \sum_i X_{ji}$, udávající počty roků s nehodou za uplynulých n let u jednotlivých řidičů, mají binomická rozdělení, $S_j \sim \mathrm{Bi}(n, p_j)$.

Bylo by možné odhadnout pravděpodobnosti p_j společně pro všechny řidiče, takovým kolektivním odhadem by bylo

$$\widehat{p} = (N^{-1}\sum_j S_j)/n = (1/20)\cdot 29/10 = 0{,}145.$$

Naším cílem je ale odhadnout pravděpodobnosti p_j individuálně. Také χ^2-test homogenity rozdělení veličin $X_{ji\cdot}$, $j = 1,\dots,N$, ukazuje, že je nelze považovat za shodné. Naivní individuální odhady $\widehat{p}_j = S_j/n$ ale nejsou pro přímé stanovení pojistného vhodné, např. řidičům, kteří neměli v minulosti nehodu, by podle nich náleželo nulové pojistné.

Předpokládejme, že pravděpodobnost p pro nehodu v roce je mezi řidiči rozložena s beta rozdělením, $p_j \sim \mathrm{B}(a,b)$[1], Tomu odpovídá dle (2.4) aposteriorní rozděleni $(p_j \mid S_j = k) \sim \mathrm{B}(a+k, b+n-k)$ a bayesovský odhad (2.5), který můžeme pro j-tého řidiče zapsat ve tvaru

$$\widehat{p}_j^{bay} = \frac{a+b}{a+b+n}\underbrace{\frac{a}{a+b}}_{p_0} + \frac{n}{\underbrace{a+b}_{n_0}+n}\frac{k}{n} = \frac{n_0}{n_0+n}p_0 + \frac{n}{n_0+n}\widehat{p}_j, \tag{3.2}$$

tedy jako lineární kombinaci střední hodnoty p_0 parametru p (je $p_0 = \mathrm{E}\,p = a/(a+b)$) a individuálního odhadu $\widehat{p}_j = \sum_i X_{ji}/n = k/n$. Lze ukázat, že tento odhad je nejlepším odhadem lineárním v pozorováních X_{ji}, a to dokonce při všech apriorních rozděleních parametru p, které vedou ke stejným hodnotám $\mathrm{E}\,p$, $\mathrm{E}\,\mathrm{var}(S \mid p)$ a $\mathrm{var}\,\mathrm{E}(S \mid p)$, jako jsou ty v našem případě.

Neznámé parametry a a b, vyskytující se v bayesovském odhadu (3.2) týkajícím se j-tého řidiče, nyní odhadneme z dat všech řidičů. Všimněme si, že

$$\mathrm{E}\,X_{ji} = \mathrm{E}\,\mathrm{E}(X_{ji} \mid p) = \mathrm{E}\,p = \frac{a}{a+b} = p_0,$$
$$\frac{\mathrm{E}\,\mathrm{var}(X_{ji} \mid p)}{\mathrm{var}\,\mathrm{E}(X_{ji} \mid p)} = \frac{\mathrm{E}(p(1-p))}{\mathrm{var}\,p} = a+b = n_0.$$

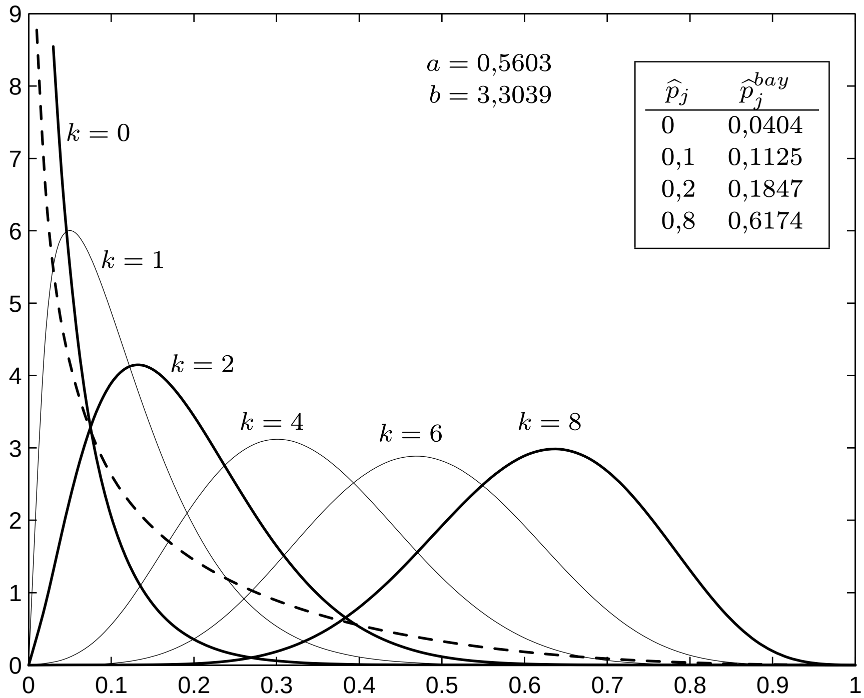
Nabízí se tedy využít zřejmých odhadů

$$\mathrm{E}\,X_{ji} = \mathrm{E}\,\mathrm{E}(X_{ji} \mid p) \approx \sum_j \widehat{p}_j/N,$$

[1]Tento model volíme s ohledem na předchozí výklad. Nepopisujeme tedy obvyklý model, kde počet nehod za určité období má u j-tého řidiče Poissonovo rozdělení s parametrem λ_j a rozdělení parametru λ mezi řidiči se řídí gama rozdělením.

$$\mathrm{E}\,\mathrm{var}(X_{ji}\,|\,p) \approx N^{-1}\sum_j\Big(\frac{n}{n-1}\widehat{p}_j(1-\widehat{p}_j)\Big),$$
$$\mathrm{var}\,\mathrm{E}(X_{ji}\,|\,p) \approx s^2_{\widehat{p}_j} - \frac{1}{nN}\sum_j\Big(\frac{n}{n-1}\widehat{p}_j(1-\widehat{p}_j)\Big),$$

kde $s^2_{\widehat{p}_j}$ označuje výběrový rozptyl mezi individuálními odhady, tj. spočtený z hodnot $\widehat{p}_1,\dots,\widehat{p}_N$. Výraz odečítaný v posledním řádku od $s^2_{\widehat{p}_j}$ upravuje odhad $\mathrm{var}\,\mathrm{E}(X_{ji}|\,p)$ tak, že i tento poslední odhad je, stejně jako předchozí dva, nestranný. Pro pozo-



$\widehat{p}_j$	$\widehat{p}_j^{bay}$
0	0,0404
0,1	0,1125
0,2	0,1847
0,8	0,6174

Obrázek 4. Apriorní hustota s odhadnutými parametry (čárkovaně) a aposteriorní hustoty pro různé počty nehodových roků $S_j = k$.

rovaná data (3.1) vychází $n_0 \approx 3{,}8643$, $p_0 \approx 0{,}1450$, a odtud $a = 0{,}5603$, $b = 3{,}3039$. Odpovídající tvary hustot parametru p_j jsou na obr. 4, hledaným odhadem (3.2) je

$$\widehat{p}_j^{bay} = 0{,}154 \cdot 0{,}145 + 0{,}846 \cdot \widehat{p}_j.$$

Odhad $\widehat{p}_j^{bay}$ tedy kombinuje spolehlivý odhad $\widehat{p} = 0{,}145$ kolektivní pravděpodobnosti p_0 s individuálním nebayesovským odhadem $\widehat{p}_j$, který byl pořízen z malého počtu pozorování z 10 roků u jednoho řidiče. Oproti $\widehat{p}_j$ je odhad $\widehat{p}_j^{bay}$ vždy posunut směrem ke kolektivnímu odhadu $\widehat{p}$.

4. Normální IQ

V předchozím příkladu jsme si při určení parametrů apriorního rozdělení museli pomoci odhady. Jindy ale kompletní informace o apriorním rozdělení vyplyne přirozeně. Nechť θ označuje inteligenčního kvocient jistého dítěte, jeho hodnotu určujeme pomocí IQ testu. Je známo, že výsledek testu X má normální rozdělení pravděpodobnosti se střední hodnotou θ a směrodatnou odchylkou 10. Dlouhodobé výzkumy ukázaly, že rozložení kvocientu u dětí příslušného věku je normální $\mathrm{N}(100, 15^2)$.

Tedy při dané hodnotě kvocientu θ má měření X normální rozdělení $\mathrm{N}(\theta, 10^2)$ s hustotou

$$f(x \mid \theta) = \frac{1}{\sqrt{2\pi \cdot 10^2}} \exp\Big(-\frac{(x-\theta)^2}{2 \cdot 10^2}\Big)$$

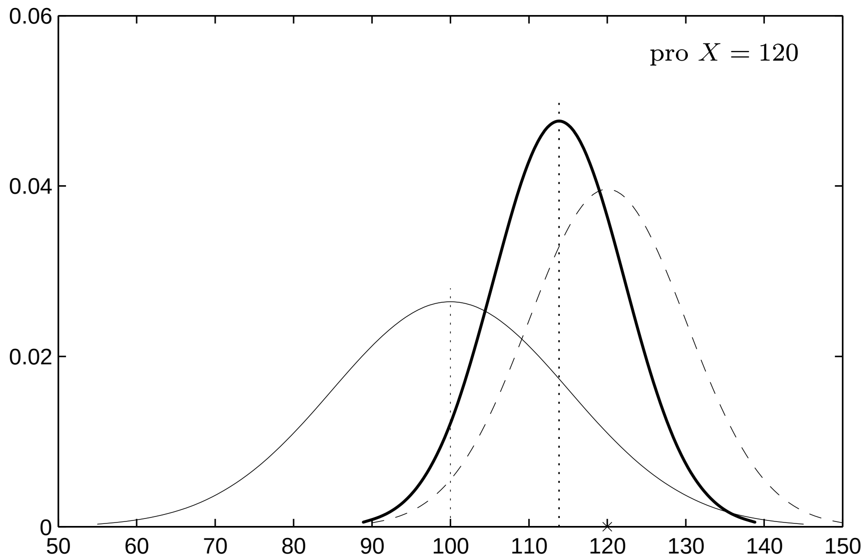
a apriorní hustota parametru θ je

$$\pi(\theta) = \frac{1}{\sqrt{2\pi \cdot 225}} \exp\Big(-\frac{(\theta-100)^2}{2 \cdot 225}\Big).$$

Aposteriorní hustotu parametru θ při daném pozorování X dostáváme podle (1.1) jako

$$\begin{aligned} \pi(\theta \mid X = x) \propto f(x \mid \theta)\pi(\theta) &\propto \exp\Big(-\frac{(x-\theta)^2}{2 \cdot 10^2} - \frac{(\theta-100)^2}{2 \cdot 225}\Big) \\ &\propto \exp\Big(-\theta^2\big(\frac{1}{10^2} + \frac{1}{225}\big) + 2\theta\big(\frac{x}{10^2} + \frac{100}{225}\big)\Big) \\ &\propto \exp\Big(-\Big(\theta - 2\big(\frac{x}{10^2} + \frac{100}{225}\big)\big(\frac{1}{10^2} + \frac{1}{225}\big)^{-1}\Big)^2 \Big/ \Big(2\big(\frac{1}{10^2} + \frac{1}{225}\big)^{-1}\Big) \\ &\sim \mathrm{N}\Big(\frac{x \cdot 225 + 100 \cdot 10^2}{225 + 10^2}, \frac{225 \cdot 10^2}{225 + 10^2}\Big) = \mathrm{N}\Big(\frac{9}{13}x + \frac{4}{13}100, \frac{900}{13}\Big). \end{aligned}$$

Na obrázku 5 je znázorněna situace v případě, že IQ test dal výsledek $X = 120$.



Obrázek 5. Apriorní hustota (tence vlevo) a aposteriorní hustota inteligenčního kvocientu θ v případě, že výsledek IQ testu je 120. Čárkovaně věrohodnostní funkce.

Bodovým odhadem inteligenčního koeficientu je jakožto střední hodnota aposteriorního rozdělení první z jeho parametrů,

$$\widehat{\theta} = \mathrm{E}(\theta \mid X) = \frac{9}{13}X + \frac{4}{13}100 \doteq 0{,}69X + 0{,}31 \cdot 100,$$

zatímco interval

$$\Big(\widehat{\theta} - u_{1-\alpha/2}\sqrt{\frac{900}{13}}, \widehat{\theta} + u_{1-\alpha/2}\sqrt{\frac{900}{13}}\Big) \doteq (\widehat{\theta} - 8{,}32 \cdot u_{1-\alpha/2}, \widehat{\theta} - 8{,}32 \cdot u_{1-\alpha/2}),$$

kde u_p označuje $100p\,\%$ kvantil standardního normálního rozdělení $\mathrm{N}(0,1)$, je bayesovskou $100(1-\alpha)\,\%$ konfidenční oblastí.

Analogickým postupem získáme aposteriorní rozdělení pro střední hodnotu normálního rozdělení v obecnějším případě s více pozorovanými hodnotami X. Jsou-li pozorování tvořena náhodným výběrem $X_1, \dots, X_n$ z normálního rozdělení $\mathrm{N}(\mu, \sigma^2)$, kde rozptyl σ^2 známe[2], pak za předpokladu, že apriorní rozdělení parametru μ je normální $\mathrm{N}(a, b^2)$, je aposteriorním rozdělením opět normální rozdělení

$$\mathrm{N}\Big(\frac{nb^2\overline{x} + \sigma^2 a}{nb^2 + \sigma^2}, \frac{\sigma^2 b^2}{nb^2 + \sigma^2}\Big) = \mathrm{N}(w\overline{x} + (1-w)a, w\sigma^2/n), \quad \text{kde} \quad w = \frac{nb^2}{nb^2 + \sigma^2}$$

a $\overline{x}$ označuje průměr pozorovaných hodnot. Aposteriorní střední hodnotu můžeme zapsat ve tvaru

$$\widehat{\mu} = w\overline{x} + (1-w)a, \quad \text{kde} \quad w = \frac{nb^2}{nb^2 + \sigma^2}, \tag{4.1}$$

a představuje bayesovský bodový odhad. Konfidenčním intervalem s pravděpodobností $100(1-\alpha)\,\%$ je

$$\Big(\widehat{\mu} - u_{1-\alpha/2}\sqrt{w}\frac{\sigma}{\sqrt{n}}, \widehat{\mu} - u_{1-\alpha/2}\sqrt{w}\frac{\sigma}{\sqrt{n}}\Big)$$

a tento interval je vzhledem k $0 < w < 1$ vždy kratší než intervalový odhad při klasické přístupu.

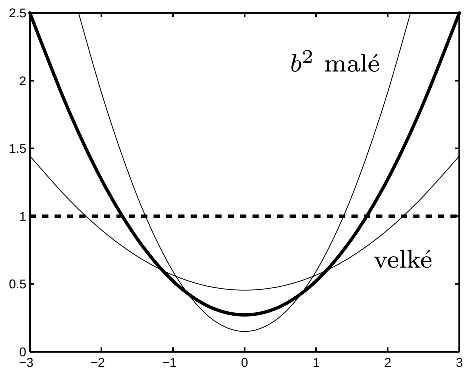
Podívejme se nyní na bayesovský odhad (4.1) z klasického hlediska. Jeho střední čtvercová chyba (MSE) je při dané hodnotě parametru μ rovna

$$\mathrm{MSE}\,\widehat{\mu} = \mathrm{E}(\widehat{\mu} - \mu)^2 = \mathrm{var}\,\widehat{\mu} + (\mathrm{E}\,\widehat{\mu} - \mu)^2 = w^2\,\mathrm{var}\,\overline{X} + (w\,\mathrm{E}\,\overline{X} + (1-w)a - \mu)^2,$$

zatímco střední čtvercová chyba klasického odhadu průměrem $\overline{X}$ je

$$\mathrm{MSE}\,\overline{X} = \mathrm{E}(\overline{X} - \mu)^2 = \mathrm{var}\,\overline{X} = \sigma^2/n \tag{4.2}$$

pro všechna $\mu \in \mathbf{R}$. Průměr $\overline{X}$ je v případě normálního rozdělení jak známo nej-



Obrázek 6. Průběh $\mathrm{MSE}\,\overline{X}$ (čárkovaně) a $\mathrm{MSE}\,\widehat{\mu}$ pro různé hodnoty b^2 v závislosti na $\mu - a$.

lepším odhadem, žádný odhad proto nemůže mít MSE lepší než (4.2) současně pro všechna μ. Přesto, podívejme se, pro která μ je u bayesovského odhadu $\mathrm{MSE}\,\widehat{\mu}$ lepší. Máme $\mathrm{MSE}\,\widehat{\mu} < \mathrm{MSE}\,\overline{X}$ právě tehdy, když

$$w^2\sigma^2/n + (1-w)^2(a-\mu)^2 < \sigma^2/n,$$

[2]Pro případ s neznámým rozptylem σ^2 viz příklad s regresí.

tj. když

$$|\mu - a| < \sqrt{2b^2 + \sigma^2/n}, \tag{4.3}$$

viz obr. 6. Víme-li při praktické aplikaci, že hodnoty parametru μ se zcela jistě budou pohybovat v nějakém omezeném intervalu, pak při vhodné volbě čísel a a b (a jako střed tohoto intervalu a b dostatečně velké, tak aby byla splněna nerovnost (4.3)) poskytuje odhad tvaru (4.1) v tomto oboru hodnot parametru μ lepší odhad (měřeno klasicky), než je odhad klasický. K této volbě a a b přitom nebyla použita žádná apriorní informace s výjimkou znalosti oboru hodnot parametru μ.

5. Regrese

Dalším zobecněním odhadu střední hodnoty normálního rozdělení je odhad parametrů v lineárním regresním modelu s normálně rozdělenými chybami. Uvažujme standardní model s k parametry $\beta = (\beta_1, \ldots, \beta_k)'$ a n pozorováními $y = (y_1, \ldots, y_n)'$, $n < k$, popsaný po složkách vztahy

$$y_i = \sum_j x_{ij}\beta_j + e_i, \quad i = 1, \ldots, n,$$

nebo maticově

$$y = X\beta + e. \tag{5.1}$$

Matici regresorů $X = (x_{ij})$ předpokládáme s plnou hodností k. Označme klasický odhad metodou nejmenších čtverců jako

$$b = (X'X)^{-1}X'y. \tag{5.2}$$

Chyby $e = (e_1, \ldots, e_n)$ předpokládáme nezávislé, normálně rozdělené s nulovou střední hodnotu a rozptyly σ^2. Pozorování jsou tedy také normálně rozdělená, y má n-rozměrné normální rozdělení $\mathrm{N}_n(X\beta, \sigma^2 I)$ s hustotou

$$f(y; \beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \mathrm{e}^{-(y-X\beta)'(y-X\beta)/2\sigma^2}.$$

Namísto parametru σ^2 počítejme nadále s převrácenou hodnotou rozptylu σ^{-2} (parametr přesnosti).

Nejprve se podíváme na případ s neinformativní apriorní hustotou tvaru

$$\pi(\beta, \sigma^{-2}) \propto 1 \cdot (\sigma^{-2})^{-1} = (\sigma^{-2})^{-1}, \quad \beta \in \mathbf{R}, \ \sigma^{-2} > 0. \tag{5.3}$$

Je to nevlastní hustota, kde složka β má rovnoměrnou hustotu a je nezávislá se σ^{-2}. Počítáme aposteriorní rozdělení:

$$\begin{aligned}
\pi(\beta, \sigma^{-2} \mid y) &= f(y \mid \beta, \sigma^{-2}) \cdot \pi(\beta, \sigma^{-2}) \\
&\propto (\sigma^{-2})^{n/2} \exp\bigl[-(y - X\beta)'(y - X\beta)\sigma^{-2}/2\bigr] \cdot (\sigma^{-2})^{-1} \\
&= (\sigma^{-2})^{k/2} \exp\bigl[-(\beta - b)'X'X(\beta - b)\sigma^{-2}/2\bigr] \qquad (5.4)\\
&\qquad \times (\sigma^{-2})^{(n-k)/2-1} \exp\bigl[-\underbrace{(y - Xb)'(y - Xb)}_{S_b}\sigma^{-2}/2\bigr], \qquad (5.5)
\end{aligned}$$

když jsme při přechodu na třetí řádek využili kolmosti vektorů $y - Xb$ a $X\beta - Xb$; S_b označuje reziduální součet čtverců odpovídající odhadu metodou nejmenších čtverců b. Toto rozdělení se označuje jako normální-gama rozdělení, protože podmíněné rozdělení vektoru β při daném σ^{-2} (viz (5.4)) je normální rozdělení a marginálním rozdělení složky σ^{-2} (viz (5.5)) je gama rozdělení, konkrétně

$$(\beta \mid \sigma^{-2}, y) \sim \mathrm{N}_k(b, \sigma^2 (X'X)^{-1}),\qquad (\sigma^{-2} \mid y) \sim \mathrm{G}(S_b/2, (n-k)/2). \tag{5.6}$$

Marginální rozdělení vektoru β v aposteriorním rozdělení dostaneme zintegrováním,

$$\pi(\beta \mid y) = \int_0^\infty \pi(\beta, \sigma^{-2} \mid y)\pi(\sigma^{-2})\, \mathrm{d}\sigma^{-2} \propto ((\beta-b)'X'X(\beta-b)+S_b)^{-n/2}.$$

Až na lineární transformaci jde o hustotu k-rozměrného t-rozdělení o $n-k$ stupních volnosti a u jednotlivých složek o (jednorozměrné) rozdělení t_{n-k}.

Bayesovský odhad vektoru β můžeme určit prostřednictvím podmíněné střední hodnoty už z (5.6) jako

$$\widehat{\beta} = \mathrm{E}(\beta \mid y) = b = (X'X)^{-1}X'y$$

a shoduje se s odhadem metodou nejmenších čtverců, podobně bayesovské konfidenční intervaly pro jednotlivé parametry β_i, resp. oblast pro celý vektor β vycházejí

$$b_i \pm t_{1-\alpha/2}(n-k)\sqrt{((X'X)^{-1})_{ii}S_b/(n-k)},$$

$$\text{resp.}\quad \Big\{\beta;\ \frac{(\beta-b)'X'X(\beta-b)/k}{S_b/(n-k)} \leqq F_{1-\alpha}(k, n-k)\Big\},$$

tedy znovu stejně jako intervaly spolehlivosti, resp. oblast standardně používané v nebayesovské statistice.

Pokud bychom normální-gama rozdělení zvolili už jako apriorní rozdělení, konkrétně

$$(\beta \mid \sigma^{-2}) \sim \mathrm{N}_k(a_0, \sigma^2 M_0^{-1}) \qquad \text{a} \qquad \sigma^{-2} \sim \mathrm{G}(S_0, n_0),$$

pak aposteriorní rozdělení je opět stejného typu s parametry

$$a_1 = M_1^{-1}(M_0 a_0 + X'Xb) = \widehat{\beta}, \quad M_1 = M_0 + X'X,$$
$$n_1 = n_0 + n, \quad S_1 = S_0 + S_b + (b-a_0)'(M_0^{-1} + (X'X)^{-1})^{-1}(b-a_0). \tag{5.7}$$

Výše uvažovaný případ s neurčitostní apriorní hustotou (5.3) odpovídá formálně hodnotám parametrů $M_0 = 0$ (resp. $M_0^{-1} = \infty$), $S_0 = 0$ a $n_0 = -k$. Samozřejmě, volíme-li počet složek vektoru β roven $k = 1$, odhad parametrů regrese se redukuje na problém odhadu střední hodnoty, jímž jsme se zabývali v předchozí části.

Úvahy o bayesovských odhadech v regresi uzavřeme příkladem, kde navíc bude potřeba ještě odhadnout neznámý parametr apriorního rozdělení. Pro jednoduchost předpokládejme, že v modelu (5.1) rozptyl σ^2 chybového členu známe, a pro regresní koeficienty β volíme apriorní rozdělení

$$\beta \sim \mathrm{N}_k(0, \tau^2(X'X)^{-1}),$$

kde τ^2 je nějaká kladná konstanta. Máme aposteriorní rozdělení $(\beta|y) \sim \mathrm{N}_k(\tau^2(\sigma^2 + \tau^2)b, \sigma^2\tau^2/(\sigma^2+\tau^2)(X'X)^{-1})$ (srov. s (5.7)) a bayesovský odhad

$$\widehat{\beta} = \mathrm{E}(\beta \mid y) = \Big(1 - \frac{\sigma^2}{\sigma^2+\tau^2}\Big)b. \tag{5.8}$$

Jestliže parametr τ apriorního rozdělení neznáme, budeme muset výraz, v němž se vyskytuje odhadnout. Odhad založíme opět na pozorováních y. Využijeme, že nepodmíněné rozdělení transformovaného vektoru b je

$$z = (X'X)^{-1/2}X'y \sim \mathrm{N}_k(0, (\sigma^2+\tau^2)I_k),$$

odkud za předpokladu $k > 2$

$$\mathrm{E}\big(z'z/(\sigma^2+\tau^2)\big)^{-1} = \mathrm{E}(1/\chi^2_k) = 1/(k-2)$$

jako moment χ^2 rozdělení veličiny $z'z/(\sigma^2+\tau^2)$. Střední hodnotu nahradíme pozorováním a odhadujeme

$$\frac{1}{\sigma^2+\tau^2} \approx \frac{k-2}{z'z}.$$

Dosazením tohoto odhadu do (5.8) dostáváme

$$\widehat{\beta}_{JS} = \Big(1 - \frac{(k-2)\sigma^2}{b'X'Xb}\Big)b,$$

což je Jamesův-Steinův odhad (původní článek Stein (1956)). Tento zkrácený odhad je z klasického hlediska lepší než odhad metodou nejmenších čtverců b a to v celém oboru hodnot parametru β — pro všechny hodnoty parametrů β je střední čtvercová chyba předpovědi $X\widehat{\beta}_{JS}$ menší než u předpovědi Xb.

6. Neparametrické bayesovské odhady

Na závěr nastíníme bayesovskou variantu neparametrického přístupu. Opět, uvažujme pozorování $X_1, \dots, X_n$ z rozdělení s distribuční funkcí $F(x)$, $x \in \mathbf{R}$, která je tentokrát kompletně neznámá. Jejím neparametrickým odhadem může být např. empirická distribuční funkce $F_e(x) = \#\{i; X_i \leqq x\}/n$.

Při bayesovském přístupu považujeme parametry, v tomto případě tedy hodnoty distribuční funkce $F(x)$, $x \in \mathbf{R}$ (parametrů tedy máme nekonečně mnoho), za náhodné veličiny, tj. celou distribuční funkci F za náhodný proces indexovaný reálným čísly. Apriorní rozdělení tohoto procesu, což je rozdělení na množině všech distribučních funkcí, vyjadřuje apriorní informaci.

Asi nejznámějším apriorním procesem je Dirichletův proces (Ferguson (1973)). Nechť F_0 je nějaká distribuční funkce a $n_0 > 0$. Proces F je Dirichletův s parametrem n_0F_0, píšeme $F \sim \mathcal{D}(n_0F_0)$, jestliže pro každé dělení $-\infty = t_0 < \dots < t_k = \infty$ reálné osy je

$$(U_1, \dots, U_k) \sim \mathrm{D}(a_1, \dots, a_k),$$

kde $U_i = F(t_i) - F(t_{i-1})$ a $a_i = n_0F(t_i) - n_0F(t_{i-1})$ označují přírůstky procesu F, resp. distribuční funkce F_0 na intervalech vzniklých dělením a $\mathrm{D}(a_1, \dots, a_k)$ označuje k-rozměrné Dirichletovo rozdělení s hustotou

$$\pi(u_1, \dots, u_k) = \frac{\Gamma(a_1+\dots+a_k)}{\Gamma(a_1)\cdot\dots\cdot\Gamma(a_k)} u_1^{a_1-1}\cdot\dots\cdot u_1^{a_k-1}, \quad u_i \geqq 0, \quad \sum u_i = 1.$$

Dirichletovo rozdělení je vícerozměrnou obdobou beta rozdělení, složky mají beta rozdělení, $U_i \sim \mathrm{B}(a_i, n_0 - a_i)$. Při apriorním rozdělení $F \sim \mathcal{D}(n_0F_0)$ je $\mathrm{E}\,F(t) = F_0(t)$ a $\operatorname{var} F(t) = F_0(t)(1-F_0(t))/(n_0+1)$, z parametrů apriorního Dirichletova procesu F_0 představuje centrální hodnotu F a n_0 popisuje míru koncentrace apriorního rozdělení kolem ní — větší hodnoty n_0 odpovídají soustředěnější apriorní informaci. Aposteriorní rozdělení parametru F je

$$(F \mid X_1, \dots, X_n) \sim \mathcal{D}(n_0F_0 + nF_e),$$

tedy opět Dirichletův proces. Bayesovským odhadem, tj. střední hodnotou aposteriorního rozdělení, je

$$\widehat{F}(t) = \frac{n_0 F_0(t) + n F_e(t)}{n_0 + n}.$$

Tento odhad je váženým průměrem mezi středem apriorního rozdělení F_0 a klasickým neparametrickým odhadem pomocí empirické distribuční funkce F_e. Složky v průměru jsou zastoupeny v poměru apriorní přesnosti n_0 a rozsahu pozorování n.

Ne vždy máme k dispozici úplné pozorování náhodného výběru. Např. v analýze dat o přežití, kde X_i představují doby života jedinců nebo třeba doby do poruchy zařízení, jsou často pozorování cenzorována zprava (sledovaný jedinec se odstěhuje, nebo zemře z jiné než sledované příčiny, čekání na poruchu je z časových důvodů ukončeno). I cenzorovaná pozorování ale obsahují informaci o sledované veličině. Jako ukázku zvolme náhodné cenzorování, kdy s každou veličinou X_i je svázána ještě další náhodná veličina Y_i, časový cenzor, a i-té pozorování je ukončeno v okamžiku X_i, nebo v okamžiku cenzorování Y_i, podle toho, který nastane dříve. V případě ukončení cenzorování máme alespoň částečnou informaci $X_i \geqq Y_i$. Pozorování můžeme shrnout pomocí veličin

$$Z_i = \min(X_i, Y_i) \quad \text{a} \quad \delta_i = I_{[X_i \leqq Y_i]} = \begin{cases} 1, & X_i \leqq Y_i \\ 0, & X_i > Y_i \end{cases}.$$

Ukazuje se (Susarla and Van Ryzin (1976)), že bayesovským odhadem funkce přežití $S(t) = 1 - F(t)$ při daných hodnotách cenzorů $Y_1, \dots, Y_n$, je-li apriorně $F \sim \mathcal{D}(n_0 F_0)$, je

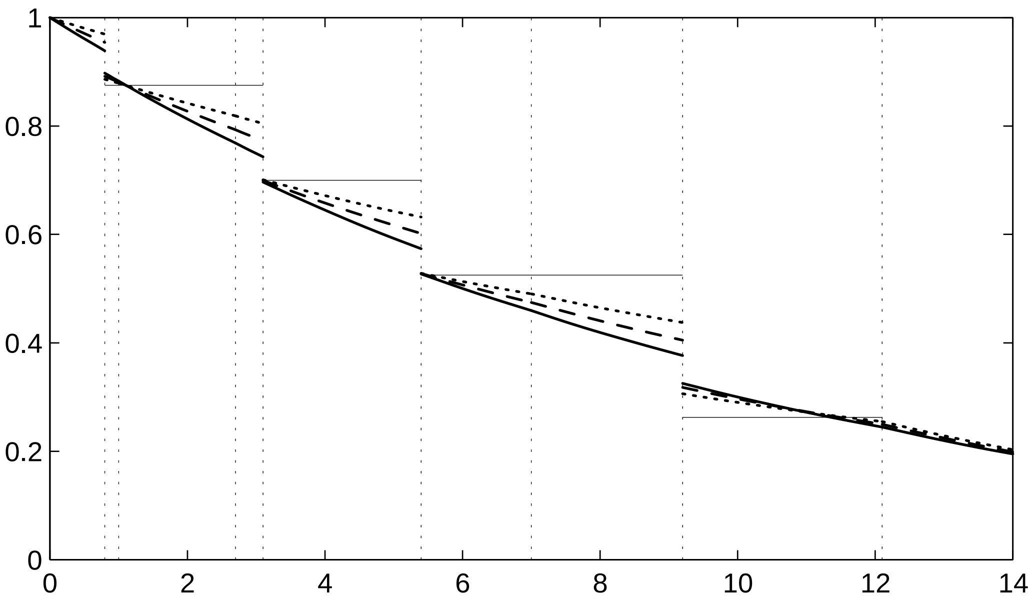
$$\widehat{S}(t) = \frac{n_0 S_0(t) + N_t}{n_0 + n} \prod_s \frac{n_0 S_0(s-) + N_{s-} - u(s)}{n_0 S_0(s-) + N_s}, \quad t > 0, \tag{6.1}$$

kde $S_0 = 1 - F_0$, součin probíhá přes okamžiky s cenzorovanými pozorováními $\{s, \exists_i\, Z_i \leqq t, \delta_i = 0\}$, dále $N_s = \#\{i; Z_i \leqq s\}$ označuje počet pozorování nepřekračujících s a $u(s)$ je počet necenzorovaných pozorování v okamžiku s.

Obrázek 7 zobrazuje tvar odhadu (6.1) na příkladu miniaturních dat uvedených v článku Kaplan and Meier (1958). Data tvoří necenzorovaná pozorování s časy 0,8, 3,1, 5,4, 9,2 a pozorování cenzorovaná v časech 1,0, 2,7, 7,0 12,1. Jako centrální distribuční funkce F_0 apriorního Dirichletova procesu je zvoleno exponenciální rozdělení, $1 - F_0(x) = \mathrm{e}^{-0{,}12x}$, $x > 0$, a jako n_0 postupně 4, 8 a 16. Odhad má v okamžicích s necenzorovanými pozorováními skoky, v okamžicích cenzorovaných pozorování jen není hladký, derivace zleva a zprava se liší. S rostoucí neurčitostí apriorní informace ($n_0 \to 0$) se odhad (6.1) blíží k neparametrickému Kaplanovu-Meierovu odhadu.

Literatura

1. Berger, J. O., *Statistical decision theory and Bayesian analysis*, 2. ed., Springer, New York, 1985.
2. Ferguson, T. S., *A Bayesian analysis of some nonparametric problems*, Ann. Statist. **1** (1973), no. 2, 209–230.
3. Kaplan, E. L., and Meier, P., *Nonparametric estimation from incomplete estimation*, J. Amer. Statist. Assoc. **53** (1958), 457–481.
4. Hušková, M., *Bayesovské metody*, skripta, Univerzita Karlova, 1985.
5. Stein, C., *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I, Univ. of California Press, Berkeley, 1956, pp. 197–206.



Obrázek 7. Neparametrický bayesovský odhad funkce přežití (pro $n_0 = 4$ tečkovaně, $n_0 = 8$ čárkovaně, $n_0 = 16$ plně) a Kaplanův-Meierův odhad (po částech konstantní, tence). Svisle naznačeny okamžiky pozorování.

6. Susarla, V., and Van Ryzin, J., *Nonparametric Bayesian estimation of survival curves from incomplete observations*, J. Amer. Statist. Assoc. **71** (1976), no. 356, 897–902.