Kapitola 4

Jádrové odhady distribuční funkce

Výstupy z výukové jednotky

Student

- bude znát základní typy jádrových odhadů distribuční funkce a jejich statistické vlastnosti.
- získá přehled o metodách pro volbu vyhlazovacího parametru.
- bude schopen navrhnout a implementovat proceduru pro zpracování reálných dat.
- se naučí používat příslušný toolbox v Matlabu a dokáže zkonstruovat jádrový odhad distribuční funkce pro daná reálná data.

1 Motivace

Distribuční funkce popisuje rozložení pravděpodobnosti náhodné veličiny (budeme předpokládat spojitost náhodné veličiny). Stejně jako při rekonstrukci hustoty z množiny pozorovaných dat lze distribuční funkci odhadnout parametrickými nebo neparametrickými metodami. Zaměříme se výhradně na neparametrické metody, kdy předpokládáme pouze jistou hladkost odhadované distribuční funkce.

Nejužívanějším neparametrickým odhadem distribuční funkce F je empirická distribuční funkce F_n. Ovšem F_n je schodovitá funkce i v případě, že F je spojitá. Nadaraya (1964) navrhl „hladkou" alternativu k F, a to jádrový odhad $\widehat{F}$, který se získá integrací známého jádrového odhadu hustoty (3.1)

2 Základní typy neparametrických odhadů

Nechť $X_1, \ldots, X_n$ jsou nezávislé náhodné proměnné, které mají tutéž spojitou hustotu f a distribuční funkci F. Nejjednodušší neparametrický dohad distribuční funkce F je *empirická distribuční funkce* $\widehat{F}_n$ definovaná v bodě x vztahem

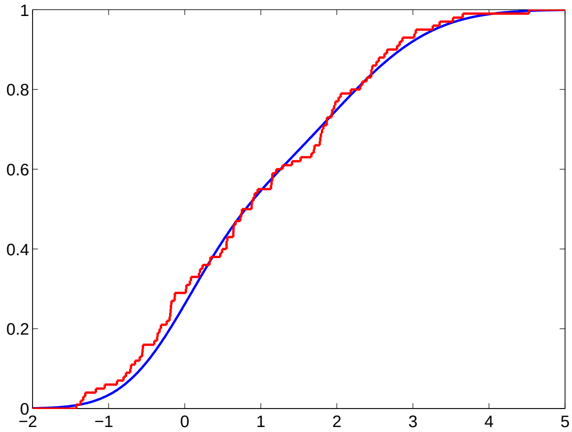
$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty,x]}(X_i).$$

Tento odhad má sice dobré statistické vlastnosti, ale je to schodovitá funkce (viz obr. 4.1, a proto se budeme zabývat postupy, které umožní zkonstruovat „hladký" odhad distribuční funkce F.

Příklad 2.1. Mějme dán náhodný výběr o velikosti $n = 100$ ze směsi dvou normálních hustot $N(0; 4/9)$ a $N(2; 1)$ s hustotou

$$f(x) = 0{,}5\frac{3}{2\sqrt{2\pi}}\,\mathrm{e}^{-\frac{9x^2}{8}} + 0{,}5\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{(x-2)^2}{2}}.$$

(Data jsou v tabulce 6.3.) Z obrázku 4.1 je patrné, že schodovitá funkce nevystihuje plně charakter distribuční funkce.



Obrázek 4.1: Empirická distribuční funkce (červeně) a skutečná distribuční funkce (modře) pro data z příkladu 2.1

Nejznámější postup spočívá v integraci jádrového odhadu hustoty, t.j.

$$\widehat{F}(x,h) = \int_{-\infty}^{x} \hat{f}(t,h)\,\mathrm{d}t = \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{x} K\left(\frac{t-X_i}{h}\right)\mathrm{d}t.$$

Užijeme-li substituce $y = (t - X_i)/h$, dostaneme

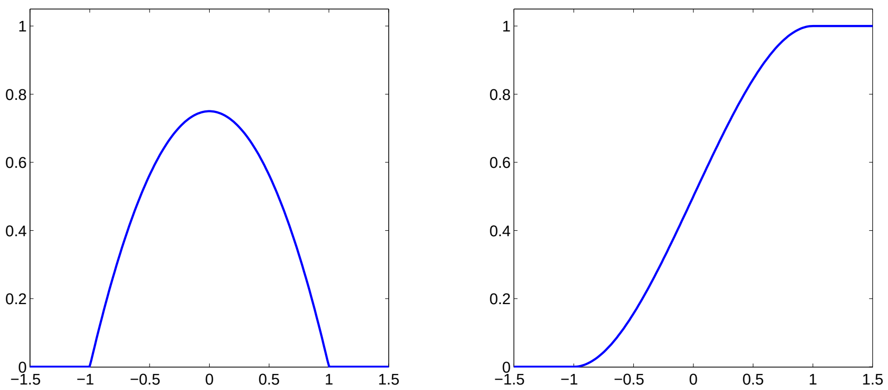
$$\widehat{F}(x,h) = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\frac{x-X_i}{h}} K(y)\,\mathrm{d}y = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{x-X_i}{h}\right).$$

To znamená, že odhad F v bodě $x \in \mathbb{R}$ je definován takto

$$\widehat{F}(x,h) = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{x-X_i}{h}\right), \quad W(x) = \int_{-\infty}^{x} K(t)\,\mathrm{d}t. \tag{4.1}$$

Zde předpokládáme, že $K \in S_{02}$, $K(x) \geq 0$ pro $x \in [-1, 1]$. Níže jsou shrnuty základní vlastnosti funkce W:

1. $W(x) = 0$ pro $x \in (-\infty, -1]$ a $W(x) = 1$ pro $x \in [1, \infty)$,
2. $\int_{-1}^{1} W^2(x)\,\mathrm{d}x \leq \int_{-1}^{1} W(x)\,\mathrm{d}x = 1$,
3. $\int_{-1}^{1} W(x)K(x)\,\mathrm{d}x = \frac{1}{2}$,
4. $\int_{-1}^{1} xW(x)K(x)\,\mathrm{d}x = \frac{1}{2}\left(1 - \int_{-1}^{1} W^2(x)\,\mathrm{d}x\right)$.

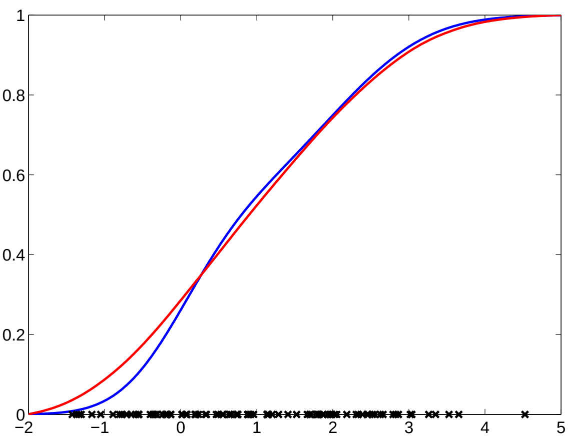


Obrázek 4.2: Epanečnikovo jádro K (vlevo) a k němu příslušná funkce W (vpravo)

Příklad 2.2. Použijeme-li Epanečnikovo jádro $K(x) = \frac{3}{4}(1 - x^2)I_{[-1,1]}(x)$, pak funkce W je tvaru

$$W(x) = \begin{cases} 0 & x \leq -1, \\ \frac{1}{4}(-x^3 + 3x + 2) & |x| < 1, \\ 1 & x \geq 1. \end{cases}$$

Pro data z příkladu 2.1 je jádrový odhad distribuční funkce zachycen na obrázku 4.3.



Obrázek 4.3: Jádrový odhad distribuční funkce s parametrem $h = 1{,}5$

3 Statistické vlastnosti odhadu

Kvalitu jádrového odhadu lze lokálně popsat pomocí střední kvadratické chyby MSE:

$$\begin{aligned} \mathrm{MSE}\,\widehat{F}(x,h) &= E(\widehat{F}(x,h) - F(x))^2 \\ &= \underbrace{(E\widehat{F}(x,h) - F(x))^2}_{\mathrm{bias}^2} + \underbrace{E(\widehat{F}(x,h))^2 - (E\widehat{F}(x,h))^2}_{\mathrm{var}}. \end{aligned}$$

Spočítejme nejdříve hodnotu $E\widehat{F}(x,h)$ v bodě $x \in \mathbb{R}$:

$$
\begin{aligned}
E\widehat{F}(x,h) &= \int W\left(\frac{x-y}{h}\right) f(y)\,\mathrm{d}y \\
&= h\int_{-\infty}^{1} W(t)f(x-ht)\,\mathrm{d}t + h\int_{1}^{\infty} W(t)f(x-ht)\,\mathrm{d}t.
\end{aligned}
$$

Předpokládejme dále, že $F \in C^2$. Označme první integrál I_1 a druhý I_2. Integrál I_1 počítáme metodou per partes a využijeme vlastnosti funkce $W(t)$

$$
\begin{aligned}
I_1 &= h\int_{-\infty}^{1} W(t)f(x-ht)\,\mathrm{d}t \\
&= \left| \begin{array}{ll} u = W(t) & u' = W'(t) = K(t) \\ v' = f(x-ht)h & v = -F(x-ht) \end{array} \right| \\
&= [-F(x-ht)]_{-1}^{1} + \int_{-1}^{1} F(x-ht)W'(t)\,\mathrm{d}t \\
&= -F(x-h) + \int_{-1}^{1} K(t)F(x-ht)\,\mathrm{d}t. \qquad (4.2)
\end{aligned}
$$

Dále použijeme Taylorův rozvoj

$$F(x-ht) = F(x) - htF'(x) + \frac{h^2t^2}{2}F''(x) + o(h^2),$$

tedy

$$I_1 = -F(x-h) + F(x) + \frac{1}{2}F''(x)h^2\beta_2(K) + o(h^2).$$

Počítejme nyní integrál I_2:

$$I_2 = h\int_{1}^{\infty} W(t)f(x-ht)\,\mathrm{d}t.$$

Uvažujeme-li substituce $x - ht = z$, dostaneme

$$I_2 = -\int_{x-h}^{-\infty} f(z)\,\mathrm{d}z = \int_{-infty}^{x-h} f(z)\,\mathrm{d}z = F(x-h).$$

Vychýlení odhadu je tedy tvaru

$$\operatorname{bias}\widehat{F}(x,h) = \frac{1}{2}F''(x)h^2\beta_2(K) + o(h^2).$$

Poznámka 3.1. Vztahy (4.2) a (**??**) dávají zajímavý vztah pro vychýlení

$$E\widehat{F}(x,h) - F(x) = \int_{-1}^{1} K(t)F(x-ht)\,\mathrm{d}t - F(x).$$

Odtud plyne

$$E\widehat{F}(x,h) = \int_{-1}^{1} K(t)F(x-ht)\,\mathrm{d}t$$

a také (z Taylorova vzorce)

$$E\widehat{F}(x,h) = F(x) + o(h).$$

Nyní dokážeme tvar rozptylu.

$$\operatorname{var} \widehat{F}(x,h) = \frac{1}{n}\left(EW^2\left(\frac{x-X}{h}\right) - E^2W\left(\frac{x-X}{h}\right)\right).$$

Zde $E^2W\left(\frac{x-X}{h}\right) = \left(EW\left(\frac{x-X}{h}\right)\right)^2 = (F(x)+o(h))^2$. Počítáme tedy pouze integrál I_3:

$$\begin{aligned} I_3 &= \frac{1}{n}\int_{-\infty}^{\infty} W^2\left(\frac{x-y}{h}\right) f(y)\,\mathrm{d}y \\ &= |\text{substituce: } x-y=th| \\ &= \frac{1}{n}\left(\int_{-\infty}^{1} W^2(t) f(x-ht)\,\mathrm{d}t + \underbrace{h\int_{1}^{\infty} f(x-ht)\,\mathrm{d}t}_{=F(x-h)}\right). \end{aligned}$$

První integrál počítáme metodou per partes a máme

$$\begin{aligned} I_3 &= \frac{1}{n}\left[-F(x-ht)W^2(t)\right]_{-1}^{1} + \frac{2}{n}\int F(x-ht)W(t)W'(t)\,\mathrm{d}t + \frac{1}{n}F(x-h) \\ &= -\frac{1}{n}F(x-h) + \frac{2}{n}\int F(x-ht)W(t)K(t)\,\mathrm{d}t + \frac{1}{n}F(x-h) \end{aligned}$$

použijeme nyní Taylorův rozvoj

$$\begin{aligned} &= \frac{2}{n}\int W(t)K(t)\left(F(x) - htF'(x) + o(h)\right)\mathrm{d}t \\ &= \frac{2}{n}F(x)\int_{-1}^{1} W(t)K(t)\,\mathrm{d}t - \frac{2}{n}hF'(x)\int_{-1}^{1} tW(t)K(t)\,\mathrm{d}t + o\left(\frac{h}{n}\right) \end{aligned}$$

užitím vlastností funkce W a $F'(x) = f(x)$ dostaneme

$$= \frac{1}{n}\left[F(x) - hf(x)\left(1 - \int_{-1}^{1} W^2(t)\,\mathrm{d}t\right)\right] + o\left(\frac{h}{n}\right).$$

Rozptyl je tedy tvaru

$$\begin{aligned} \operatorname{var} \widehat{F}(x,h) &= \frac{1}{n}\left[F(x) - hf(x)\left(1 - \int_{-1}^{1} W^2(t)\,\mathrm{d}t\right)\right] + o\left(\frac{h}{n}\right) - (F(x)+o(h))^2 \\ &= \frac{1}{n}F(x)(1-F(x)) - \frac{h}{n}f(x)\left(1 - \int_{-1}^{1} W^2(t)\,\mathrm{d}t\right) + o\left(\frac{h}{n}\right). \end{aligned}$$

Výše uvedené výsledky můžeme nyní zformulovat v následující větě:

Věta 3.1. *Nechť $F \in C^2$, $h \to 0, nh \to \infty$ pro $n \to \infty$. Pak*

$$\operatorname{MSE} \widehat{F}(x,h) = \frac{1}{n}F(x)(1-F(x)) - \frac{1}{n}hf(x)\left(1 - \int_{-1}^{1} W^2(t)\,\mathrm{d}t\right) + \frac{1}{4}(F''(x))^2 h^4 \beta_2^2(K) + o\left(\frac{h}{n} + h^4\right). \tag{4.3}$$

Globální pohled na kvalitu odhadu lze získat prostřednictvím střední integrální kvadratické chyby (MISE).

Věta 3.2. *Nechť* $F \in C^2$, $V(F'') = \int (F''(x))^2 \,\mathrm{d}x < \infty$, $K \in S_{02}$, $\lim_{n\to\infty} h = 0$ *a* $\lim_{n\to\infty} nh = \infty$. *Pak*

$$\mathrm{MISE}\,\widehat{F}(\cdot,h) = \frac{1}{n}\int F(x)(1-F(x))\,\mathrm{d}x - c_1\frac{h}{n} + c_2h^4 + o\left(\frac{h}{n}+h^4\right), \tag{4.4}$$

kde

$$c_1 = 1 - \int_{-1}^{1} W^2(t)\,\mathrm{d}t, \quad c_2 = \frac{1}{4}\beta_2^2(K)V(F'').$$

Naším cílem je nalézt takovou hodnotu vyhlazovacího parametru, pro kterou bude MISE nabývat minimální hodnoty. Ale uvedený tvar MISE není pro takovou analýzu vhodný, a proto (stejně jako při odhadu hustoty a regresní funkce) budeme uvažovat asymptotickou střední integrální kvadratickou chybu AMISE, která v tomto případě je tvaru:

$$\mathrm{AMISE}\,\widehat{F}(\cdot,h) = \mathrm{AMISE}(h) = \underbrace{\frac{1}{n}\int F(x)(1-F(x))\,\mathrm{d}x - c_1\frac{h}{n}}_{\mathrm{AIV}} + \underbrace{c_2h^4}_{\mathrm{AISB}}. \tag{4.5}$$

Nyní už lze standardními metodami matematické analýzy nalézt takovou hodnotu h, pro kterou $\mathrm{AMISE}(h)$ nabývá minimální hodnoty. Je snadné ukázat, že

$$h_{opt,0,2} = n^{-1/3}\left(\frac{c_1}{4c_2}\right)^{1/3} = O(n^{-1/3}) \tag{4.6}$$

a pak

$$\mathrm{AMISE}(h_{opt,0,2}) = \frac{1}{n}\int F(x)(1-F(x))\,\mathrm{d}x - \frac{3}{c_2^{1/3}}\left(\frac{c_1}{4}\right)^{4/3} n^{-4/3}. \tag{4.7}$$

Poznámka 3.2. Optimální hodnota vyhlazovacího parametru pro odhad distribuční funkce je řádu $n^{-1/3}$, zatímco pro odhad hustoty s jádrem $K \in S_{02}$ je vyhlazovací parametr řádu $n^{-1/5}$.

4 Volba jádra

I v tomto případě je volba jádra méně důležitá než volba vyhlazovacího parametru. Lze doporučit jádra třídy S_{02}, např.

- Epanečnikovo $K(x) = \frac{3}{4}(1-x^2)I_{[-1,1]}(x)$,
- kvartické $K(x) = \frac{15}{16}(1-x^2)^2 I_{[-1,1]}(x)$,
- triweight $K(x) = \frac{35}{32}(1-x^2)^3 I_{[-1,1]}(x)$.

5 Volba vyhlazovacího parametru

5.1 Metody křížového ověřování

Metody křížového ověřování patří k nejužívanějším metodám pro volbu vyhlazovacího parametru.

Zde uvedeme pouze metodu navrženou A. Bowmanem (1998). Funkce křížového ověřování je v tomto případě tvaru

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n}\int \left(I_{(-\infty,x]}(X_i) - \widehat{F}_{-i}(x,h)\right)^2 \mathrm{d}x,$$

kde $\widehat{F}_{-i}(x,h)$ je jádrový odhad distribuční funkce s vynecháním bodu X_i. Pak

$$h_{\mathrm{CV}} = \arg\min_{h\in H_n} \mathrm{CV}(h),$$

přičemž $H_n = [an^{-1/3}, bn^{-1/3}]$ pro vhodná $0 < a < b < \infty$.

5.2 Princip maximálního vyhlazení

Myšlenka této metody je stejná jako pro odhad hustoty. Užijeme-li faktu, že

$$\int (F''(x))^2 \, \mathrm{d}x = \int (f'(x))^2 \, \mathrm{d}x,$$

můžeme aplikovat Terrelovu větu 5.1 pro $k = 1$. V tomto případě je

$$g_1(x) = \frac{15}{16}(1 - x^2)^2 I_{[-1,1]}(x),$$

a tedy

$$h_{opt,0,2} = n^{-1/3} \left(\frac{c_1}{\beta_2^2(K) V(f')} \right)^{1/3} \leq n^{-1/3} \left(\frac{c_1}{\beta_2^2(K)} \right)^{1/3} \frac{\sigma}{\sigma_1} V(g_1')^{-1/3},$$

kde $\sigma_1 = \int x^2 g_1(x) \, \mathrm{d}x = \frac{1}{7}$, $V(g_1') = \frac{15}{7}$. Odtud plyne, že

$$h_{\mathrm{MS}} = n^{-1/3} \left(\frac{7c_1}{15\beta_2^2(K)} \right)^{1/3} \sqrt{7}\widehat{\sigma}, \tag{4.8}$$

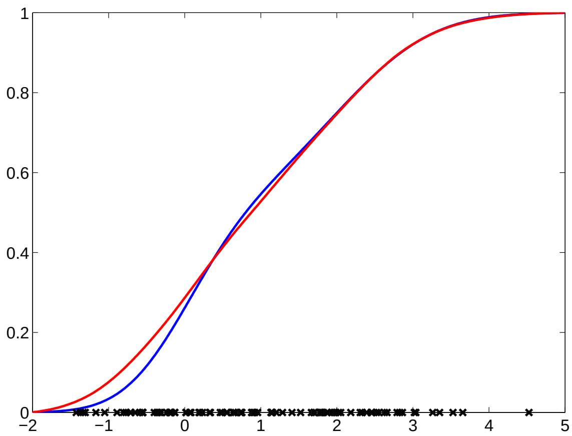
$\widehat{\sigma}$ je odhadem σ (viz rovnice (3.11) a (3.12)).

Hodnota h_{MS} může sloužit jako horní hranice pro množinu vyhlazovacích parametrů volených podle metody křížového ověřování. Tedy $H_n = [h_\ell, h_{\mathrm{MS}}]$, kde h_ℓ je nejmenší vzdálenost mezi po sobě jdoucími body $X_i, i = 1, \dots, n$.

Příklad 5.1. Pro data z příkladu 2.1 zvolme Epanečnikovo jádro. Pak hodnoty potřebné pro odhad vyhlazovacího parametru metodou maximálního vyhlazení jsou následující:

$$n = 100, \qquad \widehat{\sigma} = 1{,}3426, \qquad \beta_2(K) = \frac{1}{5}, \qquad c_1 = 1 - \int_{-1}^{1} W^2(x) \, \mathrm{d}x = 0{,}2571.$$

Pak platí $h_{\mathrm{MS}} = 1{,}1037$ a na obrázku 4.4 je zobrazen odhad distribuční funkce.



Obrázek 4.4: Odhad distribuční funkce s $h_{\mathrm{MS}} = 1{,}1037$, odhad (—), původní funkce (—)

5.3 Plug-in metoda

Společným cílem metod typu plug-in (PI) je odhadnout $V(F'')$. Za předpokladu dostatečné hladkosti funkce f užitím metody per partes dostaneme vztah

$$V(F'') = \int (F''(x))^2 \, \mathrm{d}x = - \int f''(x) f(x) \, \mathrm{d}x.$$

Tudíž se budeme dále zabývat odhadem funkcionálu

$$\psi_1 = \int f''(x) f(x)\, \mathrm{d}x.$$

Je zřejmé, že $\psi_1 = Ef''(X)$, což vede k metodě založené na odhadu druhé derivace hustoty f. Vztah (3.7) použijeme k odhadu druhé derivace s jádrem $K^{(2)} = K_{opt,2,4} \in S_{24}$. Pak

$$\widehat{\psi}_1 = n^{-1} \sum_{i=1}^{n} \hat{f}''(X_i, h) = n^{-2} h^{-3} \sum_{i=1}^{n} \sum_{j=1}^{n} K^{(2)} \left(\frac{X_i - X_j}{h} \right),$$

kde podle vztahu (3.9) je

$$h_{opt,2,4} = 10^{1/9} \frac{\delta_{24}}{\delta_{04}} h_{opt,0,4}.$$

Pak

$$\widehat{c_2} = -\frac{1}{4} \beta_2^2(K) \widehat{\psi}_1.$$

Shrnutím předchozích úvah dostaneme proceduru pro odhad distribuční funkce F:

Krok 1 Najděte optimální vyhlazovací parametr $\hat{h}_{opt,0,4}$ pro odhad hustoty s optimálním jádrem $K_{opt,0,4} \in S_{04}$.

Krok 2 Najděte optimální vyhlazovací parametr $\hat{h}_{opt,2,4}$ pro odhad druhé derivace hustoty podle vztahu (3.9) s $k = 4$ a optimálním jádrem $K^{(2)} \in S_{24}$.

Krok 3 Vypočtěte odhad funkcionálu $\widehat{\psi}_1$ s využitím hodnoty $\hat{h}_{opt,2,4}$ získané v kroku 2.

Krok 4 Vyčíslete optimální hodnotu vyhlazovacího parametru

$$h_{\mathrm{PI}} = n^{-1/3} \left(\frac{c_1}{-\widehat{\psi}_1 \beta_2^2(K)} \right)^{1/3}$$

Krok 5 Použijte parametry z předchozích kroků ke konstrukci optimálního jádrového odhadu distribuční funkce $\widehat{F}(x, h)$ s daným jádrem $K \in S_{02}$.

Příklad 5.2. S použitím funkce toolboxu zjistíme, že pro data z příkladu 2.1 je vyhlazovací parametr určený plug-in metodou roven $h_{\mathrm{PI}} = 0{,}5717$. Na obrázku 4.5 je odhad distribuční funkce společně se skutečnou distribuční funkcí.

6 Aplikace na reálná data

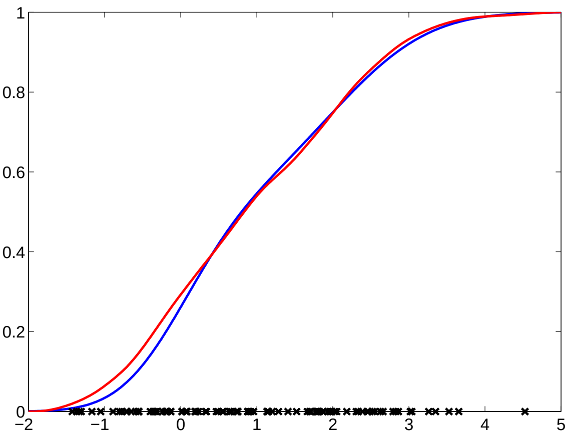
Datový soubor pochází z rozsáhlé studie, v níž autoři studovali vliv substituentů v 2,4-diamino-5-(substituovaný benzyl)pyrimidinech. Biologická aktivita při inhibici dihydrofolát reduktázy byla měřena pomocí asociační konstanty. Data jsou v tabulce 6.8 a jsou dostupná na osobních stránkách Dennise D. Boose[1], kde je také odkaz na původní článek Jonathana D. Hirsta z roku 1994.

Užitím výše uvedených metod jsme (při použití Epanečnikova jádra) dostali následující hodnoty vyhlazovacích parametrů:

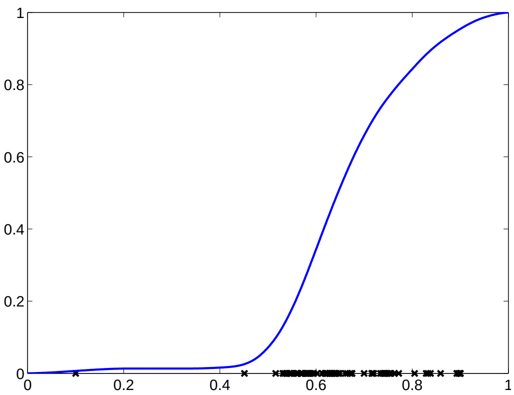
$$h_{\mathrm{MS}} = 0{,}1139, \quad h_{\mathrm{PI}} = 0{,}1931.$$

Na obrázku 4.6 jsou uvedeny odhady distribuční funkce s těmito parametry a také je zde pro srovnání uvedena empirická distribuční funkce.

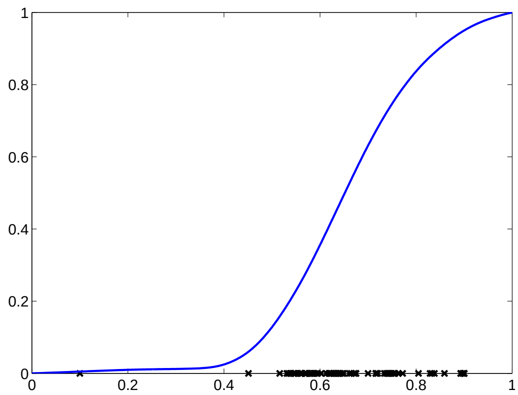
[1] http://www4.stat.ncsu.edu/~boos/var.select/pyrimidine.html



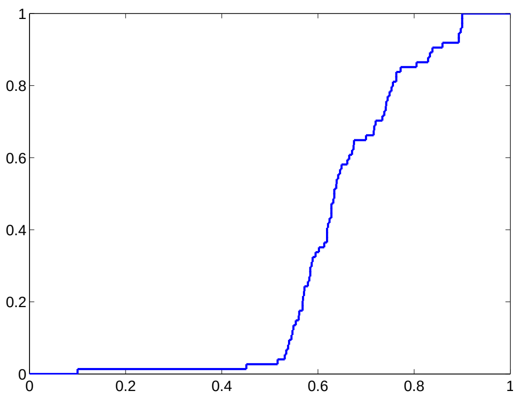
Obrázek 4.5: Odhad distribuční funkce s $h_{\mathrm{PI}} = 0{,}5717$, odhad (—), původní funkce (—)



(a) MS



(b) PI



(c) Empirická distribuční funkce

Obrázek 4.6: Odhadnuté distribuční funkce

Shrnutí

Odhad distribuční funkce $F(x)$ v bodě x je tvaru

$$\widehat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{x - X_i}{h}\right), \quad W(x) = \int_{-\infty}^{x} K(t)\,\mathrm{d}t.$$

Asymptotická střední kvadratická chyba jádrového odhadu distribuční funkce je součtem asymptotického tvaru rozptylu (AIV) a druhé mocniny vychýlení (AISB)

$$\mathrm{AMISE}(h) = \underbrace{\frac{1}{n}\int F(x)(1 - F(x))\,\mathrm{d}x - c_1\frac{h}{n}}_{\mathrm{AIV}} + \underbrace{c_2 h^4}_{\mathrm{AISB}},$$

kde

$$c_1 = 1 - \int_{-1}^{1} W^2(t)\,\mathrm{d}t, \quad c_2 = \frac{1}{4}\beta_2^2(K)V(F'').$$

Optimální vyhlazovací parametr vzhledem k AMISE pro odhad distribuční funkce je tvaru

$$h_{opt,0,2} = n^{-1/3}\left(\frac{c_1}{4c_2}\right)^{1/3},$$

t.j. $h_{opt,0,2} = O(n^{-1/3})$.

Metody pro odhad optimální hodnoty vyhlazovacího parametru h

- metoda křížového ověřování $h_{\mathrm{CV}} = \arg\min_{h\in H_n} \mathrm{CV}(h)$

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n}\int \left(I_{(-\infty,x]}(X_i) - \widehat{F}_{-i}(x,h)\right)^2 \mathrm{d}x,$$

- metoda maximálního vyhlazení

$$h_{\mathrm{MS}} = n^{-1/3}\left(\frac{7c_1}{15\beta_2^2(K)}\right)^{1/3}\sqrt{7}\widehat{\sigma},$$

- plug-in metoda

$$h_{\mathrm{PI}} = n^{-1/3}\left(\frac{c_1}{-\widehat{\psi}_1\beta_2^2(K)}\right)^{1/3}.$$

Dodatky a cvičení

1. Odvoďte tvar funkce $W(x)$ pro kvartické jádro $K(x) = \frac{15}{16}(1 - x^2)^2 I_{[-1,1]}(x)$.
2. Dokažte vlastnosti 2, 3 a 4 funkce W.
3. Dokažte vztahy (4.6) a (4.7).
4. Odvoďte tvar vyhlazovacího parametru podle metody maximálního vyhlazení pro Epanečnikovo jádro.

5. Odvoďte tvar vyhlazovacího parametru podle plug-in metody pro Epanečnikovo a pro kvartické jádro.

6. Aplikujte metodu maximálního vyhlazení a plug-in metodu na simulovaná i reálná data.