ハザード関数のカーネル型推定量の漸近表現とその応用
Asymptotic representation of a kernel type hazard function estimator and its application

Yoshihiko MAESONO　Faculty of Mathematics, Kyushu University

1　序

$X_1, \cdots, X_n$ を分布関数 $F(x)$ からの無作為標本とする．データは生存時間を表すとすると，点 x_0 でのハザード関数は

$$\frac{f(x_0)}{1-F(x_0)}$$

と定義される．ここで $f(x)$ は $F(x)$ の密度関数である．このハザード関数の推定量としては密度関数と分布関数を様々な方法で推定し代入したものが提案されている．本講演ではノンパラメトリック法な方法の１つであるカーネル型推定量を利用した下記の推定量を考察する．$K(u)$ をカーネルとするとき

$$\begin{aligned}\widehat{f}_n(x_0) &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}K\left(\frac{x_0-X_i}{h_n}\right),\\ \widehat{F}_n(x_0) &= \frac{1}{n}\sum_{i=1}^{n}W\left(\frac{x_0-X_i}{h_n}\right)\end{aligned}$$

がそれぞれの密度及び分布関数の推定量である．ただし

$$W(t)=\int_{-\infty}^{t}K(u)du$$

であり，h_n はバンド幅で $h_n \to 0$ とする．ハザード関数のカーネル型推定量は

$$\frac{\widehat{f}_n(x_0)}{1-\widehat{F}_n(x_0)}$$

で与えられる．この推定量は分母の収束のオーダーは $n^{1/2}$ であるが，分子のオーダーは $(nh_n)^{1/2}$ である．したがって平均二乗誤差は分子の推定量の平均二乗誤差が支配的となる．

カーネル関数 $K(u)$ は２次のシンメトリックカーネル，すなわち

$$\int_{-\infty}^{\infty}K(u)du=1, \qquad \int_{-\infty}^{\infty}K(u)u^2du\neq 0, \qquad K(-u)=K(u) \tag{1}$$

を満たすとする．またバンド幅は

$$h_n = cn^{-d},\ \ (1/4\leq d<1/2) \tag{2}$$

を仮定する．このとき漸近正規性が成立する．$\widehat{f}_n$ の分散を求めると

$$V\left[\widehat{f}_n(x_0)\right]=\frac{f(x_0)}{nh_n}\int K^2(z)dz+\cdots$$

となり，$1-\widehat{F}(x_0)$ の分散は

$$V\left[1-\widehat{F}_n(x_0)\right]=\frac{F(x_0)(1-F(x_0))}{n}+\cdots$$

であるから，標準化するときの収束のオーダーは $(nh_n)^{1/2}$ になる．

[定理 1] 密度関数 f の 3 回微分が有界 $|f^{(3)}|\le M$ であると仮定する．$n\to\infty$ のとき

$$(nh_n)^{1/2}\left[\frac{\widehat{f}_n(x_0)}{1-\widehat{F}_n(x_0)}-\frac{f(x_0)}{1-F(x_0)}\right]\quad\longrightarrow\quad N(0,\tau^2)$$

が成り立つ．ただし

$$\tau^2=\frac{f(x_0)}{\{1-F(x_0)\}^2}\int K^2(u)du$$

である．

以下では分布の近似の精密化を議論する．

2　カーネル型推定量の漸近表現

$(nh_n)^{1/2}$ のオーダーで漸近分布を持つように標準化するので，$o(n^{-1/2})$ までの高次漸近理論を構成するには，カーネル関数と密度関数に対する適当な条件の下で，下記のような漸近表現が必要になる．

[定理 2] 適当な正則条件の下で

$$\begin{aligned}
&\frac{\widehat{f}_n(x_0)}{1-\widehat{F}_n(x_0)}-\frac{f(x_0)}{1-F(x_0)}\\
=\;&\frac{b_{1,n}}{1-F(x_0)}+\frac{f(x_0)b_{2,n}}{\{1-F(x_0)\}^2}+\frac{1}{n}\sum_{i=1}^{n}\left[\frac{f(x_0)}{\{1-F(x_0)\}^2}Z_{i,n}+\frac{1}{1-F(x_0)}Y_{i,n}\right]\\
&+\frac{1}{n^2}\sum_{1\le i<j\le n}\frac{1}{\{1-F(x_0)\}^2}(Y_{i,n}Z_{j,n}+Y_{j,n}Z_{i,n})+(nh_n)^{-1/2}R_n
\end{aligned}$$

が成り立つ．ただし

$$\begin{aligned}
Y_{i,n}&=\frac{1}{h_n}\left[K\left(\frac{x_0-X_i}{h_n}\right)-E\left\{K\left(\frac{x_0-X_i}{h_n}\right)\right\}\right],\\
Z_{i,n}&=W\left(\frac{x_0-X_i}{h_n}\right)-E\left\{W\left(\frac{x_0-X_i}{h_n}\right)\right\},\\
b_{1,n}&=\frac{h_n^2}{2}f''(x_0)\int z^2K(z)dz,\\
b_{2,n}&=\frac{h_n^2}{2}f'(x_0)\int z^2K(z)dz
\end{aligned}$$

である．ここで残差項は $E|R_n|^p=O(n^{-1/2-p/2-\delta})$ となる $p\ge 2,\ \delta>0$ が存在することが示せる．

この漸近表現を利用すると，残差項 $o(n^{-1/2})$ までのエッジワース展開を求めることができる．ここで次の記号を準備する．

$$\begin{aligned} b_n &= \frac{b_{1,n}}{1-F(x_0)} + \frac{f(x_0)b_{2,n}}{\{1-F(x_0)\}^2}, \\ \alpha_{1,n}(X_i) &= \frac{f(x_0)}{\{1-F(x_0)\}^2} Z_{i,n} + \frac{1}{1-F(x_0)} Y_{i,n} \\ \alpha_{2,n}(X_i,X_j) &= \frac{1}{\{1-F(x_0)\}^2}(Y_{i,n}Z_{j,n} + Y_{j,n}Z_{i,n}). \end{aligned}$$

このとき

$$\begin{aligned} & \frac{\widehat{f}_n(x_0)}{1-\widehat{F}_n(x_0)} - \frac{f(x_0)}{1-F(x_0)} \\ =\; & b_n + n^{-1}\sum_{i=1}^{n} \alpha_1(X_i) + n^{-2}\sum_{1\le i<j\le n} \alpha_2(X_i,X_j) + (nh_n)^{-1/2}R_n \end{aligned} \tag{3}$$

となり，漸近 U-統計量の形を求めることができる．

最初に分散の漸近表現を求める．テイラー展開と $f(x)$ について微分可能性を仮定すると

$$\begin{aligned} & E[\alpha_{1,n}^2(X_1)] \\ =\; & \frac{f(x_0)}{h_n\{1-F(x_0)\}^2}\int K^2(u)du \\ & +h_n\left\{-\frac{2f^3(x_0)A_{1,1}}{\{1-F(x_0)\}^4} - \frac{2f(x_0)f'(x_0)A_{1,1}}{\{1-F(x_0)\}^3} + \frac{f''(x_0)}{2\{1-F(x_0)\}^2}\int u^2K^2(u)du\right\} \\ & +O(h_n^2), \end{aligned}$$

$$E[\alpha_{2,n}^2(X_1,X_2)] = O(h_n^{-1})$$

となる．ここで

$$A_{1,1} = \int uK(u)W(u)du$$

である．したがって

$$\begin{aligned} \sigma_n^2 &= Var\left(\frac{\widehat{f}_n(x_0)}{1-\widehat{F}_n(x_0)} - \frac{f(x_0)}{1-F(x_0)}\right) \\ &= \frac{1}{nh_n}\left[\tau^2 + h_n^2\left\{-\frac{2f^3(x_0)A_{1,1}}{\{1-F(x_0)\}^4} - \frac{2f(x_0)f'(x_0)A_{1,1}}{\{1-F(x_0)\}^3}\right.\right. \\ &\qquad \left.\left. + \frac{f''(x_0)}{2\{1-F(x_0)\}^2}\int u^2K^2(u)du\right\} + O(h_n^3)\right] \end{aligned} \tag{4}$$

となることが示せる．

式 (3) の表現は漸近 U-統計量であるから，Lai and Wang (1993) を利用して残差項 $o(n^{-1/2})$ エッジワース展開を求めることができる．

[**定理 3**] 適当な条件の下で

$$\left| P\left(\frac{1}{\sigma_n}\left\{\frac{\widehat{f}_n(x_0)}{1-\widehat{F}_n(x_0)}-\frac{f(x_0)}{1-F(x_0)}-b_n\right\}\leq x\right)-Q_n(x)\right|=o(n^{-1/2})$$

が成り立つ．ただし

$$\begin{aligned}\kappa_{3,n} &= (n^{1/2}\sigma_n)^{-3}\{E[\alpha_{1,n}^3(X_1)]+3E[\alpha_{1,n}(X_1)\alpha_{1,n}(X_2)\alpha_{2,n}(X_1,X_2)]\},\\ Q_n(x) &= \Phi(x)+n^{-1/2}\phi(x)\frac{\kappa_{3,n}(x^2-1)}{6}\end{aligned}$$

で，$\Phi(x)$, $\phi(x)$ は標準正規分布 $N(0,1)$ の分布関数及び密度関数である．$\kappa_{3,n}$ の具体的な表現は密度関数に関する条件を仮定して次の近似を求めることができる．

$$\begin{aligned}&(n^{1/2}\sigma_n)^{-3}E[\alpha_{1,n}^3(X_1)]=h_n^{-1/2}\tau^{-3}\frac{f(x_0)}{\{1-F(x_0)\}^3}\int K^3(u)du+O(h_n^{1/2}),\\ &(n^{1/2}\sigma_n)^{-3}3E[\alpha_{1,n}(X_1)\alpha_{1,n}(X_2)\alpha_{2,n}(X_1,X_2)]=O(h_n^{1/2})\end{aligned}$$

3　補足：キュムラントの近似

ここではバイアス，分散及びキュムラントの具体的な表現をどのように求めるかを示しておく．変数変換 $z=(x_0-y)/h_n$ と部分積分を使うと

$$\begin{aligned}E[\widehat{F}_n(x_0)] &= \int W\left(\frac{x_0-y}{h_n}\right)f(y)dy=\int h_nW(z)f(x_0-h_nz)dz\\ &= \int K(u)F(x_0-h_nu)du\end{aligned}$$

が得られる．ここで F を点 x_0 の周りでテーラー展開すると

$$\begin{aligned}E[\widehat{F}_n(x_0)] &= \int K(z)F(x_0-h_nz)dz\\ &= F(x_0)\int K(z)dz+f(x_0)h_n\int zK(z)dz+\frac{1}{2}h_n^2f'(x_0)\int z^2K(z)dz+\cdots\end{aligned}$$

が得られる．またカーネルの条件 (1) より

$$E[\widehat{F}_n(x_0)]=F(x_0)+\frac{1}{2}h_n^2f'(x_0)\int z^2K(z)dz+O(h_n^3)$$

となる．同様にして

$$\begin{aligned}E[\widehat{f}_n(x_0)] &= \int K(z)f(x_0-h_nz)dz\\ &= f(x_0)\int K(u)du+f'(x_0)h_n\int zK(z)dz+\frac{1}{2}h_n^2f''(x_0)\int z^2K(z)dz+\cdots\\ &= f(x_0)+\frac{1}{2}h_n^2f''(x_0)\int z^2K(z)dz+O(h_n^3)\end{aligned}$$

が得られる．この表現と Maesono (2005) を使って，ハザード関数の漸近表現 (3) を求めることができる．

分散についても同様に，変数変換とテーラー展開を利用して具体的な表現を求めることができる．定義より

$$
\begin{aligned}
&E[\alpha_{1,n}^2(X_1)]\\
=&\ \frac{1}{\{1-F(X_0)\}^2}E(Y_{1,n}^2)+\frac{2f(x_0)}{\{1-F(X_0)\}^3}E(Y_{1,n}Z_{1n})+\frac{f^2(x_0)}{\{1-F(X_0)\}^4}E(Z_{1,n}^2)
\end{aligned}
$$

となる．ここで変数変換とテーラー展開より

$$
\begin{aligned}
E(Y_{1,n}^2) &= \frac{1}{h_n^2}\int K^2\left(\frac{x_0-y}{h_n}\right)f(y)dy\\
&= h_n^{-1}\int K^2(z)f(x_0-h_ny)dy\\
&= h_n^{-1}f(x_0)\int K^2(z)dz-f'(x_0)\int zK^2(z)dz\\
&\qquad +\frac{h_n}{2}f''(x_0)\int z^2K^2(z)dz-f^2(x_0)+O(h_n^2)\\
&= h_n^{-1}f(x_0)\int K^2(z)dz-f^2(x_0)+\frac{h_n}{2}f''(x_0)\int z^2K^2(z)dz-f^2(x_0)+O(h_n^2)
\end{aligned}
$$

が得られる．同様に

$$
\begin{aligned}
E(Y_{1,n}Z_{1,n}) &= \frac{1}{2}f(x_0)-f(x_0)F(x_0)-h_nf'(x_0)A_{1,1}+O(h_n^2)\\
E(Z_{1,n}^2) &= F(x_0)\{1-F(x_0)\}-2h_nf(x_0)A_{1,1}+O(h_n^2)
\end{aligned}
$$

となる．これらを代入すると $E[\alpha_{1,n}^2(X_1)]$ の近似を求めることができる．

また分散 σ_n^2 に含まれる $E[\alpha_{2,n}^2(X_1,X_2)]$ についても

$$
\begin{aligned}
&E[\alpha_{2,n}^2(X_1,X_2)]\\
=&\ \frac{1}{\{1-F(x_0)\}^4}E[\{Y_{1,n}Z_{2,n}+Y_{2,n}Z_{1,n}\}^2]\\
=&\ \frac{2}{\{1-F(x_0)\}^4}\left[E(Y_{1,n}^2)E(Z_{2,n}^2)+\{E(Y_{1,n}Z_{1,n})\}^2\right]\\
=&\ O(h_n^{-1})
\end{aligned}
$$

が示せる．以上をまとめると分散 σ_n^2 の近似 (4) を求めることができる．

3 次のキュムラントについても同様に近似を求めることができる．定義より

$$
\begin{aligned}
&E[\alpha_{1,n}^3(X_1)]\\
=&\ \frac{E(Y_{1,n}^3)}{\{1-F(x_0)\}^3}+\frac{3f(x_0)E(Y_{1,n}^2Z_{1,n})}{\{1-F(x_0)\}^4}+\frac{3f^2(x_0)E(Y_{1,n}Z_{1,n}^2)}{\{1-F(x_0)\}^5}+\frac{f^3(x_0)E(Z_{1,n}^3)}{\{1-F(x_0)\}^6}
\end{aligned}
$$

となる．分散のときと同様にして

$$\begin{aligned}
E(Y_{1,n}^3) &= h_n^{-2} f(x_0) \int K^3(u)du - 3h_n^{-1} f^2(x_0) \int K^2(u)du + O(1), \\
E(Y_{1,n}^2 Z_{1,n}) &= h_n^{-1} f(x_0) \int K^2(u)du \left[\frac{1}{2} - F(x_0)\right] + O(1) \\
E(Y_{1,n} Z_{1,n}^2) &= O(1), \\
E(Z_{1,n}^3) &= O(1)
\end{aligned}$$

の近似が得られる．係数として $h_n^{3/2}$ を掛けるから

$$h_n^{3/2} E[\alpha_{1,n}^3(X_1)] = h_n^{-1/2} f(x_0) \int K^3(u)du + O(h_n^{1/2})$$

となる．同様にして

$$E[\alpha_{1,n}(X_1)\alpha_{1,n}(X_2)\alpha_{2,n}(X_1, X_2)] = O(h_n^{1/2})$$

となることが示せる．これらを利用すると，残差項が $o(n^{-1/2})$ のエッジワース展開を求めることができる．

参考文献

[1] Estévez-Pérez, G, Quitela-del-Río, P and Vieu, P. (2002), *Convergnec rate for cross-validatory bandwidth in kernel hazard estimation from dependent samples.* Jour. Stat. Plan. Inf., Vol.104, 1-30.

[2] Huang, Z. and Maesono, Y. (2014), *Edgeworth expansion for kernel estimators of a distribution function.* Bulletin of Informatics and Cybernetics to appear

[3] Lai, T.L. and Wang, J.Q. (1993), *Edgeworth expansions for symmetric statistics with applications to bootstrap methods.* Statistica Sinica, Vol.3., 517-542.

[4] Maesono, Y. (2005), *Asymptotic representations of ratio statistics and their mean squared errors.* Journal of the Japan Statistical Society, Vol.35, pp.73-97.

[5] Patil, P. N. (1993), *Nonparametric hazard rate estimation by orthogonal wavelet methods.* Jour. Stat. Plan. Inf., Vol.35, 15-30.

[6] Umeno, S. and Maesono, Y. (3013) *Improvement of normal approximation for kernel density estimator.* Bulletin of Informatics and Cybernetics, Vol.45, pp.11-24.