Inverse problems in Statistics Chapter 7: Asymptotic distribution theory

Bruno Gagliano

June 20, 2007

7.2 Plug-in: Wicksell

Recall quickly the plug-in estimator for the function V in Wicksell's problem. Given an i.i.d sequence $Z_1, Z_2, \dots$ of squared circle radii, the plug in estimator for the function V based on $Z_1, Z_2, \dots, Z_n$ is defined by

$$\tilde{V}_n(x) = \int_{(x,\infty)} (z-x)^{-1/2} dG_n(z) = \frac{1}{n}\sum_{i=1}^{n}(Z_i - x)^{-1/2} 1_{(x,\infty)}(Z_i),$$

for each $x \geq 0$.
The goal of this subsection is to find a the asymptotic distribution of the plug-in estimator V_n for the distribution function. We will see that

$$\sqrt{\frac{n}{\log n}}\big(\tilde{V}_n(x) - V(x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, g(x)).$$

In section 6.2 we saw that this estimator is pointwise consistent by the strong law of large numbers:

$$\tilde{V}_n(x) = \int_{(x,\infty)} \frac{dG_n(z)}{\sqrt{z-x}} \longrightarrow \int_{(x,\infty)} \frac{dG(z)}{\sqrt{z-x}} = V(x),$$

for each $x \geq 0$, as $n \to \infty$.

Now we fix x and note that

$$\tilde{V}_n(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

where

$$Y_i = \frac{1_{(x,\infty)}(Z_i)}{\sqrt{Z_i - x}}.$$

Now suppose that x is a point such that g is strictly positive in a neighbourhood of x. Then $Var(Y_i) = \infty$. Further, consider the random variable Y and note that

$$P[Y > y] = P[Z > x \wedge Z < x + y^{-2}] = G(x + y^{-2}) - G(x),$$

so that the r.v. Y has a point mass of $G(x)$ at zero, i.e.

$$P[Y \leq y] = 1 - P[Y > y] = 1 - \underbrace{G(x + y^{-2})}_{\to 1\,(y \to 0)} + G(x) \longrightarrow G(x)\,(y \to 0).$$

Y has a density of

$$h(y) = \frac{d}{dy}H(y) = \frac{d}{dy}\big(1 - G(x + y^{-2}) + G(x)\big) = 2y^{-3}g(x + y^{-2}).$$

The variance of Y therefore again is $Var(Y) = \infty$.
This is rather inconvenient, since the classical CLT cannot be used since it requires finite variance. However, there is a solution to our problem, stated in the following lemma

Lemma 7.1. 7.2.1 Let $Y_1, Y_2, \ldots$ be i.i.d. with distribution function H. Denote by Φ the standard normal distribution function. Then

$$\lim_{n\to\infty} P\Big[\frac{1}{B_n}\sum_{i=1}^{n} Y_i - A_n < x\Big] = \Phi(x)$$

for some $B_n > 0$ and A_n, iff

$$\lim_{c\to\infty} \frac{c^2 \int_{|y|>c} dH(y)}{\int_{|y|<c} y^2 dH(y)} = 0.$$

If this condition holds, the normalising sequences may be chosen as

$$B_n = \sup\Big\{c :\; c^{-2}\int_{|y|<c} y^2 dH(y) \geq 1/n\Big\}$$

and

$$A_n = \frac{n}{B_n}\int_{|y|>c} y dH(y)$$

Now we apply the above lemma to our situation. We have

$$\int_{|y|>c} dH(y) = P[Y > c] = G(x + c^{-2}) - G(x),$$

such that $c^2 \int_{|y|>c} dH(y)$ tends to $g(x)$ as $c \to \infty$. The denominator $\int_{|y|<c} y^2 dH(y)$ can be written as

$$\begin{aligned}\int_{|y|<c} y^2 dH(y) &= \int_0^c y^2 h(y)dy = \int_0^c 2y^{-1}g(x + y^{-2})dy \\ &= \int_0^1 2y^{-1}g(x + y^{-2})dy + \int_1^c 2y^{-1}g(x + y^{-2})dy.\end{aligned}$$

The first term is finite since, changing variables $y^{-2} \to u$

$$2\int_0^1 y^{-1}g(x + y^{-2})dy = \int_1^\infty \frac{g(x + u)}{u}du \leq \int_0^\infty g(u)du = 1.$$

For the second term, using that g is continuous at x,

$$2\int_1^c y^{-1}g(x+y^{-2})dy \sim 2g(x)\log c.$$

This satisfies the condition in the lemma. Using the expressions for B_n and A_n given in the lemma we see that we may take

$$B_n = \sqrt{g(x)n\log n} \quad \text{and} \quad A_n = \sqrt{\frac{n}{g(x)\log n}}V(x).$$

This shows that

$$\sqrt{\frac{n}{\log n}}\big(\tilde{V}_n(x) - V(x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, g(x)).$$

7.5 A worked example

In this section we will derive the asymptotic distribution of the isotonic inverse estimator in the exponential deconvolution problem. In the exponential deconvolution problem we have

$$U(x) = G(x) + \int_0^x G(x)dx \quad \text{and} \quad U_n(x) = G_n(x^-) + \int_0^x G_n(x)dx,$$

where G_n is the empirical distribution function based on the first n random variables of an i.i.d. sequence generated by the density

$$g(z) = \int_0^z e^{x-z}dF(x)$$

with corresponding distribution function G. $G_n(x^-)$ stands for $\lim_{y\uparrow x} G_n(y)$. The choice of $G_n(x^-)$ instead of $G_n(x)$ does not influence the convex minorant and its derivative. The estimator $\tilde{F}_n(x)$ is the right derivative of the convex minorant of U_n evaluated at x. The goal in this example is to show that

$$n^{1/3}f(x)^{-1/3}g(x)^{-1/3}2^{-2/3}\big(\tilde{F}_n(x) - F(x)\big) \xrightarrow{\mathcal{D}} \operatorname{argmin}_{t\in\mathbb{R}}\big(W(t) + t^2\big),$$

where f is the density of the distribution function F, g is the density of the distribution function G and W is the standard Brownian Motion.
Now fix $x > 0$ such that $F'(y) = f(y)$ for a continuous, strictly positive f.
From lemma 7.3.1 it follows that

$$\delta_n^{-1}\big(\tilde{F}_n(x) - F(x)\big) < v \iff \operatorname{argmin}_s^- Z_n(s) > 0,$$

where

$$Z_n(s) = W_n(s) + D_n(s) - vs.$$

Our goal for now is to investigate $\operatorname{argmin}_s^- Z_n(s)$. To that purpose we have a closer look at $W_n(s)$ and $D_n(s)$. For $s > 0$

$$\begin{aligned} W_n(s) &= \delta_n^{-2}\big(U_n(x+\delta_n s - U_n(x) - U(x+\delta_n s) + U(x)\big) \\ &= \delta_n^{-2}\int_{]x,x+\delta_n s]} d(G_n - G)(z) + \delta_n^{-2}\int_{]x,x+\delta_n s]}(G_n(z) - G(z))dz \end{aligned}$$

and

$$\begin{aligned} D_n(s) &= \delta_n^{-2}\big(U(x+\delta_n s - U(x) - F(x)\delta_n s\big) \\ &= \delta_n^{-2}\int_x^{x+\delta_n s}\big(F(y)-F(x)\big)dy = \frac{1}{2}f(x)s^2 + R_n(s), \end{aligned}$$

where $R_n \downarrow 0$ uniformly on compacta, when $\delta_n \downarrow 0$. The case $s < 0$ proceeds analogously and will not be shown here.
Now we need to determine the rate of δ_n, such that W_n does not become asymptotically degenerate. This is equivalent to saying that W_n must be of $O_P(1)$. The second term of W_n is of $\delta_n^{-1}n^{-1/2}$, since $\sup_{x\le z\le x+\delta_n s}|G_n(z)-G(z)| = O_P(n^{-1/2})$, as seen in chapter 6. The first term is an increment of the empirical process over an interval of length $\delta_n s$, multiplied by $n^{-1/2}\delta_n^{-2}$. This will therefore be $O_P(\delta^{-3/2}n^{-1/2})$. The first term dominates the second one, and to make W_n asymptotically non degenerate we need $\delta_n \sim n^{-1/3}$.
We now study W_n with $\delta_n \sim n^{-1/3}$

$$\begin{aligned} W_n(s) &= n^{2/3}\int_{]x,x+n^{-1/3}s]} d(G_n-G)(z) + n^{2/3}\int_{]x,x+n^{-1/3}s]}(G_n(z)-G(z))dz \\ &= n^{2/3}\int_{]x,x+n^{-1/3}s]} d(G_n-G)(z) + R_n^{(2)}(s), \end{aligned}$$

where for each $K < \infty$

$$\sup_{0\le s\le K} R_n^{(2)}(s) \le n^{2/3}n^{-1/3}K\|G_n-G\|_\infty = O_P(n^{-1/6}),$$

as $n\to\infty$. Using the theorem (7.4.2) for each $K<\infty$

$$W_n \rightsquigarrow \sqrt{g(x)}W \,\text{in}\, l^\infty([-K,K]),$$

where W is the two-sided Brownian Motion on $\mathbb{R}$. Therefore,

$$Z_n \rightsquigarrow Z \,\text{in}\, l^\infty([-K,K]).$$

Here:

$$Z(s) = \sqrt{g(x)}W(s) + \frac{1}{2}f(x)s^2 - vs$$

Since $Var\big(Z(s)-Z(t)\big) \neq 0$ for all $s \neq t$ the process Z has a.s. a unique minimiser. We will call this minimiser

$$\hat{h} := \text{argmin}_s Z(s).$$

The next step in establishing the asymptotic distribution of $\tilde{F}(x)$ is to show the tightness of the argmins. To this end we will apply theorems (7.4.3) and (7.4.4). The result of these applications is

$$\hat{\theta}_n = \text{argmin}^-_\theta \mathbb{M}_n(\theta) = O_P(n^{-1/3}),$$

where

$$\begin{aligned}\mathbb{M}_n(\theta) &= U_n(x+\theta) - U_n(x) - \theta F(x) - v\theta n^{-1/3} \\ &= \int_0^{x+\theta} (1+x-\theta-z)dG_n(z) - \int_0^{x+} (1+x-z)dG_n(z) - \theta F(x) - v\theta n^{-1/3}\end{aligned}$$

and

$$\mathbb{M}(\theta) = U(x+\theta) - U(x) - \theta F(x),$$

$\theta_0 = 0$, and $\mathbb{M}_n(0) = \mathbb{M}(0) = 0$.
Since

$$Z_n(t) = n^{2/3}\mathbb{M}_n(n^{-1/3}t),$$

this means that

$$\hat{h}_n = \text{argmin}_h^- Z_n(h) = O_P(1).$$

This establishes the uniform tightness of $\hat{h}_n$, which here simply means that the $\hat{h}_n$ are bounded. Now - finally - theorem (7.4.1) can be used to conclude that

$$\hat{h}_n \xrightarrow{\mathcal{D}} \hat{h} = \text{argmin}_s \sqrt{g(x)}W(s) + \frac{1}{2}f(x)s^2.$$

Using the scaling property of Brownian motion, we see that for each $a > 0$ and $b \in \mathbb{R}$

$$\text{argmin}_{t\in\mathbb{R}}\big(aW(t) + (t-b)^2\big) \stackrel{\mathcal{D}}{=} a^{2/3}\text{argmin}_{t\in\mathbb{R}}\big(W(t)+t^2\big) + b$$

implying that

$$\text{argmin}_{t\in\mathbb{R}}\big(\sqrt{g(x)}W(t) + \frac{1}{2}f(x)t^2 - vt\big) \stackrel{\mathcal{D}}{=} \frac{g(x)^{1/3}2^{2/3}}{f(x)^{2/3}}\text{argmin}_{t\in\mathbb{R}}\big(W(t)+t^2\big) + \frac{v}{f(x)}$$

using the fact that the argmin function is invariant under multiplication by $c > 0$ and addition of any $d \in \mathbb{R}$. Hence

$$\lim_{n\to\infty} P\big[n^{1/3}\big(\tilde{F}_n(x) - F(x)\big) < v\big] = P\big[\text{argmin}_{t\in\mathbb{R}}\big(W(t)+t^2\big) > -vf(x)^{-1/3}g(x)^{-1/3}2^{-2/3}\big]$$

to show that

$$n^{1/3}f(x)^{-1/3}g(x)^{-1/3}2^{-2/3}\big(\tilde{F}_n(x) - F(x)\big) \xrightarrow{\mathcal{D}} \text{argmin}_{t\in\mathbb{R}}\big(W(t)+t^2\big).$$