Statistics 583, Problem Set 5 Solutions

Wellner; 5/3/2006

1. Suppose that we observe $X_1, \ldots, X_n$ i.i.d. P on $\mathbb{R}^+ = [0, \infty)$ and assume that $P \in \mathcal{P}_0 \equiv \{P_\theta : (dP_\theta/d\lambda) = p_\theta,\ \theta \in \Theta\}$ where $\theta = (\alpha, \beta) \in (0, \infty)^2$ and $p_\theta = p_{\alpha,\beta}$ is the Weibull density given by $p_\theta(x) = (\beta/\alpha)(x/\alpha)^{\beta-1}\exp(-(x/\alpha)^\beta)1_{(0,\infty)}(x)$. From Lehmann and Romano, TPE, Example 6.6.1 (page 468) and problems 6.6.1 - 6.6.3 (page 509), we know that the maximum likelihood estimator $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ exists and is unique if $0 < X_{(1)} < X_{(n)}$.

   (a) If, in fact, $P \notin \mathcal{P}_0$, to what function of P do you expect $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ converges (in probability)? [Hint: use the development in example 6.6.1 of Lehmann and Romano to show that the solution of the population version of the score equations (with sampling from $P \notin \mathcal{P}$) leads to $\alpha(P) = \{E_P(X^\beta)\}^{1/\beta}$ where β is the solution of
$$\frac{E_P(X^\beta \log X)}{E_P X^\beta} - \frac{1}{\beta} = E_P(\log X),$$
   assuming that $E_P(X^\beta|\log X|) < \infty$.]

   (b) Show heuristically that $\theta(P) = (\alpha(P), \beta(P))$ minimizes $K(P, P_\theta)$ over Θ.

   (c) Show in particular that if P has density $p(x) = 2\phi(x)1_{(0,\infty)}(x)$ where ϕ is the standard normal density (so p is the "half-normal" density), then $(\alpha(P), \beta(P)) = (0.855417\ldots, 1.26276\ldots)$. Plot $p(x)$ and the "best-fitting" member of the Weibull family $p_{(\alpha(P),\beta(P))}(x)$ as functions of x.

   Solution: (a) From our discussions in class we expect that $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ will converge to the solution $T(P) \equiv \theta(P) = (\alpha(P), \beta(P))$ of
$$P\dot{\mathbf{l}}_\theta(X; \theta(P)) = \int \dot{\mathbf{l}}_\theta(x; \theta(P))dP(x) = 0 \tag{1}$$
   where $\theta(P) = (\alpha(P), \beta(P)) \in \mathbb{R}^2$ and $\dot{\mathbf{l}}_\theta(\cdot; \theta)$ is the vector of score functions for the Weibull model. Since
$$\dot{\mathbf{l}}_\theta(x; \theta) = \begin{pmatrix} \frac{\beta}{\alpha}\left(\left(\frac{x}{\alpha}\right)^\beta - 1\right) \\ \frac{1}{\beta}\left(1 - \log\left\{\left(\frac{x}{\alpha}\right)^\beta\right\}\left\{\left(\frac{x}{\alpha}\right)^\beta - 1\right\}\right) \end{pmatrix},$$
   it follows that $\theta(P) = (\alpha(P), \beta(P))$ satisfies
$$\int (x/\alpha)^\beta dP(x) = 1, \qquad \text{and}$$
$$\int \left(1 - \log\left\{\left(\frac{x}{\alpha}\right)^\beta\right\}\left\{\left(\frac{x}{\alpha}\right)^\beta - 1\right\}\right) dP(x) = 0, .$$

The first equation yields

$$\alpha(P) = \{E_P X^\beta\}^{1/\beta},$$

and hence the second equation can be re-written as

$$\begin{aligned} 1 &= E_P\left\{\beta \log X\left(\left(\frac{X}{\alpha}\right)^\beta - 1\right)\right\} \\ &= \beta\left\{\frac{E_P(X^\beta \log X)}{E_P X^\beta} - E_P \log X\right\} \end{aligned}$$

or, equivalently,

$$\frac{E_P(X^\beta \log X)}{E_P X^\beta} - \frac{1}{\beta} = E_P(\log X). \tag{2}$$

Letting

$$h(\beta) \equiv \frac{E_P(X^\beta \log X)}{E_P X^\beta} - \frac{1}{\beta},$$

we have

$$\begin{aligned} h'(\beta) &= \frac{E_P\{X^\beta(\log X)^2\}}{E(X^\beta)} - \frac{[E_P(X^\beta \log X]^2}{[E(X^\beta)]^2} + \frac{1}{\beta^2} \\ &\geq Var_Q(\log X) \quad \text{under } \ Q \ \text{ with } \ dQ(x) \equiv \frac{x^\beta}{E_P(X^\beta)} dP(x) \\ &> 0 \end{aligned}$$

unless P is degenerate. Thus if P is non-degenerate $h(\beta)$ is monotone increasing with

$$-\infty = \lim_{\beta \downarrow 0} h(\beta) < E_P(\log X) < \lim_{\beta \to \infty} h(\beta) = \sup\{\text{support}(P)\}.$$

Thus, in parallel to the finite sample situation as in the problems in Lehmann and Romano, the equation (2) has a unique finite solution in $(0, \infty)$ if P is non-degenerate, does not have positive mass at 0 and provided that X has sufficiently many moments: e.g. if $E_P X^r < \infty$ for some $r > \beta(P)$.

(b) We did this in class. Here is th heuristic argument showing why this should be true: Note that for many cases we have

$$\begin{aligned} \hat{\theta}_n &= \text{argmax}_\theta n^{-1} l_n(\theta) = \text{argmax}_\theta \mathbb{P}_n(\log \theta) \\ &\to_p \text{argmax}_\theta P(\log \theta) = \text{argmax}_\theta \int \log p_\theta(x) dP(x). \end{aligned}$$

Now

$$\begin{aligned} P(\log p_\theta) &= P(\log p) + P\log\left(\frac{p_\theta}{p}\right) \\ &= P(\log p) - P\log\left(\frac{p}{p_\theta}\right) \\ &= P(\log p) - K(P, P_\theta). \end{aligned}$$

Thus

$$\text{argmax}_\theta \int \log p_\theta(x) dP(x) = \text{argmin}_\theta K(P, P_\theta) \equiv \theta(P).$$

If we can interchange differentiation and integration it follows that

$$\nabla_\theta K(P, P_\theta) = \int p(x) \dot{\mathbf{l}}_\theta(x;\theta) d\mu(x) = \int \dot{\mathbf{l}}_\theta(x;\theta) dP(x),$$

so the relation (1) is obtained by setting this gradient vector equal to 0.
(c) When P has density p given by $p(x) = 2\phi(x)1_{(0,\infty)}(x)$ we compute

$$\begin{aligned} E_P X^\beta &= \sqrt{\frac{2}{\pi}} \int_0^\infty x^\beta e^{-x^2/2} dx = \sqrt{\frac{2}{\pi}} \int_0^\infty x^{\beta-1} e^{-x^2/2} x dx \\ &= \sqrt{\frac{2}{\pi}} \int_0^\infty (2y)^{(\beta-1)/2} e^{-y} dy = \sqrt{\frac{2}{\pi}} 2^{(\beta-1)/2} \Gamma((\beta+1)/2), \qquad (3) \\ E_P(\log X) &= \sqrt{\frac{2}{\pi}} \int_0^\infty \log x e^{-x^2/2} dx \\ &= \sqrt{\frac{2}{\pi}} \int_0^\infty (2y)^{-1/2} \log(\sqrt{2y}) e^{-y} dy \\ &= \sqrt{\frac{2}{\pi}} 2^{-1/2} \left\{ (1/2)(\log 2)\Gamma(1/2) + (1/2) \int_0^\infty y^{-1/2} \log(y) e^{-y} dy \right\} \\ &= \sqrt{\frac{2}{\pi}} 2^{-3/2} \Gamma(1/2) \{\log 2 + \psi(1/2)\}, \qquad \text{and} \qquad (4) \\ E_P(X^\beta \log X) &= \sqrt{\frac{2}{\pi}} \int_0^\infty x^\beta (\log x) e^{-x^2/2} dx = \sqrt{\frac{2}{\pi}} \int_0^\infty x^{\beta-1} (\log x) e^{-x^2/2} x dx \\ &= \sqrt{\frac{2}{\pi}} \int_0^\infty (2y)^{(\beta-1)/2} \log(\sqrt{2y}) e^{-y} dy \\ &= \sqrt{\frac{2}{\pi}} 2^{(\beta-1)/2-1} \left\{ (\log 2)\Gamma((\beta+1)/2) + \int_0^\infty y^{(\beta-1)/2} \log(y) e^{-y} dy \right\} \\ &= \sqrt{\frac{2}{\pi}} 2^{(\beta-1)/2-1} \Gamma((\beta+1)/2) \{(\log 2) + \psi(\beta+1)/2)\} \qquad (5) \end{aligned}$$

where $\psi(x) = (\log \Gamma(x))'$ is the digamma function. Thus (2) becomes, using $\Gamma(1/2) = \sqrt{\pi}$,

$$\frac{1}{2}\left\{(\log 2) + \psi((\beta+1)/2)\right\} - \frac{1}{\beta} = \frac{1}{2}\left\{(\log 2) + \psi(1/2)\right\},$$

or, equivalently,

$$\psi((\beta+1)/2) - \psi(1/2) = 2/\beta.$$

The left side is an increasing function of β, while the right side is a decreasing function of β, and they have a unique intersection at $\beta = 1.26276... = \beta(P)$. Then $\alpha(P) = \{E_P X^\beta\}^{1/\beta} = 0.855417...$ by using (3) and $\beta = 1.26276...$, and the resulting Kullback-Leibler distance is $K(P, P_{\theta(P)}) = 0.01061....$ The following figure shows p and the Weibull density $p(\cdot;\theta(P))$ "closest" to p.

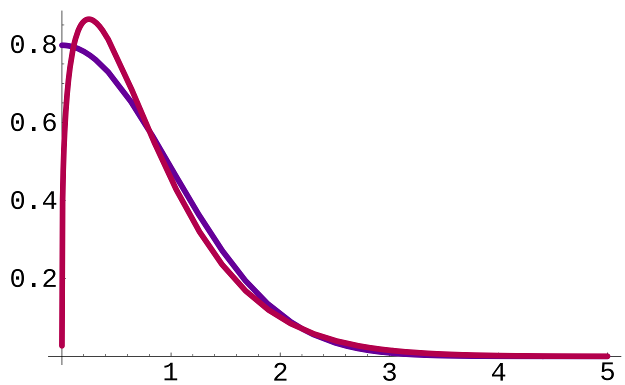

Figure 1: The density p (blue) and the Weibull density $p(\cdot;\theta(P))$ (red) minimizing the Kullback-Leibler distance to p

2. Suppose that $\hat{\theta}_n$ is the MLE for the Weibull family as in problem 1 above, and that $P \notin \mathcal{P}_0$. Heuristically we expect that

$$\sqrt{n}(\hat{\theta}_n - \theta(P)) \rightarrow_d N_2(0, \Sigma(P)) \tag{6}$$

for some covariance matrix $\Sigma = \Sigma(P)$ as $n \to \infty$.
(a) What is the form of Σ that you expect in (6)?
(b) What methods could be used to make these heuristics precise?

Solution: (a) From our heuristic computation of the influence function of a $Z-$estimator and from Huber's Z-theorem, we expect $\Sigma(P)$ to be of the form

$$\Sigma(P) = \dot{\Psi}(P)^{-1} V(P) [\dot{\Psi}(P)^{-1}]^T$$

where $\dot\Psi(P) \equiv \dot\Psi(\theta(P))$ is the derivative of $\Psi(\theta) = P(\dot{\mathbf{l}}_\theta(X,\theta))$ at $\theta(P)$, i.e.

$$\dot\Psi(\theta(P)) = \int \ddot{\mathbf{l}}_{\theta\theta}(x,\theta(P))dP(x),$$

and

$$\begin{aligned} V(P) &= E_P(\dot{\mathbf{l}}(X;\theta(P))\dot{\mathbf{l}}(X;\theta(P))^T) \\ &= \begin{pmatrix} \left(\frac{\beta}{\alpha}\right)^2 E_P\left(\left(\frac{X}{\alpha}\right)^\beta - 1\right)^2 & -\frac{1}{\alpha}E_P\left\{\left(\left(\frac{X}{\alpha}\right)^\beta - 1\right)^2 \log\left(\frac{X}{\alpha}\right)^\beta\right\} \\ -\frac{1}{\alpha}E_P\left\{\left(\left(\frac{X}{\alpha}\right)^\beta - 1\right)^2 \log\left(\frac{X}{\alpha}\right)^\beta\right\} & \frac{1}{\beta^2}E_P\left\{\left(1-\log\left(\frac{X}{\alpha}\right)^\beta\left(\left(\frac{X}{\alpha}\right)^\beta - 1\right)\right)^2\right\} \end{pmatrix} \end{aligned}$$

The matrix $\dot\Psi(P)$ can be written somewhat more explicitly by first calculating $\ddot{\mathbf{l}}_{\theta\theta}$, and then computing the necessary expectations: from the forms of $\dot{\mathbf{l}}_\alpha$ and $\dot{\mathbf{l}}_\beta$ we get

$$\begin{aligned} \ddot{\mathbf{l}}_{\alpha,\alpha}(x,\theta) &= -\frac{\beta}{\alpha^2}\left(\left(\frac{x}{\alpha}\right)^\beta - 1\right) - \left(\frac{\beta}{\alpha}\right)^2\left(\frac{x}{\alpha}\right)^\beta, \\ \ddot{\mathbf{l}}_{\alpha,\beta}(x,\theta) &= \frac{1}{\alpha}\left(\left(\frac{x}{\alpha}\right)^\beta - 1\right) + + \frac{1}{\alpha}\left(\frac{x}{\alpha}\right)^\beta \log\left(\frac{x}{\alpha}\right)^\beta, \\ \ddot{\mathbf{l}}_{\beta,\beta}(x,\theta) &= -\frac{1}{\beta^2}\left(1 - \log\left\{\left(\frac{x}{\alpha}\right)^\beta\right\}\left\{\left(\frac{x}{\alpha}\right)^\beta - 1\right\}\right) \\ &\quad + \frac{1}{\beta}\left\{-\log\left\{\left(\frac{x}{\alpha}\right)^\beta\right\}\left(\frac{x}{\alpha}\right)^\beta \log\left\{\left(\frac{x}{\alpha}\right)\right\}\right. \\ &\qquad \left. - \left\{\left(\frac{x}{\alpha}\right)^\beta - 1\right\}\log\left\{\left(\frac{x}{\alpha}\right)\right\}\right\} \\ &= -\frac{1}{\beta^2}\left(1 - \log\left\{\left(\frac{x}{\alpha}\right)^\beta\right\}\left\{\left(\frac{x}{\alpha}\right)^\beta - 1\right\}\right) \\ &\quad + \frac{1}{\beta^2}\left\{-\log\left\{\left(\frac{x}{\alpha}\right)^\beta\right\}^2\left(\frac{x}{\alpha}\right)^\beta\right. \\ &\qquad \left. 1 - \left\{\left(\frac{x}{\alpha}\right)^\beta - 1\right\}\log\left\{\left(\frac{x}{\alpha}\right)^\beta\right\} - 1\right\}. \end{aligned}$$

Thus we compute

$$\dot\Psi(P) = \begin{pmatrix} -(\beta/\alpha)^2 & \alpha^{-1}E_P(X/\alpha)^\beta\log(X/\alpha)^\beta \\ \alpha^{-1}E_P(X/\alpha)^\beta\log(X/\alpha)^\beta & -\beta^{-2}E_P\{1+(X/\alpha)^\beta\left(\log(X/\alpha)^\beta\right)^2\} \end{pmatrix}$$

where $(\alpha,\beta) \equiv (\alpha(P),\beta(P))$.

3. (a) Exercise 2.7.5, Wasserman, page 24: Suppose that $|T(F) - T(G)| \le C\|F - G\|_\infty$ for some constant $0 < C < \infty$ where $\|F - G\|_\infty = \sup_x |F(x) - G(x)|$. (a1) Prove that $T(\mathbb{F}_n) \to_{a.s.} T(F)$.
(a2) Suppose that $|X| \le M < \infty$. Show that $T(F) = \int x dF(x)$ satisfies the hypothesis $|T(F) - T(G)| \le C\|F - G\|_\infty$.
(b) What can you say about $T(F) = \int x dF(x)$ using this same line of reasoning when the hypothesis $|X| \le M < \infty$ in (a) fails?

Solution: (a1) It follows easily that $|T(\mathbb{F}_n) - T(F)| \le C\|\mathbb{F}_n - F\|_\infty \to_{a.s.} 0$ by the Glivenko-Cantelli theorem. Thus $T(\mathbb{F}_n) \to_{a.s.} T(F)$.
(a2) By Proposition 1.4.2, page 18, of Chapter 1,

$$T(F) = \int x dF(x) = E_F(X) = -\int_{-\infty}^{0} F(x)dx + \int_0^\infty (1 - F(x))dx,$$

and hence for distribution functions F and G with support contained in $[-M, M]$

$$\begin{aligned} T(F) - T(G) &= -\int_{-\infty}^{0} (F(x) - G(x))dx + \int_0^\infty [(1 - F(x)) - (1 - G(x))]dx \\ &= -\int_{-\infty}^{\infty} (F(x) - G(x))dx = -\int_{-M}^{M} (F(x) - G(x))dx \end{aligned}$$

Thus

$$|T(F) - T(G)| = |\int_{-M}^{M} (F(x) - G(x))dx| \le \int_{-M}^{M} |F(x) - G(x)|dx \le 2M\|F - G\|_\infty.$$

(b) If F is not supported on $[-M, M]$ for some $M < \infty$, but $\int |x|dF(x) < \infty$, then we may consider truncation:

$$T(F) = \int_{|x|\le M} x dF(x) + \int_{|x|>M} x dF(x) \equiv T_M(F) + \int_{|x|>M} x dF(x)$$

where we may choose M so large that $\int_{|x|>M} |x|dF(x) < \epsilon$. Similarly, for the same M we have $\int_{|x|>M} |x|d\mathbb{F}_n(x) \to_{a.s.} \int_{|x|>M} |x|dF(x) < \epsilon$ by the strong law of large numbers, so we have

$$\limsup_{n\to\infty} \int_{|x|>M} |x|d\mathbb{F}_n(x) \le 2\epsilon.$$

Then we can apply the argument of (a) to $T_M(F)$ to conclude that $T(\mathbb{F}_n) \to_{a.s.} T(F)$. Note that this argument is slightly circular, however: we used the strong law of large numbers to handle the term $\int_{|x|>M} |x|d\mathbb{F}_n(x)$. Nevertheless this type of truncation argument can produce useful results, in particular in proving Glivenko - Cantelli theorems.

4. (a) Exercise 3.8.1, Wasserman, page 39. [Hint: see Wasserman, page 29.]
(b) Under what additional hypotheses can we establish $\sqrt{n}(T(\mathbb{F}_n) - T(F)) \to N(0, E_F\psi_F^2(X))$? (Here my ψ_F equals Wasserman's L_F.)

Solution: (a) To find the influence function of $T(F) = \int(x-\mu)^3 dF(x)/\sigma(F)^3$, let $F_t \equiv (1-t)F + tG$. Then we need to compute $(d/dt)T(F_t)|_{t=0}$. But, by using the calculations in examples 7.4.2 and 7.4.3,

$$\begin{aligned}
\frac{d}{dt}T(F_t)|_{t=0} &= \frac{d}{dt}\frac{\int(x-\mu(F_t))^3 dF_t(x)}{[\sigma^2(F_t)]^{3/2}}\Big|_{t=0} \\
&= \frac{\int(x-\mu(F_t))^3 d(G-F)(x)}{[\sigma^2(F_t)]^{3/2}}\Big|_{t=0} \\
&\qquad -\frac{3}{2}\frac{\int(x-\mu(F_t))^3 dF_t(x)}{[\sigma^2(F_t)]^{5/2}}\frac{d}{dt}\sigma^2(F_t)\Big|_{t=0} \\
&\qquad -3\frac{\int(x-\mu(F_t))^2 dF_t(x)}{\sigma(F_t)^3}\frac{d}{dt}\mu(F_t)\Big|_{t=0} \\
&= \int\left(\frac{x-\mu(F)}{\sigma(F)}\right)^3 d(G-F)(x) \\
&\qquad -\frac{3}{2}T(F)\frac{1}{\sigma^2(F)}\left\{\int(x-\mu(F))^2 dG(x) - \sigma^2(F)\right\} \\
&\qquad -3\int\left(\frac{x-\mu(F)}{\sigma(F)}\right) dG(x) \\
&= \int\left(\frac{x-\mu(F)}{\sigma(F)}\right)^3 dG(x) - T(F) \\
&\qquad -\frac{3}{2}T(F)\int\left\{\left(\frac{x-\mu(F)}{\sigma(F)}\right)^2 - 1\right\} dG(x) \\
&\qquad -3\int\left(\frac{x-\mu(F)}{\sigma(F)}\right) dG(x).
\end{aligned}$$

Hence by taking $G = \delta_x$ we find the influence function of $T(F)$:

$$\begin{aligned}
\dot{T}(F;\delta_x - F) &= \left(\frac{x-\mu(F)}{\sigma(F)}\right)^3 - T(F) - \frac{3}{2}T(F)\left\{\left(\frac{x-\mu(F)}{\sigma(F)}\right)^2 - 1\right\} \\
&\qquad -3\left(\frac{x-\mu(F)}{\sigma(F)}\right) \\
&\equiv \psi_F(x). \qquad (7)
\end{aligned}$$

Note that this derivation does not seem to agree with the result stated on page 29 of Wasserman: the third term here does not appear in Wasserman's claimed influence function.

(b) Here is a direct calculation to see the result in (a) another way. Write

$$\begin{aligned}
&\sqrt{n}(T(\mathbb{F}_n) - T(F)) \\
&= \frac{1}{\sigma(\mathbb{F}_n)^3}\sqrt{n}\left\{\int (x-\mu(\mathbb{F}_n))^3 d\mathbb{F}_n(x) - \int (x-\mu(F))^3 dF(x)\right\} \\
&\qquad + \int (x-\mu(F))^3 dF(x)\sqrt{n}\left\{\frac{1}{\sigma(\mathbb{F}_n)^3} - \frac{1}{\sigma(F)^3}\right\} \\
&\equiv A_n + B_n.
\end{aligned}$$

To understand A_n, write

$$\begin{aligned}
(x-\mu(\mathbb{F}_n))^3 &= (x-\mu(F) - (\mu(\mathbb{F}_n) - \mu(F)))^3 \equiv (a-b)^3 \\
&= a^3 - 3a^2 b + 3ab^2 + b^3 \\
&= (x-\mu(F))^3 - 3(x-\mu(F))^2(\mu(\mathbb{F}_n) - \mu(F)) \\
&\qquad + 3(x-\mu(F))(\mu(\mathbb{F}_n) - \mu(F))^2 + (\mu(\mathbb{F}_n) - \mu(F))^3.
\end{aligned}$$

Thus we see that

$$\begin{aligned}
A_n &= \frac{1}{\sigma(F)^3}\left\{\sqrt{n}\int (x-\mu(F))^3 d(\mathbb{F}_n(x) - F(x))\right. \\
&\qquad - 3\int (x-\mu(F))^2 d\mathbb{F}_n(x)\sqrt{n}(\mu(\mathbb{F}_n) - \mu(F)) \\
&\qquad \left. + 3\int (x-\mu(F)) d\mathbb{F}_n(x)\sqrt{n}(\mu(\mathbb{F}_n) - \mu(F)))^2 + \sqrt{n}(\mu(\mathbb{F}_n) - \mu(F))^3\right\} + o_p(1) \\
&= \frac{1}{\sigma(F)^3}\left\{\sqrt{n}\int (x-\mu(F))^3 d(\mathbb{F}_n(x) - F(x))\right. \\
&\qquad \left. - 3\sigma^2(F)\sqrt{n}\int (x-\mu(F)) d(\mathbb{F}_n(x) - F(x))\right\} \\
&\qquad + o_p(1).
\end{aligned}$$

For B_n we can write, with $m_3(F) \equiv \int (x-\mu(F))^3 dF(x)$

$$\begin{aligned}
B_n &= m_3(F)\sqrt{n}\left\{\frac{1}{\sigma(\mathbb{F}_n)^3} - \frac{1}{\sigma(F)^3}\right\} \\
&= -\frac{m_3(F)}{\sigma(F)^3\sigma(\mathbb{F}_n)^3}\sqrt{n}\left\{\sigma(\mathbb{F}_n)^3 - \sigma(F)^3\right\} \\
&= -\frac{m_3(F)}{\sigma^2(F)^3}\sqrt{n}\left\{\sigma^2(\mathbb{F}_n)^{3/2} - \sigma^2(F)^{3/2}\right\} + o_p(1) \\
&= -\frac{m_3(F)}{\sigma^2(F)^3}\frac{3}{2}\sigma(F)\sqrt{n}(\sigma^2(\mathbb{F}_n) - \sigma^2(F)) + o_p(1) \\
&= -\frac{m_3(F)}{\sigma^3(F)}\frac{3}{2\sigma^2(F)}\sqrt{n}\int \{(x-\mu(F))^2 - \sigma^2(F)\} d\mathbb{F}_n(x).
\end{aligned}$$

Putting the A_n and B_n pieces together we see that we have complete agreement with the result of the influence function calculation:

$$\sqrt{n}(T(\mathbb{F}_n) - T(F)) = \sqrt{n} \int \psi_F(x) d\mathbb{F}_n(x)$$

where $\psi_F(x)$ is as given in (7). It is clear (from the Central Limit Theorem) that this is asymptotically normal if $E_F X^6 < \infty$.

When I use the influence function derived here to obtain an estimator of the Standard Error of the skewness estimator for the nerve data treated in Wasserman's example 3.10, page 29, I get $\hat{se} = .163$ rather than Wasserman's estimate of .18, a slight reduction. The resulting confidence interval for the population skewness is $1.76 \pm 2(.163) = (1.434, 2.086)$.