第二讲　统计泛函和影响函数

张伟平
统计与金融系

第二讲　统计泛函与影响函数

2.1 统计泛函 2
2.2 影响函数 12
2.3 高阶导数 20

2.1 统计泛函

■ 在很多时候，感兴趣的参数往往是分布函数 F 的函数，记为 $\theta = T(F)$, 称为**统计泛函**.

- 均值: $T(F) = \int x dF(x)$
- 方差 $T(F) = \int (x-\mu)^2 dF(x)$
- 分位数 $T(F) = F^{-1}(p)$
- 相关系数 $\rho(t_1,t_2,t_3,t_4,t_5) = \frac{t_3 - t_1 t_2}{\sqrt{(t_4 - t_1^2)(t_5 - t_2^2)}}$, 其中 $T_1(F) = \int\int x dF(x,y), T_2(F) = \int\int y dF(x,y), T_3(F) = \int xy dF(x,y), T_4(F) = \int x^2 dF(x,y), T_5(F) = \int y^2 dF(x,y)$
- Mann-Whitney 泛函: $T(F,G) = P(X \le Y | X \sim F, Y \sim G) = \int F(x) dG(x)$.

记 F_n 为经验分布函数，则统计泛函 $\theta = T(F)$ 的"Plug-in" 估计为 $\hat{\theta} = T(F_n)$. Definition

■ 常见统计泛函的 plug-in 估计:

- 样本均值: $T(F_n) = \int x dF_n(x) = \frac{1}{n}\sum_{i=1}^n X_i$
- 样本方差: $T(F_n) = \int x^2 dF_n(x) - \left(\int x dF_n(x)\right)^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$
- 样本相关系数:

$$\begin{aligned}\hat{\rho} &= \rho(t_1(F_n), t_2(F_n), t_3(F_n), t_4(F_n), t_5(F_n)) \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}\end{aligned}$$

- Mann-Whitney 统计量:

$$T(F_m, G_n) = \int F_m(x) dG_n(x) = \frac{1}{nm} \sum_{i,j} 1\{X_i \le Y_j\}$$

- 多元协方差矩阵:k 维随机变量 $X \sim F$, 记 $F_n(x) = \frac{1}{n}\sum_{i=1}^n 1\{X_i \le x\} = \frac{1}{n}\sum_{i=1}^n 1\{X_{i1} \le x_1, \ldots, X_{ik} \le x_k\}$, 则由协方差矩阵

$$\begin{aligned} T(F) &= \int (x - \int y dF(y))(x - \int y dF(y))^T dF(x) \\ &= \frac{1}{2} \iint (x-y)(x-y)^T dF(x) dF(y) \end{aligned}$$

知样本协方差矩阵为

$$\begin{aligned} T(F_n) &= \int (x - \int y dF_n(y))(x - \int y dF_n(y))^T dF_n(x) \\ &= \frac{1}{2} \iint (x-y)(x-y)^T dF_n(x) dF_n(y) \end{aligned}$$

连续性

称 $T : \mathcal{F} \mapsto R$ 在 F_0 处连续，如果 $F_n \Rightarrow F_0$ 蕴含 $T(F_n) \to T(F_0)$.

Definition

例: 若 $F_n = (1-\frac{1}{n})F_0 + \frac{1}{n}\delta_{a_n}$, $a_n/n \to \infty$, 则 $T(F) = \int x dF$ 在 F_0 处不连续.

事实上, 由于对任何有界函数 ψ 有

$$\int \psi dF_n = (1-n^{-1}) \int \psi dF_0 + n^{-1}\psi(a_n) \to \int \psi dF_0$$

故有 $F_n \to F_0$. 但是

$$T(F_n) = (1-n^{-1})T(F_0) + n^{-1}a_n \to \infty$$

.

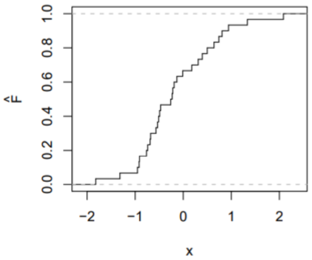
- Plug-in 估计 $T(F_n)$ 是 $T(F)$ 的一个好估计吗?

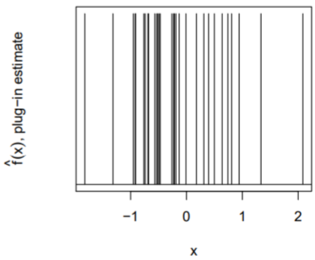
- The Glivenko-Cantelli 定理保证了 $F_n \to F, a.s.$, 是不是也意味着 $T(F_n) \to T(F), a.s.$?

- 答案是: 有时候可以, 有时候不对!

■ 例如 Plug-in 密度估计

泛函 $T(F) = \frac{d}{dx}F(x)$ 的 plug-in 估计显然不是相合估计:





■ 因此什么时候 plug-in 估计为相合估计? **需要对 $T(F)$ 做一些光滑性假设 (可导性)**

■ 如何对一个泛函求导? 我们需要推广导数的定义

Gâteaux derivative

$$\frac{dT(F)}{dF}\text{"} = \text{"}\lim_{\epsilon\to 0}\left[\frac{T(F_\epsilon)-T(F)}{\epsilon}\right]$$

- 一个泛函 T 在 F 处沿方向 G 的 Gâteaux 导数定义为

$$L_F(T;G) = \lim_{\epsilon\to 0}\left[\frac{T((1-\epsilon)F+\epsilon G)-T(F)}{\epsilon}\right]$$

- 一个等价定义为先定义 $D = G - F$，从而

$$L_F(T;D) = \lim_{\epsilon\to 0}\left[\frac{T(F+\epsilon D)-T(F)}{\epsilon}\right]$$

■ 从数学上来看，Gâteaux 导数是方向导数在泛函分析中的推广

■ 从统计上来看，Gâteaux 导数表示的是一个统计泛函在一小部分被分布 G 污染下的变化速率

例：密度 $f(x)$ 的 Gâteaux 导数

■ 设 F 为连续的 CDF，G 为在点 x_0 处退化的随机变量的分布函数

■ 那么 $T(F) = f(x_0)$ 的 Gâteaux 导数是什么？

$$
\begin{aligned}
L_F(T;G) &= \lim_{\epsilon\to 0}\left[\frac{\frac{d}{dx}\{(1-\epsilon)F(x)+\epsilon G(x)\}_{x=x_0} - \frac{d}{dx}F(x)|_{x=x_0}}{\epsilon}\right] \\
&= \lim_{\epsilon\to 0}\left[\frac{(1-\epsilon)f(x_0)+\epsilon g(x_0) - f(x_0)}{\epsilon}\right] \\
&= \infty
\end{aligned}
$$

■ 因此，即便 F 和 F_ϵ 相差无限小，但是 $T(F_n)$ 和 $T(F)$ 可以相差无穷大

■ 本例中 Glivenko-Cantelli 定理并不蕴含 $T(F_n) \to T(F), a.s.$

Hadamard 导数

■ 事实上假设 Gâteaux 可导也还是比较弱，不能保证 $T(F_n) \to T(F), a.s.$

■ 即使是 Gâteaux 导数存在，也可能不是在全域上唯一存在的。这就是 Hadamard 可导性所想解决的

Definition

一个泛函 T 称为是 **Hadamard** 可微的，如果对任意序列 $\epsilon_n \to 0$ 和满足 $\sup_x |D_n(x) - D(x)| \to 0$ 的函数 D_n，有

$$\frac{T(F + \epsilon_n D_n) - T(F)}{\epsilon_n} \to L_F(T, D)$$

定理 1. 如果泛函 T 是 *Hadamard* 可微的，则 $T(F_n) \to T(F), in\ P.$

Gâteaux 导数与 Hadamard 导数

■ 设 $\mathcal{F}$ 表示所有分布函数，$\mathcal{D}$ 表示由 $\mathcal{F}$ 所生成的线性空间。记 $T((1-\epsilon)F+\epsilon G)=T(F+\epsilon D)$，其中 $D=G-F\in\mathcal{D}$.

■ 由 Gâteaux 导数 $L_F(D)$ 的定义知

$$\lim_{\epsilon\downarrow 0}\left|\frac{T(F+\epsilon D)-T(F)}{\epsilon}-L_F(D)\right|\to 0.$$

因此 $T(F+\epsilon D)-T(F)\approx\epsilon L_F(D)+o(\epsilon)$，其中误差项 $o(\epsilon)\to 0$ 当 $\epsilon\to 0$。

■ Hadamard 导数则要求误差项 $o(\epsilon)$ 为一个紧集上的一致小量. 对线性空间 $\mathcal{D}$ 赋一度量 d，此时泛函 $T(F)$ 在 F 处是 Hadamard 可导的，如果存在 $\mathcal{D}$ 上的线性泛函 L_F，使得对任意 $\epsilon_n\to 0$ 和 $\{D,D_1,D_2,\ldots\}\subset\mathcal{D}$ 满足 $d(D_n,D)\to 0$ 和 $F+\epsilon_n D_n\in\mathcal{F}$ 有

$$\lim_{n\to 0}\left(\frac{T(F+\epsilon_n D_n)-T(F)}{\epsilon_n}-L_F(D_n)\right)=0.$$

2.2 影响函数

污染一个点质量

- 在稳健统计中，通过额外的一小部分点来污染一个总体分布已经有很长的历史
- 统计学家一般并不直接使用 Gâteaux 导数，而是使用一种特殊的场合，此时称为**影响函数** (Influence function)，即 G 为在点 x 处退化的随机变量之分布函数:

$$G(u) = \delta_x(u) = \begin{cases} 0, if \;\; u < x \\ 1, if \;\; u \geq x \end{cases}$$

- 影响函数常记为 x 的函数:

$$IF(x; T, F) = \lim_{\epsilon \to 0} \left[\frac{T((1-\epsilon)F + \epsilon\delta_x) - T(F)}{\epsilon} \right]$$

- 与其密切相关的是**经验影响函数**:

$$IF(x;T,F_n)=\lim_{\epsilon\to 0}\left[\frac{T((1-\epsilon)F_n+\epsilon\delta_x)-T(F_n)}{\epsilon}\right]$$

- 例: 设 $T(F)=F(A)$, 其中 A 为给定的事件. 则记 $F_t=(1-t)F+t\delta_x$, 从而

$$\frac{T(F_\epsilon)-T(F)}{\epsilon}=1_A(x)-F(A)$$

- 例: 设 $T(F)=\mu(F)=\int xdF$, 则

$$\frac{T(F_\epsilon)-T(F)}{\epsilon}=x-T(F)$$

- 例: 设 $T(F)=Var_F(X)=\int(x-\mu(F))^2dF(x)$, 则

$$\begin{aligned}\left.\frac{d}{dt}T\left(F_t\right)\right|_{t=0}&=\frac{d}{dt}\int\left(x-\mu\left(F_t\right)\right)^2dF_t(x)\\&=\int(x-\mu(F))^2d(G-F)(x)\end{aligned}$$

$$
\begin{aligned}
&+2\int (x-\mu(F))(-1)\dot{\mu}(F;G-F)dF(x) \\
&=\int (x-\mu(F))^2 d(G-F) \\
&=\int \left\{(x-\mu(F))^2-\sigma_F^2\right\} dG(x) \\
&=(x-\mu(F))^2-\sigma_F^2
\end{aligned}
$$

■ 例. $T(F)=F^{-1}(1/2)$, 假设 F 有密度 f 且满足 $f(F^{-1}(1/2))>0$, 则影响函数

$$
\left.\frac{d}{dt}T\left(F_t\right)\right|_{t=0}=\left.\frac{d}{dt}F_t^{-1}(1/2)\right|_{t=0}
$$

注意 $F_t(F_t^{-1}(1/2))=1/2$, 因此

$$
\begin{aligned}
0&=\left.\frac{d}{dt}F_t\left(F_t^{-1}(1/2)\right)\right|_{t=0} \\
&=\left.\frac{d}{dt}\left\{F\left(F_t^{-1}(1/2)\right)+t(G-F)\left(F_t^{-1}(1/2)\right)\right\}\right|_{t=0}
\end{aligned}
$$

$$= f\left(F^{-1}(1/2)\right) IF(x;T,F) + (G-F)\left(F^{-1}(1/2)\right) + 0$$

故有

$$\begin{aligned} IF(x;T,F) &= -\frac{(G-F)\left(F^{-1}(1/2)\right)}{f\left(F^{-1}(1/2)\right)} \\ &= -\frac{\int \left(1_{\left(-\infty,F^{-1}(1/2)\right)}(x) - 1/2\right) dG(x)}{f\left(F^{-1}(1/2)\right)} \\ &= -\frac{1}{f\left(F^{-1}(1/2)\right)} \left\{1_{\left(-\infty,F^{-1}(1/2)\right]}(x) - 1/2\right\} \end{aligned}$$

■ 例. 设 F 有正的密度 f，则 $T(F) = F^{-1}(p)$ 的影响函数为

$$\begin{aligned} IF(x;T,F) &= -\frac{\int \left(1_{\left(-\infty,F^{-1}(p)\right]}(x) - p\right) dG(x)}{f\left(F^{-1}(p)\right)} \\ &= \begin{cases} \frac{p-1}{f(F^{-1}(p))}, x \leq F^{-1}(p) \\ \frac{p}{f(F^{-1}(p))}, x > F^{-1}(p) \end{cases} \end{aligned}$$

称 $T(F)=\int a(x)dF(x)$ 为**线性泛函**. Definition

■ 线性泛函的 Plog-in 估计为

$$T(F_n)=\int a(x)dF_n(x)=\frac{1}{n}\sum_{i=1}^{n}a(X_i)$$

■ 对**线性泛函**$T(F)=\int a(x)dF(x)$，我们有

- 影响函数 $IF(x;T,F)=a(x)-T(F)$, 经验影响函数 $IF(x;T,F_n)=a(x)-T(F_n)$.
- 对任意 G,

$$T(G)=T(F)+\int IF(x;T,F)dG(x).$$

- $\int IF(x;T,F)dF(x)=0$

- 记 $\tau^2 = \int IF^2(x;T,F)dF(x)$, 则 $\tau^2 = \int (a(x) - T(F))^2 dF(x)$, 且当 $\tau^2 < \infty$ 时

$$\sqrt{n}(T(F) - T(F_n)) \rightsquigarrow N(0, \tau^2).$$

- 记

$$\hat{\tau}^2 = \frac{1}{n}\sum_{i=1}^{n} IF^2(X_i;T,F_n) = \frac{1}{n}\sum_{i=1}^{n}(a(X_i) - T(F_n))^2$$

则 $\hat{\tau}^2 \xrightarrow{P} \tau^2$, 以及 $\hat{se}/se \xrightarrow{P} 1$, 其中 $\hat{se} = \hat{\tau}/\sqrt{n}$, $se = \sqrt{Var(T(F_n))}$.

- $\frac{\sqrt{n}(T(F)-T(F_n))}{\hat{\tau}} \rightsquigarrow N(0,1)$.

■ 对非线性泛函有

定理 2. 设统计泛函 $T(F)$ 相对于距离 $d(F,G)=\sup_x|F(x)-G(x)|$ 是 *Hadamard* 可微的，则

$$\sqrt{n}(T(F_n)-T(F)) \rightsquigarrow N(0,\tau^2),$$

其中 $\tau^2=\int IF^2(x;T,F)(x)dF(x)$. 同时有

$$\frac{(T(F_n)-T(F))}{\hat{se}} \rightsquigarrow N(0,1),$$

其中 $\hat{se}=\hat{\tau}/\sqrt{n}$ 以及 $\hat{\tau}=\frac{1}{n}\sum_{i=1}^n IF^2(X_i;T,F_n)$.

■ **非参数 Delta 方法**: 泛函 $T(F)$ 的 $1-\alpha$ 逐点渐近置信区间为 $T(F_n)\pm z_{\alpha/2}\hat{se}$.

■ **Chain Rule** 设统计泛函 $T(F)$ 对某些函数 $a(t_1,\ldots,t_m)$ 有形式 $T(F)=a(T_1(F),\ldots,T_m(F))$, 则其影响函数为

$$IF(x;T,F)=\sum_{i=1}^m \frac{\partial a}{\partial t_i} IF_i(x;T,F)$$

其中

$$IF_i(x;T,F)=\lim_{\epsilon\to 0}\frac{T_i((1-\epsilon)F+\epsilon\delta_x)-T_i(F)}{\epsilon}.$$

■ 例如样本相关系数 $T(F)=a(T_1(F),T_2(F),T_3(F),T_4(F),T_5(F))$ 的影响函数为

$$IF(x,y;T,F)=\tilde{x}\tilde{y}-\frac{1}{2}T(F)(\tilde{x}^2+\tilde{y}^2),$$

其中

$$\tilde{x}=\frac{x-\int xdF}{\sqrt{\int x^2dF-(\int xdF)^2}},\tilde{y}=\frac{y-\int ydF}{\sqrt{\int y^2dF-(\int ydF)^2}}.$$

2.3 高阶导数

■ 设 $X_1, \ldots, X_n\ i.i.d \sim B(1,p)$, 则 $\sqrt{n}(\bar{X}-p) \rightsquigarrow N(0,p(1-p))$, 若 $g(p)=p(1-p)$, 则由 Delta 方法知道

$$\sqrt{n}(g(\bar{X})-g(p)) \rightsquigarrow N(0,p(1-p)(1-2p)^2)$$

特别, 若 $p=1/2$, 则由于 $g'(1/2)=0$, 从而 $\sqrt{n}(g(\bar{X})-g(1/2)) \rightsquigarrow 0$. 因此需要考虑高阶导数。注意到

$$g(p)=g(1/2)+0\cdot(p-1/2)+\frac{1}{2}(-2)(p-1/2)^2+o((p-1/2)^2)$$

从而

$$n(g(\bar{X})-g(1/2))=-n(\bar{X}-1/2)^2 \rightsquigarrow -\frac{1}{4}\chi_1^2$$

■ 考虑泛函 $T:\mathcal{F}\mapsto\mathcal{R}$, 利用实函数 $g(t)=T(F_t), F_t=F+tD$ 在 $t=0$ 处的 Taylor 展开式:

$$\begin{aligned}g(t)&=g(0)+g'(0)t+\frac{1}{2}g''(0)t^2+\cdots+\frac{1}{m!}g^{(m)}(0)t^m+o(t^m)\\&=T(F)+tT'_F(F,D)+\cdots+\frac{1}{m!}T_F^{(m)}(F,D)+o(t^m)\end{aligned}$$

其中 $T_F^{(m)}(F,D)=\frac{d^m}{dt^m}T(F_t)\Big|_{t=0}$.

■ 令 $t=\frac{1}{\sqrt{n}}$, $D=\sqrt{n}(F_n-F)=G_n$, 则得到**Von Mises**展开:

$$T(F_n)-T(F)=\frac{1}{\sqrt{n}}T'_F(F,G_n)+\cdots+\frac{1}{m!}\frac{1}{n^{m/2}}T_F^{(m)}(F,G_n)+\cdots$$

■ 在很多时候, m 阶导数 $T_F^{(m)}(F,G-F)$ 可以表示为

$$\begin{aligned}T_F^{(m)}(F,G-F)&=\int\cdots\int\psi_m(x_1,\ldots,x_m)d(G-F)(x_1)\cdots d(G-F)(x_m)\\&=\int\cdots\int\psi_{m,F}(x_1,\ldots,x_m)dG(x_1)\cdots dG(x_m)\end{aligned}$$

其中 $\psi_{m,F}$ 由 ψ_m 中心化而成:

$$\psi_{1,F} = \psi_1(x) - \int \psi_1 dF,$$

$$\psi_{2,F} = \psi_2(x_1, x_2) - \int \psi_2(x_1, x_2) dF(x_1)$$
$$- \int \psi_2(x_1, x_2) dF(x_2) + \iint \psi_2(x_1, x_2) dF(x_1) dF(x_2)$$

■ 当 $m = 1$ 时候，一般 $T_F^{(m)}(F, D)$ 为一线性泛函，则

$$T(F_n) - T(F) \approx \frac{1}{\sqrt{n}} T'_F(F, G_n) = \frac{1}{n} \sum_{i=1}^{n} \psi_{1,F}(X_i) = \frac{1}{n} \sum_{i=1}^{n} IF(X_i; T, F)$$

从而

$$\sqrt{n}(T(F_n) - T(F)) \approx \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(X_i; T, F)$$

利用中心极限定理，即可得出极限分布

$$\sqrt{n}(T(F_n) - T(F)) \rightsquigarrow N(0, \tau^2)$$

其中 $\tau^2 = \int IF^2(x; T, F)(x) dF(x)$.

■ 注意到 Fisher 信息量为

$$I\left(F_{\theta_0}\right) = \int \left(\frac{\partial}{\partial \theta} \ln f_\theta(x) \bigg|_{\theta_0} \right)^2 dF_{\theta_0}$$

The Cramér-Rao information inequality: 对任意平方可积的统计量 T, 有

$$\frac{d}{d\theta} \mathbb{E}_\theta T = \int \frac{\partial}{\partial \theta} T f_\theta \mu(dx)$$

$$Var\left(T; F_\theta\right) \geq \frac{\left[\frac{\partial}{\partial \theta} \mathbb{E}_\theta T\right]^2}{I\left(F_\theta\right)}$$

因此 $T(F_n)$ 是渐近有效的, 仅当

$$IF(x;T,F) = I\left(F_{\theta_0}\right)^{-1} \frac{\partial}{\partial\theta}\left(\ln f_\theta(x)\right)\Big|_{\theta_0}$$

- Jackknife 方差的估计与影响函数之间有渐近关系

对统计泛函 $T_n = T(F_n)$, 定义第 j 个 pseudovalue:

$$T^*_{nj} = nT(F_n) - (n-1)T(F_{(j)})$$

其中 $F_{(j)}$ 为去掉第 j 个点后的经验分布函数. 故 Jackknife 估计为

$$T^*_n = \frac{1}{n}\sum_{j=1}^{n} T^*_{nj}$$

这些 pseudovalue 的样本方差为

$$V_n = \frac{1}{n-1}\sum_{j=1}^{n}\left(T^*_{nj} - T^*_n\right)^2$$

T_n 方差的 Jackknife 估计为 $\frac{1}{n}V_n$. 带入到影响函数的定义并取 $\epsilon = \frac{-1}{n-1}$ 有

$$\begin{aligned}
&\frac{n-1}{-1}\left(T\left(\left(1-\frac{-1}{n-1}\right)F_n+\frac{-1}{n-1}\delta_{x_j}\right)-T\left(F_n\right)\right) \\
&=(n-1)\left[T\left(F_n\right)-T\left(F_{(j)}\right)\right] \\
&=T_{nj}^{*}-T\left(F_n\right)
\end{aligned}$$

注意 $\left(1-\frac{-1}{n-1}\right)F_n+\frac{-1}{n-1}\delta_{x_j}=F_{(j)}$. 因此,Jackknife pseudovalue 给出了影响函数的一个近似.

- 基于影响函数的稳健性度量

Gross error sensitivity of T at F

$$\gamma^{*}=\gamma^{*}(T,F)=\sup_{\{x: IF(x;T,F) \text{ exists }\}}|\mathrm{IF}(x;T,F)|$$

用来度量给 F 在一个点上的小扰动后 T 的最大变化. 若 $\gamma^{*}(T,F)<\infty$, 则称 T 在 F 处是 B-robust 的.(B is for bias).

Local shift sensitivity

$$\lambda^* = \lambda^*(T,F) = \sup_{\{x\neq y:\mathbb{F}(x;T,F)\text{ and } \text{IF}(y;T,F) \text{ both exist }\}} \frac{|\,\text{IF}(y;T,F) - \text{IF}(x;T,F)|}{|y-x|}$$

Reject point若 F 是对称的 (且对称中心为原点), 则称

$$\rho^* = \rho^*(T,F) = \inf\{r>0 : \text{IF}(x;T,F) = 0, \quad \forall |x|>r\}$$

为拒绝点. 若不存在这样的 r, 则 $\rho^* = \infty$. 所有大于 ρ^* 的点都被完全拒绝.

Breakdown point 记 $F_\epsilon = (1-\epsilon)F + \epsilon\delta_x$. 则

$$\epsilon^* = \inf\{\epsilon > 0 : \sup_x |T(F) - T(F_\epsilon)| = \infty\}$$

称为估计量 T 的崩溃点.

■ 例. 考虑统计泛函 $T = T(F) = \int x dF$, 则 $IF(x;T,\Phi) = x - T(F)$. 从而 $\gamma^* = \infty$, $\lambda^* = 1$, $\rho^* = \infty$, $\epsilon^* = 0$, 故均值受舍入误差影响不大, 但是对异常点比较敏感.

■ 在大部分情况下，展开到一阶导数足够了，此时极限分布为正态分布。但是，在一些例子中一阶导数为 0，此时就需要展开到二阶导数项。

■ **Serfling's Condition** A_m 假设

$$Var_F(\psi_{k,F}(X_1,\ldots,X_k))\begin{cases} =0 & for\ k<m \\ >0 & for\ k=m \end{cases}$$

$$Let\ R_{mn}=T(F_n)-T(F)-\frac{1}{m!}T_F^{(m)}(F_n-F),$$

$$then\ n^{m/2}R_{mn}=o_p(1).$$

定理 3 (Serfling's theorem A). *设 $X_1, \ldots, X_n$ $i.i.d \sim F$, 泛函 T 满足 Serfling 条件 A_1. 记 $\mu(T, F) = E_F \psi_{1,F}(X_1) = 0$ 以及 $\tau^2 = Var(\psi_{1,F}(X_1))$, 假设 $\tau^2 < \infty$, 则*

$$\sqrt{n}(T(F_n) - T(F)) \rightsquigarrow N(0, \tau^2)$$

证明. 根据条件 A_1 的第二条,

$$\begin{aligned}\sqrt{n}(T(F_n) - T(F)) &= o_p(1) + \sqrt{n} T'_F(F, G_n) \\ &= o_p(1) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{1,F}(X_i) \rightsquigarrow N(0, \tau^2).\end{aligned}$$

□

定理 4 (Serfling's theorem B). 设 $X_1,\ldots,X_n$ *i.i.d* $\sim F$, 泛函 T 满足 *Serfling* 条件 A_2 且 $\psi_{2,F}(x,y)=\psi_{2,F}(y,x)$, 以及 $E_F\psi^2_{2,F}(X_1,X_2)<\infty$, $E_F|\psi_{2,F}(X_1,X_1)|<\infty$, $E_F\psi_{2,F}(x,X_2)=0$, 定义 $A:\mathcal{L}_2(F)\mapsto\mathcal{L}_2(F)$:

$$Ag(x)=\int\psi_{2,F}(x,y)g(y)dF(y), g\in\mathcal{L}_2(F)$$

记 $\{\lambda_k\}$ 为 A 的特征根, 则有

$$n(T(F_n)-T(F))\rightsquigarrow\sum_{k=1}^{\infty}\lambda_k Z_k^2$$

其中 $\{Z_k\}$ 为 *i.i.d* 标准正态随机变量, λ_k 依赖于 F 和泛函 T。

证明. 由条件 A_2 我们有

$$\begin{aligned}n(T(F_n)-T(F))&=n\Big\{T(F_n)-T(F)-\frac{1}{2!}T''_F(T,F_n-F)\Big\}\\&\quad+\frac{n}{2!}T''_F(T,F_n-F)\end{aligned}$$

$$= o_p(1) + \frac{n}{2} \iint \psi_{2,F}(x_1, x_2) dF_n(x_1) dF_n(x_2)$$

记运算 A 的正交特征向量函数和相应的特征根分别为 $\{\phi_k\}$ 和 $\{\lambda_k\}$，则 $A\phi_k = \lambda_k \phi_k$，以及

$$\psi_{2,F}(x, y) = \sum_{k=1}^{\infty} \lambda_k \phi_k(x) \phi_k(y)$$

在 $\mathcal{L}_2(F \times F)$ 中成立。因此

$$\begin{aligned}
&\frac{n}{2} \iint \psi_{2,F}(x_1, x_2) dF_n(x_1) dF_n(x_2) \\
&\quad = \frac{n}{2} \iint \sum_{k=1}^{\infty} \lambda_k \phi_k(x) \phi_k(y) dF_n(x) dF_n(y) \\
&\quad = \frac{n}{2} \sum_{k=1}^{\infty} \lambda_k \Big\{ \int \phi_k dF_n \Big\}^2 = \frac{1}{2} \sum_{k=1}^{\infty} \lambda_k \Big\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_k(X_i) \Big\}^2 \\
&\quad \rightsquigarrow \frac{1}{2} \sum_{k=1}^{\infty} \lambda_k Z_k^2
\end{aligned}$$

其中 $\{Z_i\}$ 为 $i.i.d$ 标准正态分布随机变量，因为 $E_F\phi_k(X_i) = 0$, $E_F\phi_k^2(X_i) = 1$, $E_F\phi_j(X_i)\phi_k(X_i) = 0$. $\square$