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

■ 在很多时候，感兴趣的量 (参数) 可以表示成分布函数 F 的函数，记为 $T(F): \mathcal{F}_0 \mapsto R$, 称为**统计泛函**. 其中 $\mathcal{F}_0$ 为一类 CDF 组成的集合, 满足

(i) $F \in \mathcal{F}_0 \Rightarrow \mathbb{F}_n \in \mathcal{F}_0$

(ii) $\mathcal{F}_0$ 为凸集 (ie, $F, G \in \mathcal{F}_0$ 蕴含 $F + \lambda(G - F) \in \mathcal{F}_0$, 对任意 $\lambda \in [0, 1]$)

■ 例子

- 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 $T_1(F) =$

$\iint xdF(x,y), T_2(F) = \iint ydF(x,y), T_3(F) = \int xydF(x,y), T_4(F) = \int x^2dF(x,y), T_5(F) = \int y^2dF(x,y)$

- Mann-Whitney 泛函: $T(F,G) = P(X \leq Y | X \sim F, Y \sim G) = \int F(x)dG(x)$.

Definition

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

■ 常见统计泛函的 plug-in 估计:

- 样本均值: $T(\mathbb{F}_n) = \int xd\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^n X_i$
- 样本方差: $T(\mathbb{F}_n) = \int x^2d\mathbb{F}_n(x) - \left(\int xd\mathbb{F}_n(x)\right)^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$

- 样本相关系数:

$$\hat{\rho} = \rho(t_1(\mathbb{F}_n), t_2(\mathbb{F}_n), t_3(\mathbb{F}_n), t_4(\mathbb{F}_n), t_5(\mathbb{F}_n))$$
$$= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

- 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 $\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^n 1\{X_i \le x\} = \frac{1}{n}\sum_{i=1}^n 1\{X_{i1} \le x_1, \ldots, X_{ik} \le x_k\}$, 则由协方差矩阵

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

知样本协方差矩阵为

$$T(\mathbb{F}_n) = \int (x - \int y d\mathbb{F}_n(y))(x - \int y d\mathbb{F}_n(y))^T d\mathbb{F}_n(x)$$
$$= \frac{1}{2} \iint (x-y)(x-y)^T d\mathbb{F}_n(x) d\mathbb{F}_n(y)$$

- (M 估计): 设 $\rho : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$, Θ 为 $\mathbb{R}$ 上的开集, 则

$$T(F) = \arg\min_{t\in\Theta} \int \rho(x,t) dF(x), \quad F \in \mathcal{F}_0$$

为一个 M-泛函. $T(\mathbb{F}_n) = \arg\min_{t\in\Theta} \sum_{i=1}^n \rho(X_i, t)$ 为 $T(F)$ 的 M 估计.

连续性

称 $T:\mathcal{F}\mapsto R$ 在 F_0 处连续，如果 $\mathbb{F}_n$ 弱收敛到 $F_0 \Rightarrow T(\mathbb{F}_n)\to T(F_0)$.

Definition

例: $T(F)=(1-2\alpha)^{-1}\int_\alpha^{1-\alpha}F^{-1}(u)du$ ($0<\alpha<1/2$) 在每个 F_0 处连续: $\mathbb{F}_n\Rightarrow F_0$ 蕴含 $\mathbb{F}_n^{-1}(t)\to F_0^{-1}(t)$ a.e. Lebesgue. 因此

$$T\left(\mathbb{F}_n\right)=(1-2\alpha)^{-1}\int_\alpha^{1-\alpha}\mathbb{F}_n^{-1}(u)du$$
$$\to(1-2\alpha)^{-1}\int_\alpha^{1-\alpha}F_0^{-1}(u)du=T\left(F_0\right)$$

例: 若 $\mathbb{F}_n = (1-\frac{1}{n})F_0 + \frac{1}{n}\delta_{a_n}$, $a_n/n \to \infty$, 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

$$\int \psi d\mathbb{F}_n = (1-n^{-1})\int \psi dF_0 + n^{-1}\psi(a_n) \to \int \psi dF_0$$

故有 $\mathbb{F}_n \Rightarrow F_0$. 但是

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

.

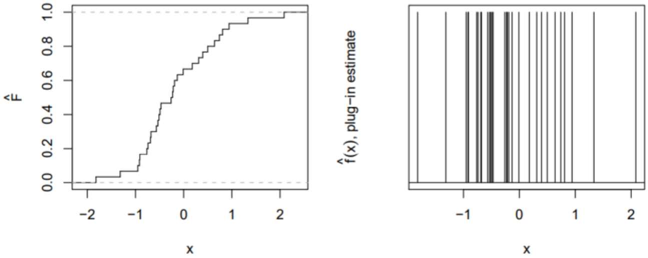
- Plug-in 估计 $T(\mathbb{F}_n)$ 是 $T(F)$ 的一个好估计吗?

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

- 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

导数

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

Gâteaux derivative 设 $F,G\in\mathcal{F}_0$, 对 $F_t=(1-t)F+tG$, $t\in[0,1]$, 泛函 T 在 F 处沿方向 G 的 Gâteaux 导数 $\dot{T}(F;\cdot)$ 定义为

$$\begin{aligned}\dot{T}(F;G-F)&=\frac{d}{dt}T(F+t(G-F))|_{t=0}=\lim_{t\to 0}\left[\frac{T(F_t)-T(F)}{t}\right]\\&=\int\psi(x)d(G(x)-F(x))=\int\psi_F(x)dG(x),\end{aligned}$$

如果存在与 G 无关的可测函数 $\psi_F(x)=\psi-\int\psi dF(x)$ 使得上式成立.

■ 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？

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

其中 $g(x_0) = dG(x)|_{x=x_0} = \infty$. 因此，即便 F 和 F_t 相差无限小，但是 $T(\mathbb{F}_n)$ 和 $T(F)$ 可以相差无穷大

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

2.1.1 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:

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

■ 与其密切相关的是**经验影响函数**:

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

记 $F_t=(1-t)F+t\delta_x$, 则

例: 设 $T(F)=F(A)$, 其中 A 为给定的事件. 从而

$$\frac{T(F_t)-T(F)}{t}=1_A(x)-F(A)$$

例: 设 $T(F)=\mu(F)=\int xdF$, 则

$$\dot{\mu}(F;\delta_x-F)=\lim_{t\to 0}\frac{T(F_t)-T(F)}{t}=x-T(F)$$

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

$$\begin{aligned}\left.\frac{d}{dt}T\left(F_t\right)\right|_{t=0} &= \frac{d}{dt}\int\left(u-\mu\left(F_t\right)\right)^2 dF_t(u)\\ &= \int(u-\mu(F))^2 d(\delta_x-F)(u)\\ &+2\int(u-\mu(F))(-1)\dot{\mu}(F;\delta_x-F)dF(u)\\ &= \int(u-\mu(F))^2 d(\delta_x-F)\end{aligned}$$

$$
\begin{aligned}
&= \int \left\{ (u-\mu(F))^2 - \sigma_F^2 \right\} d\delta_x(u) \\
&= (x-\mu(F))^2 - \sigma_F^2
\end{aligned}
$$

注意 $\dot{\mu}(F;\delta_x - F) = x - \mu(F)$ 仅与退化点 x 有关，与积分元无关.

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

$$
\begin{aligned}
0 &= \left.\frac{d}{dt} F_t\left(F_t^{-1}(1/2)\right)\right|_{t=0} \\
&= \left.\frac{d}{dt}\left\{F\left(F_t^{-1}(1/2)\right) + t(\delta_x - F)\left(F_t^{-1}(1/2)\right)\right\}\right|_{t=0} \\
&= f\left(F^{-1}(1/2)\right) IF(x;T,F) + (\delta_x - F)\left(F^{-1}(1/2)\right) + 0
\end{aligned}
$$

故有

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

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

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

■ 线性泛函的 Plog-in 估计为

$$T(\mathbb{F}_n) = \int a(x) d\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} a(X_i)$$

■ 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 $IF(x;T,\mathbb{F}_n) = a(x) - T(\mathbb{F}_n)$.
- 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

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

- 记

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

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

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

如果泛函 $T(G)$ 是 Gâteaux 可导的, 则 $T((1-\epsilon)F+\epsilon G)$ 可以利用在点 $\epsilon=0$ 处 Taylor 展开得到一阶近似 (取 $\epsilon=1$):

$$\begin{aligned} T(G)-T(F) &= \int \psi_F(x)dG(x) + \mathrm{Rem} \\ &= \int \psi_F(x)d[G(x)-F(x)] + \mathrm{Rem} \end{aligned}$$

使用 $\mathbb{F}_n$ 代替 G, 我们有

$$\begin{aligned} \sqrt{n}\left[T\left(\mathbb{F}_n\right)-T(F)\right] &= \sqrt{n}\int \psi_F(x)d\mathbb{F}_n(x) + R_n \\ &= n^{-1/2}\sum_{i=1}^{n}\psi_F(X_i) + R_n \end{aligned}$$

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

$$IF(x;T,F) = -\frac{\int \left(1_{(-\infty,F^{-1}(p)]}(u)-p\right)d\delta_x(u)}{f\left(F^{-1}(p)\right)}$$

因此由于

$$
n^{-1/2}\sum_{i=1}^{n}\psi_F(X_i)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{p-I_{\left(-\infty,F^{-1}(p)\right]}(X_i)}{f\left(F^{-1}(p)\right)}
$$

$$
\xrightarrow{\mathcal{L}} N\left(0,\frac{p(1-p)}{f^2\left(F^{-1}(p)\right)}\right)
$$

可以证明 $R_n=O_{a.s.}(n^{-1/4}logn)=o_p(1)$ (Bahadur,1966), 故知 $\sqrt{n}\left[T\left(\mathbb{F}_n\right)-T(F)\right]\xrightarrow{\mathcal{L}} N\left(0,\frac{p(1-p)}{f^2\left(F^{-1}(p)\right)}\right)$.

■ 事实上假设 Gâteaux 可导也还是比较弱，不能保证 $R_n=o_p(1)$.

■ 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

Hadamard 导数

一个泛函 T 称为在 F 处相对度量 ρ 是 Hadamard 可微的，如果存在一个连续的线性泛函 $\dot{T}(F;\cdot)$，对任意序列 $t_n \to 0$ 和 $\rho(D_n, D) \to 0$ 的函数 D_n，有

$$\frac{T(F + t_n D_n) - T(F) - \dot{T}(F; t_n D_n)}{t_n} \to 0$$

Definition

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

定理 2. 假设 $T : \mathcal{F} \to \mathbb{R}$ 在 F 处相对于 $\|\cdot\|_\infty$ 是 *Hadamard* 可微的，则

$$\sqrt{n}\left(T\left(\mathbb{F}_n\right) - T(F)\right) \to_d N\left(0, E\left(\dot{T}^2\left(F; 1_{(-\infty,\cdot]}(X) - F\right)\right)\right)$$

并且，

$$\sqrt{n}\left(T\left(\mathbb{F}_n\right)-T(F)\right)-\dot{T}\left(F;\sqrt{n}\left(\mathbb{F}_n-F\right)\right)=o_p(1)$$

Proof. 可以用经验过程的 Skorokhod 构造或扩展的连续映射定理来证明，Gill(1989) 使用了 Skorokhod 方法；Wellner(1989) 指出扩展的连续映射证明。Gill(1989) 使用了 Skorokhod 方法；Wellner(1989) 指出了扩展连续映射证明。 □

Fréchet 导数

Definition

一个泛函 T 称为是 ρ-**Fréchet** 可微的，如果存在一个连续的线性泛函 $\dot{T}(F;\cdot)$，对任意满足 $\rho(G_n, F) \to 0$ 的函数序列 G_n，有

$$\frac{T(G_n) - T(F) - \dot{T}(F; G_n - F)}{\rho(G_n, F)} \to 0$$

Fréchet 可导 $\Rightarrow$ Hadamard 可导 $\Rightarrow$ Gateaux 可导

定理 3. 设 ρ 是一种弱收敛度量，则

(1) 如果 $\dot{T}$ 在 *Fréchet* 可导意义下存在，则其是唯一存在的，并且 T 是 *Gateaux* 可导的，导数为 $\dot{T}$.

(2) 如果 $\dot{T}$ 在 F 处是 *Fréchet* 可导的，则 T 在 F 处连续.

(3) $\dot{T}(F; G - F) = \int \psi d(G - F) = \int \left(\psi - \int \psi dF\right) dG$，其中 ψ

是有界连续函数.

证明见 Huber (1981), proposition 5.1, page 37.

定理 4. 如果 T 相对于 ρ 在 F 处是 *Fréchet* 可导的, 满足 $\sqrt{n}\rho(\mathbb{F}_n, F) = O_p(1)$, 则

$$\begin{aligned}\sqrt{n}\left(T\left(\mathbb{F}_n\right) - T(F)\right) &= \int \psi_F d(\sqrt{n}\mathbb{F}_n) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_F\left(X_i\right) + o_p(1) \\ &\to N\left(0, E\psi_F^2(X)\right)\end{aligned}$$

Proof. 由 T 的 Fréchet 可导性, 有

$$\begin{aligned}\sqrt{n}\left(T\left(\mathbb{F}_n\right)-T(F)\right) &= \sqrt{n}\int \psi_F d\mathbb{F}_n+\sqrt{n}o\left(\rho\left(\mathbb{F}_n, F\right)\right) \\ &= \sqrt{n}\int \psi_F d\mathbb{F}_n+\frac{o\left(\rho\left(\mathbb{F}_n, F\right)\right)}{\rho\left(\mathbb{F}_n, F\right)}\sqrt{n}\rho\left(\mathbb{F}_n, F\right) \\ &= \sqrt{n}\int \psi_F d\mathbb{F}_n+o(1)O_p(1)\end{aligned}$$

由可导性定义和假设立得. □

定理 5. 非参数 *Delta* 定理 设统计泛函 $T(F)$ 相对于距离 $\rho(F,G)=\sup_x|F(x)-G(x)|$ 是 *Hadamard* 可微的，则

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

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

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

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

- 泛函 $T(F)$ 的 $1-\alpha$ 逐点渐近置信区间为 $T(\mathbb{F}_n)\pm z_{\alpha/2}\hat{se}$.
- **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

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

■ 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

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

例: M 估计的渐近正态性. 如果 $\Psi(x,t)=\partial\rho(x,t)/\partial t$ 存在且

$$\lambda_F(t)=\int\Psi(x,t)dF(x)=\frac{\partial}{\partial t}\int\rho(x,t)dF(x)$$

则 $T(F)$ 为 $\lambda_F(t)=0$ 的解. 因此,

$$\lambda_F(T(F))=\int\Psi(x,T(F))dF(x)=0 \quad \text{for all } F\in\mathcal{F}_0$$

为了计算 $IF(x,T,F)=\frac{d}{dt}T\left((1-t)F+t\delta_x\right)\big|_{t=0}$, 利用

$$\begin{aligned}
0&=\lambda_{(1-t)F+t\delta_x}\{(1-t)F+t\delta_x\}\\
&=\int\Psi\left(u,T\left((1-t)F+t\delta_x\right)\right)d\left[(1-t)F(u)+t\delta_x(x)\right]\\
&=(1-t)\int\Psi\left(u,T\left((1-t)F+t\delta_x\right)\right)dF(u)+t\Psi\left(x,T\left((1-t)F+t\delta_x\right)\right)
\end{aligned}$$

两边对 t 求导并令 $t=0$ 得到

$$
\begin{aligned}
0 = & -\int \Psi(u, T(F)) dF(u) + \frac{d}{dt} \int \Psi\left(u, T\left((1-t)F + t\delta_x\right)\right) dF(u) \Big|_{t=0} \\
& + \Psi(x, T(F)) + 0 \\
= & -\lambda_F(T(F)) + \frac{d}{dt} \lambda_F\left(T\left((1-t)F + t\delta_x\right)\right)\Big|_{t=0} + \Psi(x, T(F)) \\
= & 0 + \lambda_F'(T(F)) IF(x, T, F) + \Psi(x, T(F))
\end{aligned}
$$

假设 $\lambda_F'(T(F)) \neq 0$, 我们有

$$
IF(x, T, F) = -\Psi(x, T(F)) / \lambda_F'(T(F))
$$

容易验证

$$
\mathrm{E}_F\left[IF(X, T, F)\right] = -\mathrm{E}_F[\Psi(X, T(F))] / \lambda_F'(T(F)) = -\lambda_F(T(F)) / \lambda_F'(T(F)) = 0
$$

和

$$
\sigma^2(F) = \sigma^2(F, \Psi) = \mathrm{Var}_F\left[IF(X, T, F)\right] = \frac{\int \Psi^2(x, T(F)) dF(x)}{\left\{\lambda_F'(T(F))\right\}^2}
$$

从而

$$\sqrt{n}\left[T\left(F_n\right)-T(F)\right]=n^{-1/2}\sum_{i=1}^{n}IF(X_i,T,F)+R_n \xrightarrow{\mathcal{L}} N\left(0,\sigma^2(F,\Psi)\right)$$

只要可以验证 $R_n=o_P(1)$.

■ 对最大似然估计,$\Psi(x,t)=-\partial\log f(x,t)/\partial t$, 令 $F_\theta=F(\cdot,\theta)$, $T\left(F_\theta\right)=\theta$, 则

$$\lambda_{F_\theta}'\left(T\left(F_\theta\right)\right)=\lambda_{F_\theta}'(\theta)=\int\left(-\frac{\partial^2\log f(x,t)}{\partial t^2}\right)_{t=\theta}f(x,\theta)dx=I_f(\theta),\ \text{and}$$
$$\int\Psi^2\left(x,T\left(F_\theta\right)\right)dF_\theta(x)=\int\left(\frac{\partial\log f(x,t)}{\partial t}\right)_{t=\theta}^2 f(x,\theta)dx=I_f(\theta)$$

其中 $I_f(\theta)$ 是分布族 $\{f(x,t),t\in\Theta\}$ 在 $t=\theta$ 时 Fisher 信息量. 因此 $\sigma^2\left(F_\theta,\Psi\right)=1/I_f(\theta)$.

2.1.2 高阶导数

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

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

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

$$T_F^{(m)}(F,G-F)=\int\cdots\int\psi_m(x_1,\ldots,x_m)d(G-F)(x_1)\cdots d(G-F)(x_m)$$

$$= \int \cdots \int \psi_{m,F}(x_1, \ldots, x_m) dG(x_1) \cdots dG(x_m)$$

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

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

从而

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

利用中心极限定理，即可得出极限分布

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

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

- 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

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

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

因此 $T(\mathbb{F}_n)$ 是渐近有效的, 仅当

$$IF^2(x;T,F) = \left\{I\left(F_\theta\right)^{-1}\frac{\partial}{\partial\theta}\left(\ln f_\theta(x)\right)\frac{d}{d\theta}\mathbb{E}_\theta T\right\}^2.$$

- 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

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

Reject point若 F 是对称的 (且对称中心为原点), 则称

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

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 $F_t = (1-t)F + t\delta_x$. 则

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

称为估计量 T 的崩溃点.

■ 例. 考虑统计泛函 $T = T(F) = \int x dF$, 则 $IF(x; T, F) = x - T(F)$. 从而 $\gamma^* = \infty$, $\lambda^* = 1$, $\rho^* = \infty$, $t^* = 0$, 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

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

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

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

定理 6 (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*

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

Proof. 根据条件 A_1 的第二条,

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

□

定理 7 (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

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

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。

Proof. 由条件 A_2 我们有

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

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

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

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

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