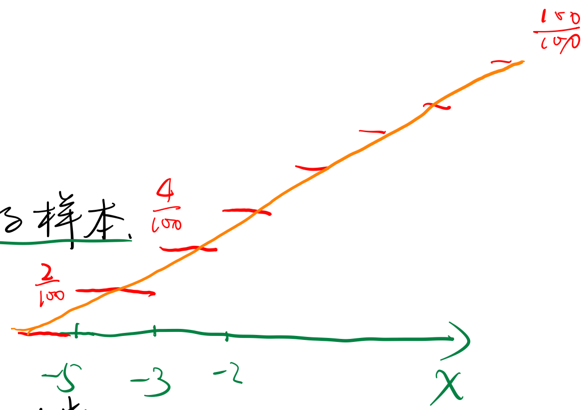
Bootstrap. 鞋带 自力更生. 自抽样



经验分布函数：$\{x_1,\cdots,x_n\}$ 是从 F 中而来的样本. ($\to F$)

$$F_n(x)=\frac{1}{n}\sum_{i=1}^{n}1(x_i\le x),\quad -\infty<x<\infty$$

1：示性函数. $\sum_{i=1}^{n}1(x_i\le x)$ 表示样本中小于等于 x 的个数.

定理：$F_n \xrightarrow{P} F$ 条件：$n\to\infty$. Mean Square Convergence.

$\begin{cases} E(\cdot)\ \text{定值} \\ Var(\cdot)\to 0 \end{cases}$ M.S.C

证明：
$$E[1(x_i\le x)] = 1\cdot Pr(x_i\le x)+0\cdot(1-Pr(x_i\le x))$$
$$= Pr(x_i\le x)$$
$$Var(1(x_i\le x)) = (1-Pr(x_i\le x))^2 Pr(x_i\le x) + (0-Pr(x_i\le x))^2(1-Pr(x_i\le x))$$
$$= p(1-p)[(1-p)+p] = p(1-p)$$
$$E\left[\frac{1}{n}\sum_{i=1}^{n}1(x_i\le x)\right]=\frac{1}{n}np=p=F(x)$$
$$Var\left(\frac{1}{n}\sum_{i=1}^{n}1(x_i\le x)\right)=\frac{1}{n^2}np(1-p)=\frac{1}{n}p(1-p)\to 0$$

故 $F_n\xrightarrow{m.s.}F$，$F_n\xrightarrow{P}F$

情形：统计量很难算.
Sargan.

Bootstrap 思想

用样本算统计量，作为对真实值的估计. 估计误差多大？

例：$\bar{x}_n=\frac{1}{n}\sum_{i=1}^{n}x_i$ 作为均值 μ 的估计. 解析解 无偏，一致

$\bar{x}_n$ 的误差（标准差）是 $se(\bar{x}_n)=\boxed{\frac{\sigma}{\sqrt{n}}}$ 因为

$Var(\bar{x}_n)=\frac{1}{n^2}(n\sigma^2)=\boxed{\frac{\sigma^2}{n}}$ 样本均值的方差.

再用 σ 的估计量，比如 $\hat{\sigma}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2$ 代入 $\frac{\sigma}{\sqrt{n}}$，即可得 $\bar{x}_n$ 的标准差.

样本方差

$\frac{\hat{\sigma}}{\sqrt{n}}$ ✓

但有时事情不这么简单.

F

$\{x_1, \cdots, x_n\}$

g

任给一统计量 $T_n = g(x_1, \cdots, x_n)$，$se(T_n)$ 未必可简单写出.

也即 $\int (T_n - E(T_n))^2 dF(x) \equiv Var_F(T_n)$ 未知.

$\frac{\sigma}{\sqrt{n}}$

用 $Var_{F_n}(T_n)$ 代替？ $\frac{\hat{\sigma}}{\sqrt{n}}$

想法不错，但 $Var_{F_n}(T_n)$ 就能简单写出吗？

n个

我们强行算吧： 有放回 $\{x_2, x_3, x_3, \cdots, x_{15}, \cdots, x_{100}\}$ ②

x_1^* F x_n^*

③

- 从 $\{x_1, \cdots, x_n\}$ ① 中再抽样，也即再次生成 F_n 的一个抽样分布 $\{x_1^*, \cdots, x_n^*\}$.

F_B^*

④

- 让 $\{x_1^*, \cdots, x_n^*\}$ 生成 B 次，$T_n^* = g(x_1^*, \cdots, x_n^*)$ 就有了

B很大 500000

⋮

T分布 $T_{n1}^*, \cdots, T_{nB}^*$. Ⓑ

- $Var_{Boot} = \frac{1}{B} \sum_{b=1}^{B} \left(T_{nb}^* - \frac{1}{B} \sum_{r=1}^{B} T_{nr}^* \right)^2$

接受 只有 $\{x_1, \cdots, x_n\}$

$$Var_F(T_n) \approx Var_{F_n}(T_n) \approx Var_{Boot}(T_n^*)$$

取决于n 取决于B

？

很接近.