大样本理论

司继春

[1]上海对外经贸大学

2024年11月

概览

收敛的概念

收敛的定义

若$\{a_n, n = 1, 2, ...\}$为实数序列，如果对于任意的$\epsilon > 0$，存在$n_0 = n_0(\epsilon)$使得：

$$|a_n - a| < \epsilon, \forall n > n_0$$

那么我们称数列$\{a_n\}$ 的极限为a，或者$\{a_n\}$ 收敛到（converges to）a，记为

$$\lim_{n\to\infty} a_n = a$$

或者

$$a_n \to a \; as \; n \to \infty$$

收敛的概念

有界的定义

若$\{a_n, n = 1, 2, ...\}$为实数序列，如果存在常数$b < \infty$，使得

$$|a_n| < b$$

那么我们称数列$\{a_n\}$为有界的（bounded），否则称之为无界的（unbounded）。

小o符号

小o符号的定义

对于两个序列$\{a_n\}, \{b_n\}$，如果随着$n \to \infty$，有：

$$\frac{a_n}{b_n} \to 0$$

那么我们记为$a_n = o(b_n)$。特别的，如果令$b_n = 1$，那么$a_n = o(1)$等价为$a_n \to 0$。

小o符号

小o实例

如果$a_n = \frac{1}{n^2}, b_n = \frac{1}{n}$，那么:

$$\frac{a_n}{b_n} = \frac{\frac{1}{n^2}}{\frac{1}{n}} = \frac{1}{n} \to 0$$

因而$\frac{1}{n^2} = o\left(\frac{1}{n}\right)$，即$\frac{1}{n^2}$以更快的速度收敛到0。如果两个序列$a_n \to 0, b_n \to 0$, 且$a_n = o\left(b_n\right)$，那么我们称$a_n$为比$b_n$高阶的无穷小量。

小o符号

小o符号经常用来对一个复杂的式子进行化简，通过将无穷小量舍掉从而减少了计算量。比如：

小o的应用

假设有两个数列，$a_n = \frac{1}{n} + \frac{6}{n^2} - \frac{8}{n^3}$而另外一个序列：$b_n = \frac{1}{n}$如果定义$R_n = \frac{6}{n^2} - \frac{8}{n^3}$，显然$R_n = o\left(\frac{1}{n}\right)$，因而$a_n = b_n + o\left(\frac{1}{n}\right)$，即：

$$\frac{a_n}{b_n} = \frac{b_n + o\left(\frac{1}{n}\right)}{b_n} \to 1$$

因而尽管两个序列a_n和b_n并不相等，但是当$n \to \infty$时，两者误差趋向于0，因而我们可以舍去无穷小量R_n，使用更简单的序列b_n去逼近a_n。

小o符号的应用：泰勒展开

当$x \to a$时，$(x-a)=o(1)$，同时我们有$(x-a)^{k+1}=o\left((x-a)^k\right)$，即当$x \to a$时，$(x-a)$的高阶幂是低阶幂的无穷小量。对于一个单变量实值函数$f(x)$且k阶可微，那么有：

$$
\begin{aligned}
f(x) &= f(a)+f'(a)(x-a)+\frac{f''(a)}{2!}(x-a)^2+\cdots \\
&+\frac{f^{(k)}(a)}{k!}(x-a)^k+o\left(|x-a|^k\right)
\end{aligned}
$$

因而对于一个难以计算的函数f，我们经常使用其前k阶泰勒多项式对其进行逼近。

泰勒展开

泰勒展开

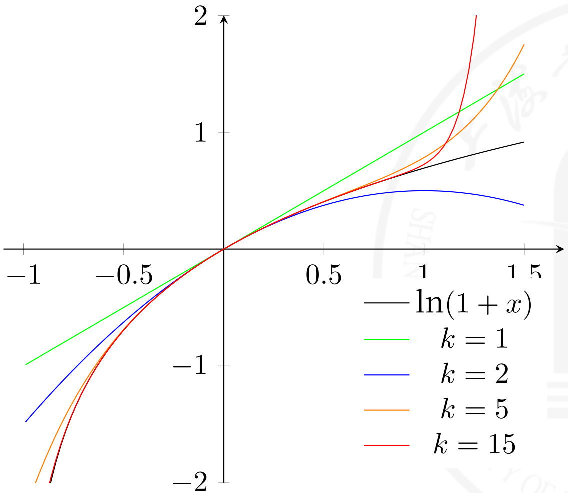
函数$f(x)=\ln(1+x)$在$x=0$处的泰勒展开为:

$$f(x)=\sum_{n=1}^{\infty}(-1)^{n+1}\frac{x^n}{n}=x-\frac{x^2}{2}+\frac{x^3}{3}-\cdots$$

因而当x充分靠近0时，我们可以使用前k阶泰勒展开对其进行逼近。特别的，如果令

$$k=1,\quad \ln(1+x)=x+o(x)\approx x$$

泰勒展开



多元函数的泰勒展开

更一般的，如果$f:\mathbb{R}^n\to\mathbb{R}$为多元实值函数，那么其泰勒级数为：

$$
\begin{aligned}
f(x)=f(a)+\frac{\partial f}{\partial x'}(a)(x-a)+\frac{1}{2!}(x-a)'\frac{\partial^2 f}{\partial x\partial x'}(a)(x-a)\\
+o\left(\|x-a\|_2^2\right)
\end{aligned}
$$

其中x和a为$n\times 1$向量。

多元函数的泰勒展开

多元函数泰勒展开示例

令$f(x) = e^{x_1}\ln(1+x_2)$，其中$x = (x_1, x_2)'$。那么:

$$\frac{\partial f}{\partial x} = \begin{bmatrix} e^{x_1}\ln(1+x_2) \\ \frac{e^{x_1}}{1+x_2} \end{bmatrix}, \frac{\partial^2 f}{\partial x \partial x'} = \begin{bmatrix} e^{x_1}\ln(1+x_2) & \frac{e^{x_1}}{1+x_2} \\ \frac{e^{x_1}}{1+x_2} & -\frac{e^{x_1}}{(1+x_2)^2} \end{bmatrix}$$

因而其在$a = (0,0)'$处的二阶泰勒展开:

$$\begin{aligned} p_2(x) &= [0,1]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{2}[x_1, x_2]\begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= x_2 + \frac{1}{2}\left(2x_1x_2 - x_2^2\right) \end{aligned}$$

小o符号

小o的性质

1. 若$a_n = o(b_n), b_n = o(c_n)$，那么$a_n = o(c_n)$
2. 对于任意的常数$c \neq 0$，及$a_n = o(b_n)$，有$ca_n = o(b_n)$
3. 对于任意的数列$c_n \neq 0$，及$a_n = o(b_n)$，有$c_n a_n = o(c_n b_n)$
4. 如果$d_n = o(b_n), e_n = o(c_n)$，那么$d_n e_n = o(b_n c_n)$
5. 如果$a_n, b_n > 0, c_n, d_n > 0$，$a_n = o(b_n), c_n = o(d_n)$，那么$a_n + c_n = o(b_n + d_n)$。

大O符号

大O符号的定义

对于两个序列$\{a_n\},\{b_n\}$，如果随着$n\to\infty$，$\left|\frac{a_n}{b_n}\right|$是有界的，即存在一个M使得:

$$\left|\frac{a_n}{b_n}\right|<M$$

那么我们记为$a_n=O(b_n)$。特别的，如果令$b_n=1$，那么$a_n=O(1)$等价为a_n是有界的。

同阶的定义

对于两个序列$\{a_n\},\{b_n\}$，如果$a_n=O(b_n)$，且同时$b_n=O(a_n)$，那么我们称两个序列是同阶的，简记为$a_n\asymp b_n$。

大O符号

大O符号示例

对于序列$a_n = \frac{1}{n} + \frac{b}{n\sqrt{n}} + \frac{c}{n^2} + \frac{d}{n^2\sqrt{n}}$，同时定义$R_n = \frac{b}{n\sqrt{n}} + \frac{c}{n^2} + \frac{d}{n^2\sqrt{n}}$那么：

1. $a_n \sim \frac{1}{n}$
2. 若$b = 0$，$R_n = O\left(\frac{1}{n^2}\right)$
3. 若$b = 0$，$R_n \asymp \frac{1}{n^2}$
4. 若$b \neq 0$，$R_n \sim \frac{b}{n\sqrt{n}}$
5. 若$b = c = 0$，$R_n = o\left(\frac{1}{n^2}\right)$

大O符号

大O符号的性质

1. 若$a_n = O(b_n), b_n = O(c_n)$，那么$a_n = O(c_n)$
2. 对于任意的常数$c \neq 0$，及$a_n = O(b_n)$，有$ca_n = O(b_n)$
3. 对于任意的数列$c_n \neq 0$，及$a_n = O(b_n)$，有$c_n a_n = O(c_n b_n)$
4. 如果$d_n = O(b_n), e_n = O(c_n)$，那么$d_n e_n = O(b_n c_n)$
5. 如果$a_n = o(b_n)$，$c_n = O(b_n)$，那么$a_n c_n = o(b_n)$
6. 如果$a_n = o(b_n)$，$c_n = O(b_n)$，那么$a_n + c_n = O(b_n)$

依概率收敛

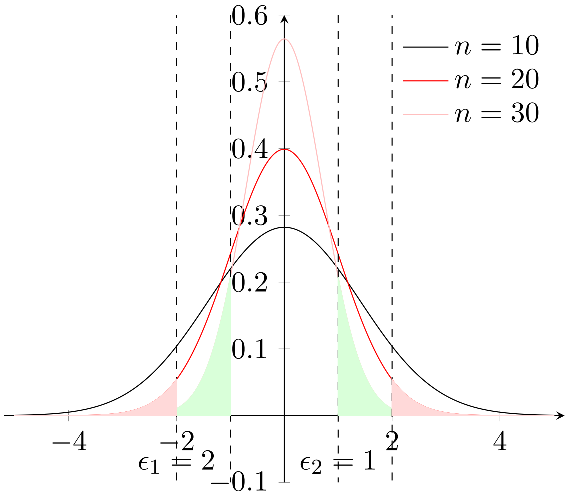
依概率收敛

如果对于任意的$\epsilon > 0$，概率空间$(\Omega, \mathscr{F}, \mathscr{P})$上的一系列随机变量序列$\{X_n\}$ 满足：

$$P\left(|X_n - X| > \epsilon\right) \to 0$$

那么我们称X_n依概率收敛于X，记为$X_n \xrightarrow{p} X$，或$\text{plim}X_n = X$。

依概率收敛



o_p符号

o_p符号

$\{X_n\}$与$\{Y_n\}$为定义在概率空间$(\Omega, \mathscr{F}, \mathscr{P})$上的两个随机变量序列，如果

$$\frac{X_n}{Y_n} \xrightarrow{p} 0$$

那么我们记为$X_n = o_p(Y_n)$。特别的，当$Y_n = 1$时，即$X_n = o_p(1)$，等价于$X_n \xrightarrow{p} 0$。

依概率收敛

简单的大数定律

现在假设有一组独立同分布的样本$x_i, i=1,2,...,N$，其方差有界：$\mathbb{V}(x_i)<M$，其样本均值为$\bar{x}_N$。根据切比雪夫不等式，随着样本量$N\rightarrow\infty$，有：

$$
\begin{aligned}
P\left(\left|\bar{x}_N-\mathbb{E}\left(x_i\right)\right|>\epsilon\right) &\leq \frac{\mathbb{E}\left[\left(\bar{x}_N-\mathbb{E}\left(x_i\right)\right)^2\right]}{\epsilon^2} \\
&= \frac{\mathbb{V}\left(\bar{x}_N\right)}{\epsilon^2}=\frac{\frac{\mathbb{V}(x_i)}{N}}{\epsilon^2} \\
&< \frac{M}{N\epsilon^2}\rightarrow 0
\end{aligned}
$$

从而$\bar{x}_N \xrightarrow{p} \mathbb{E}\left(x_i\right)$，或者$\bar{x}_N-\mathbb{E}\left(x_i\right)=o_p\left(1\right)$，或者$\bar{x}_N=\mathbb{E}\left(x_i\right)+o_p\left(1\right)$。

O_p符号

- 类似的，我们还可以定义大O_p符号。
- 与数列有界不同的是，随机变量的取值范围可能是$\mathbb{R}$，从而一定要求随机变量取值在一个有限的范围内是不可能的。
    - 比如，如果$X \sim \mathrm{N}(0,1)$，虽然X的分布范围看起来很小，但是$\mathrm{supp}(X) = \mathbb{R}$意味着其“理论上”的取值范围仍然是$(-\infty, \infty)$
    - 所以根据$|X_n| < M$这样定义有界显然是不合适的。

O_p符号

- 对于一个随机变量X，如果对于任意的$\epsilon > 0$，都能找到一个C_ϵ使得$P\left(|X| > C_\epsilon\right) < \epsilon$，那么我们就可以认为随机变量$X$是有界的。
- 根据这一定义，我们可以将其推广到随机变量序列上：

O_p符号

$\{X_n\}$与$\{Y_n\}$为定义在概率空间$(\Omega, \mathscr{F}, \mathscr{P})$上的两个随机变量序列，如果对于任意的$\epsilon > 0$，存在一个$C_\epsilon$使得：

$$\sup_n P\left(|X_n| \geq C_\epsilon \, |Y_n|\right) < \epsilon$$

那么我们记$X_n = O_p\left(Y_n\right)$。特别的，当$Y_n = 1$时，我们称$X_n$**依概率有界（bounded in probability）**。

O_p符号

- O_p符号是对O符号的推广，当X_n退化为确定性序列时，O_p符号就变成了O符号。
- 相对于随机变量的有界，随机变量序列的依概率有界需要对于所有的$\{X_n\}$找到一个共同的C_ϵ：
  - 如果$X_n = O_p(1)$，那么意味着$|X_n|$不能太大，我们总是可以以很高的概率（$1-\epsilon$）保证对于任意的n，都有$|X_n| < C_\epsilon$成立即可。
- 如果$\{X_n\}$的期望趋向于正无穷，那么自然不是依概率有界的。

O_p符号

- 注意如果$\mathbb{E}\left(X_n^2\right) < M$，对于任意的$\epsilon > 0$，取$C_\epsilon = \sqrt{M/\epsilon + 1}$，那么：

$$P\left(|X_n| \geq C_\epsilon\right) \leq \frac{\mathbb{E}\left(X_n^2\right)}{C_\epsilon^2} = \frac{\mathbb{E}\left(X_n^2\right)}{M/\epsilon + 1} < \epsilon$$

  因而$X_n = O_p(1)$。
- 如果$\mathbb{E}\left|X_n\right| < M_0$，那么$\mathbb{E}\left(X_n^2\right) < M$意味着$\mathbb{V}\left(X_n\right) = \mathbb{E}\left(X_n^2\right) - \mathbb{E}\left(X_n^2\right) < M + M_0$从而方差有界
- 从而$\{X_n\}$的期望、方差同时有界意味着$X_n = O_p(1)$。
- 实际上，只需要$\mathbb{E}\left|X_n\right| < M$即可（why?）。

依概率有界

样本均值的阶数

现在假设有一组独立同分布的样本$x_i, i=1,2,...,N$，其方差有界：$\mathbb{V}\left(x_i\right)<M$，其样本均值为$\bar{x}_N$。可以计算：

$$\mathbb{V}\left(\bar{x}_N\right)=\frac{\mathbb{V}\left(x_i\right)}{N}$$

因而$\bar{x}_N$的方差是有界的，$\mathbb{V}\left(\bar{x}_N\right)=O_p\left(1\right)$。此外：

$$\mathbb{V}\left(\sqrt{N}\bar{x}_N\right)=\mathbb{V}\left(x_i\right)$$

因而$\sqrt{N}\bar{x}_N=O_p\left(1\right)$。但是，如果我们考虑样本和：$S_N=\sum_{i=1}^N x_i$由于$\mathbb{V}\left(S_N\right)=N\mathbb{V}\left(x_i\right)$，该方差是无界的，因而不是$O_p\left(1\right)$。

O_p和o_p符号的性质

O_p和o_p的性质

如果$X_n = o_p(1), Y_n = o_p(1), Z_n = O_p(1), W_n = O_p(1)$，那么:

1. $X_n + Y_n = o_p(1)$
2. $X_n + Z_n = O_p(1)$
3. $Z_n + W_n = O_p(1)$
4. $X_nY_n = o_p(1)$
5. $X_nZ_n = o_p(1)$
6. $Z_nW_n = O_p(1)$

均方收敛

均方收敛的定义

如果概率空间$(\Omega, \mathscr{F}, \mathscr{P})$上的一系列随机变量序列$\{X_n\}$随着$n \to \infty$满足：

$$\mathbb{E}\left(|X_n - X|^2\right) \to 0$$

那么我们称X_n均方收敛于X，记为$X_n \xrightarrow{L^2} X$。

均方收敛

均方收敛与依概率收敛

如果随机变量序列$X_n \xrightarrow{L^2} X$，那么$X_n \xrightarrow{p} X$。

Proof.

据切比雪夫不等式，对于任意的$\epsilon > 0$，有：

$$P(|X_n - X| > \epsilon) \le \frac{\mathbb{E}\left(|X_n - X|^2\right)}{\epsilon^2} \to 0$$

$\square$

依分布收敛

依分布收敛的定义

令F_n,F为分布函数，如果对于每一个$F(x)$连续的点x，有：

$$\lim_{n\to\infty} F_n(x) = F(x)$$

那么我们称$F_n(x)$弱收敛于$F(x)$，记为$F_n \xrightarrow{w} F$。

如果一系列随机变量$\{X_n\}$的分布函数$F_{X_n}(x) \xrightarrow{w} F_X$，我们称$X_n$依分布收敛于$X$，记为$X_n \xrightarrow{D} X$或者：$X_n \overset{a}{\sim} F$，其中$a$代表渐近的（asymptotically），即$X_n$渐近服从分布函数为$F$的分布。

依分布收敛

依分布收敛与O_p的关系

1. 如果$X_n \neq O_p(1)$，那么分布函数极限$\lim_{n\to\infty} F_n(x) = F(x)$不是一个分布函数。
2. 如果X_n依分布收敛，那么$X_n = O_p(1)$。

因而当我们讨论依分布收敛时，一定要保证我们讨论的$X_n = O_p(1)$。

大数定律

大数定律（Law of Large Numbers, LLN）讨论样本均值的极限，即在何种条件下，以下结论：

$$\frac{S_N - \mathbb{E}(S_N)}{N} = \bar{x} - \mathbb{E}(\bar{x}) \xrightarrow{p} 0$$

成立，其中：

$$S_N = \sum_{i=1}^{N} x_i$$

大数定律

实际上，$S_N - \mathbb{E}(S_N) = \sum_{i=1}^{N}[x_i - \mathbb{E}(x_i)]$，因而：

$$\begin{aligned}
\mathbb{E}[S_N - \mathbb{E}(S_N)]^2 &= \mathbb{E}\left(\sum_{i=1}^{N}[x_i - \mathbb{E}(x_i)]\right)^2 \\
&= \mathbb{E}\left(\sum_{i=1}^{N}[x_i - \mathbb{E}(x_i)]^2\right) \\
&+ \mathbb{E}\left(2\sum_{1\leq j<i\leq N}[x_i - \mathbb{E}(x_i)][x_j - \mathbb{E}(x_j)]\right) \\
&= \sum_{i=1}^{N}\mathbb{V}(x_i) + 2\sum_{1\leq j<i\leq N}\mathbb{C}(x_i, x_j)
\end{aligned}$$

大数定律

$$\mathbb{E}\left[S_N-\mathbb{E}\left(S_N\right)\right]^2=\sum_{i=1}^{N}\mathbb{V}\left(x_i\right)+2\sum_{1\leq j<i\leq N}^{N}\mathbb{C}\left(x_i,x_j\right)$$

如果我们假设$\mathbb{V}\left(x_i\right)<M$，那么：

1. $\sum_{i=1}^{N}\mathbb{V}\left(x_i\right)<NM=O\left(N\right)$
2. 而根据Cauthy-Schwartz不等式，$\mathbb{C}\left(x_i,x_j\right)\leq\sqrt{\mathbb{V}\left(x_i\right)\mathbb{V}\left(x_j\right)}\leq M$，而上式中有$\frac{N(N-1)}{2}$个协方差，所以

$$\sum_{1\leq j<i\leq N}^{n}\mathbb{C}\left(x_i,x_j\right)=O\left(N^2\right)$$

大数定律

$$\mathbb{E}\left[S_N-\mathbb{E}\left(S_N\right)\right]^2=O\left(N\right)+O\left(N^2\right)$$

从而:

$$\mathbb{E}\left[\frac{S_N-\mathbb{E}\left(S_N\right)}{N}\right]^2=\frac{1}{N^2}O\left(N\right)+\frac{1}{N^2}O\left(N^2\right)=o\left(1\right)+O\left(1\right)$$

根据均方收敛定义，如果希望$\frac{S_N}{N}-\frac{\mathbb{E}(S_N)}{N}\xrightarrow{L^2}0$，必须

$$\mathbb{E}\left[\frac{S_N-\mathbb{E}\left(S_N\right)}{N}\right]^2=o\left(1\right)$$

大数定律

$$\mathbb{E}\left[\frac{S_N-\mathbb{E}\left(S_N\right)}{N}\right]^2=o\left(1\right)+O\left(1\right)$$

其中的$O\left(1\right)$来源于$\frac{N(N-1)}{2}$个协方差，如果$\mathbb{C}\left(x_i,x_j\right)=0$，那么自然有：

$$\mathbb{E}\left[\frac{S_N-\mathbb{E}\left(S_N\right)}{N}\right]^2=o\left(1\right)$$

从而

$$\frac{S_N}{N}-\frac{\mathbb{E}\left(S_N\right)}{N}\xrightarrow{L^2}0\Rightarrow\frac{S_N}{N}-\frac{\mathbb{E}\left(S_N\right)}{N}\xrightarrow{p}0$$

大数定律

大数定律1

如果概率空间$(\Omega, \mathscr{F}, \mathscr{P})$上的一个随机变量序列$\{x_i\}$两两不相关，且存在一个$M$使得对于所有的$i = 1, 2, ...,$ 都有$\mathbb{V}(x_i) < M$，那么：

$$\bar{x} - \mathbb{E}(\bar{x}) = \frac{S_N - \mathbb{E}(S_N)}{N} \xrightarrow{L^2} 0$$

从而

$$\bar{x} - \mathbb{E}(\bar{x}) = \frac{S_N - \mathbb{E}(S_N)}{N} \xrightarrow{p} 0$$

如果额外假设$\{x_i\}$是同分布的，那么$\mathbb{E}(\bar{x}) = \mathbb{E}(x_i) = \mu$，从而：

$$\bar{x} \xrightarrow{p} \mu$$

大数定律

以下的定理放松了二阶矩有限的假定以及独立的假定，保留了独立同分布的假定：

大数定律2

令$\{x_i\}$为概率空间$(\Omega, \mathscr{F}, \mathscr{P})$上的两两独立且同分布的随机变量序列，若

$$\mathbb{E}\left|x_i\right| < \infty$$

那么

$$S_N/N = \bar{x} \xrightarrow{p} \mu$$

其中$\mu = \mathbb{E}\left(x_i\right)$。

大数定律

以下的定理则同时放宽了同分布的假定以及二阶矩的假定。

大数定律3

令$\{x_i\}$为概率空间$(\Omega, \mathscr{F}, \mathscr{P})$上的相互独立的随机变量序列，如果存在一个常数$p \in [1, 2]$，随着$N \to \infty$，使得：

$$\frac{1}{N^p} \sum_{i=1}^{N} \mathbb{E} |x_i|^p \to 0$$

那么$S_N / N = \bar{x} \xrightarrow{p} \mu$，其中

$$\mu = \lim_{N \to \infty} \frac{\sum_{i=1}^{N} \mathbb{E}(x_i)}{N}$$

大数定律

大数定律的应用

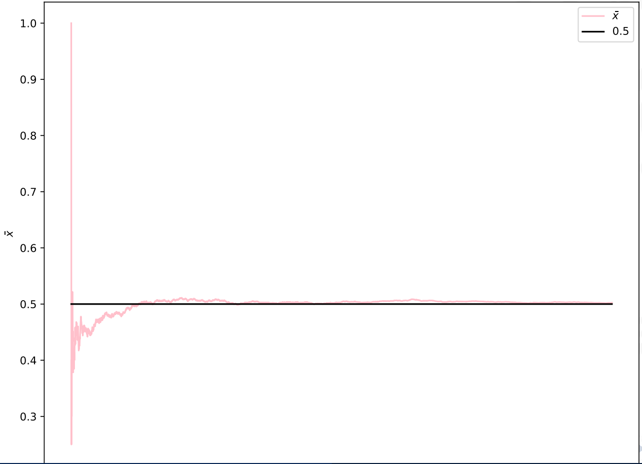
如果令$\{x_i\}$为一系列i.i.d的随机变量，且$x_i \sim Ber\left(p\right)$，那么$\mathbb{E}\left(x_i\right) = p, \mathbb{V}\left(x_i\right) = p\left(1-p\right) < \infty$，定义：

$$\hat{p} = \frac{\sum_{i=1}^{N} x_i}{N}$$

即成功的比例，那么根据上例，可以得到

$$\hat{p} \xrightarrow{p} p$$

大数定律



中心极限定理

中心极限定理（Central Limit Theorem, CLT）讨论样本均值的极限分布，即在大样本条件下，样本均值的分布情况，通常中心极限定理可以得到如下结论：

$$\sqrt{N}\left(\bar{x}-\mu\right)=\frac{S_N-N\mu}{\sqrt{N}}\overset{a}{\sim}N\left(0,\mathbb{V}\left(x_i\right)\right)$$

其中：

$$S_N=\sum_{i=1}^{N}x_i$$

即样本均值的极限分布为正态分布。

中心极限定理

前面我们提到，如果希望讨论$\sqrt{N}\left(\bar{x}-\mu\right)$的极限分布，需要保证$\sqrt{N}\left(\bar{x}-\mu\right)=O_p\left(1\right)$，实际上：如果假设$\{x_i\}$之间两两不相关，那么：

$$
\begin{aligned}
\mathbb{V}\left(S_N\right) &= \mathbb{E}\left[S_N-\mathbb{E}\left(S_N\right)\right]^2 \\
&= \mathbb{E}\left[\sum_{i=1}^{n}\left(X_i-\mathbb{E}\left(X_i\right)\right)\right]^2 \\
&= \sum_{i=1}^{n}\mathbb{V}\left(X_i\right) \\
&= O\left(N\right)
\end{aligned}
$$

从而$\mathbb{V}\left(\sqrt{N}\frac{S_N}{N}\right)=O\left(1\right)$，或者其方差有界，因而$\sqrt{N}\frac{S_N}{N}=\sqrt{N}\bar{x}=O_p\left(1\right)$。

中心极限定理

中心极限定理（标量）

令$\{x_i\}$为概率空间$(\Omega, \mathscr{F}, \mathscr{P})$上$i.i.d$的随机变量序列，且$\mathbb{E}\left(x_i\right) = \mu, \mathbb{V}\left(x_i\right) = \sigma^2$，那么：

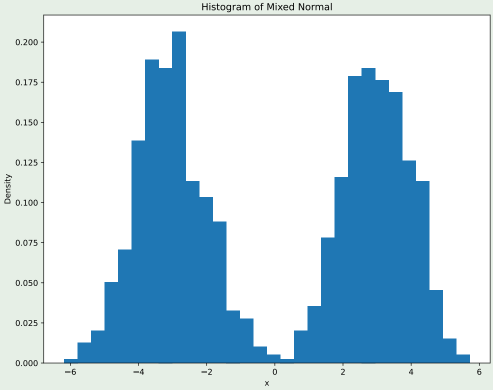
$$\sqrt{N}\left(\bar{x}_N - \mu\right) \overset{a}{\sim} \mathrm{N}\left(0, \sigma^2\right)$$

或：

$$\sqrt{N}\left(\frac{\bar{x}_N - \mu}{\sigma}\right) \overset{a}{\sim} \mathrm{N}\left(0, 1\right)$$

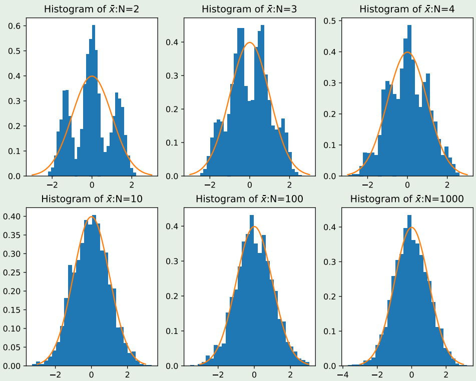
中心极限定理

x_i的分布



中心极限定理

$\sqrt{N}\left(\bar{x}-\mu\right)/\sigma$的分布



中心极限定理

中心极限定理示例

如果$\{x_i\}$为$i.i.d$的随机变量，且$x_i \sim \mathrm{Ber}\left(p\right)$，令$\hat{p}_N$如前定义，那么：

$$\sqrt{N}\left(\hat{p}_N - p\right) \overset{a}{\sim} \mathrm{N}\left(0, p\left(1-p\right)\right)$$

如果$\{x_i\}$为$i.i.d$的随机变量，且$x_i \sim \mathrm{N}\left(0,1\right)$，那么可知$\mathbb{E}\left(x_i^2\right) = 1, \mathbb{E}\left(x_i^4\right) = 3$，因而：

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} x_i^2 - 1\right) \overset{a}{\sim} \mathrm{N}\left(0, 2\right)$$

中心极限定理

中心极限定理（向量）

令$\{x_i\}$为概率空间$(\Omega, \mathscr{F}, \mathscr{P})$上$i.i.d$的随机向量序列，且$\mathbb{E}(x_i) = \mu, \mathbb{V}(x_i) = \Sigma$，那么：

$$\sqrt{N}(\bar{x}_N - \mu) \overset{a}{\sim} \mathrm{N}(0, \Sigma)$$

或：

$$\sqrt{N}\Sigma^{-\frac{1}{2}}(\bar{x}_N - \mu) \overset{a}{\sim} \mathrm{N}(0, I)$$

随机变量连续函数的收敛

连续函数的收敛

令$\{X_i\}$为k维随机向量，$g(x):\mathbb{R}^k\rightarrow\mathbb{R}^l$为连续函数，那么：

1. $X_n\stackrel{\text{a.s.}}{\rightarrow}X\Rightarrow g(X_n)\stackrel{\text{a.s.}}{\rightarrow}g(X)$
2. $X_n\stackrel{p}{\rightarrow}X\Rightarrow g(X_n)\stackrel{p}{\rightarrow}g(X)$
3. $X_n\stackrel{D}{\rightarrow}X\Rightarrow g(X_n)\stackrel{D}{\rightarrow}g(X)$

随机变量连续函数的收敛

样本相关系数的极限

对于二维随机向量(x_i, y_i)，令(x_i, y_i)为i.i.d的样本，那么在可积性条件下，

$$\begin{cases} \frac{1}{N}\sum_{i=1}^{N} x_i \xrightarrow{p} \mathbb{E}(x_i) & \frac{1}{N}\sum_{i=1}^{N} y_i \xrightarrow{p} \mathbb{E}(y_i) \\ \frac{1}{N}\sum_{i=1}^{N} x_i y_i \xrightarrow{p} \mathbb{E}(x_i y_i) & \\ \frac{1}{N}\sum_{i=1}^{N} x_i^2 \xrightarrow{p} \mathbb{E}(x_i^2) & \frac{1}{N}\sum_{i=1}^{N} y_i^2 \xrightarrow{p} \mathbb{E}(y_i^2) \end{cases}$$

从而

$$\frac{\frac{1}{N}\sum_{i=1}^{N} x_i y_i - \frac{1}{N}\sum_{i=1}^{N} x_i \frac{1}{N}\sum_{i=1}^{N} y_i}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N} y_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} y_i\right)^2}} \xrightarrow{p} \mathbb{C}\mathrm{orr}(x_i, y_i)$$

Slutsky定理

Slutsky定理

如果随机变量$X_n \xrightarrow{D} X$，$R_n = o_p(1)$，那么

$$X_n + R_n \xrightarrow{D} X$$

同时如果$Y_n \xrightarrow{p} a \neq 0$，那么

$$\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{a}$$

如果$Y_n \xrightarrow{p} a$，那么

$$X_n Y_n \xrightarrow{D} aX$$

Slutsky定理

t统计量的大样本分布

之前曾讨论过，如果$x_i \sim \mathrm{N}\left(\mu,\sigma^2\right)$ $i.i.d$，那么

$$\frac{\sqrt{N}\left(\bar{x}-\mu\right)}{\sqrt{\frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}{N-1}}} \sim \mathrm{t}\left(N-1\right)$$

现在我们不假设x_i服从正态分布，而是假设其独立同分布且具有有限的二阶矩，那么我们有$\bar{x} \xrightarrow{p} \mathbb{E}\left(x_i\right), \frac{1}{N}\sum_{i=1}^{n} x_i^2 \xrightarrow{p} \mathbb{E}\left(x_i^2\right)$

Slutsky定理

t统计量的大样本分布

因而：

$$\begin{aligned}\frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}{N-1} &= \frac{\sum_{i=1}^{N} x_i^2}{N-1}-\frac{N}{N-1}\bar{x}^2\\ &= \frac{\sum_{i=1}^{N} x_i^2-N\bar{x}^2}{N-1}\\ &\xrightarrow{p} \mathbb{E}\left(x_i^2\right)-\left[\mathbb{E}\left(x_i\right)\right]^2=\mathbb{V}\left(x_i\right)\end{aligned}$$

进而：

$$\frac{\sqrt{N}\left(\bar{x}-\mu\right)}{\sqrt{\frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}{N-1}}} \xrightarrow{p} \frac{\sqrt{N}\left(\bar{x}-\mu\right)}{\sqrt{\mathbb{V}\left(x_i\right)}}=\sqrt{N}\left(\frac{\bar{x}-\mu}{\sqrt{\mathbb{V}\left(x_i\right)}}\right) \overset{a}{\sim} N(0,1)$$

delta方法

- 中心极限定理和Slutsky定理相结合已经可以解决很多统计量的渐近分布问题，然而还有更多的统计量的分布问题无法简单套用以上两个定理。
- 这其中，有一类问题是比较特殊的，即如果我们已知一个统计量$T(x)$的分布，那么$T(x)$的一个函数$g[T(x)]$的分布是什么呢？在一些简单情况下，这个问题相对比较容易回答，比如：

对数正态

如果$x_i \sim \mathrm{N}(\mu, \sigma^2)$ $i.i.d$，那么$\bar{x} \sim \mathrm{N}(\mu, \sigma^2/N)$。现在，如果我们需要研究$e^{\bar{x}}$的分布，显然有$e^{\bar{x}} \sim \mathrm{LN}(\mu, \sigma^2/N)$。如果放弃正态性假设，即假设$x_i \sim (\mu, \sigma^2)$ $i.i.d$，那么$\bar{x} \overset{a}{\sim} \mathrm{N}(\mu, \sigma^2/N)$，从而$e^{\bar{x}} \overset{a}{\sim} \mathrm{LN}(\mu, \sigma^2/N)$。

delta方法

然而，以上指数函数是一个非常特殊的函数形式，如果$g(x)$的函数形式比较复杂，比如$g(x)=\sin x$，此时并没有可以直接使用的简单结论，为了解决这类问题，一个比较广泛适用的方法是使用**delta方法（delta method）**。

- 如果有一个随机变量序列$\{X_n\}$（如$X_n=T(x)=\bar{x}$），假设$a_n(X_n-c)\xrightarrow{D}Y$，$\lim_{a_n\to\infty}a_n=\infty$，那么

$$a_n(X_n-c)=O_p(1),X_n-c=o_p(1)$$

delta方法

- 对于任意的连续二阶可微的函数$g(x)$，都可以对其泰勒展开：

$$
\begin{aligned}
a_n[g(X_n)-g(c)] &= \frac{\partial g}{\partial x'}(c)\,a_n(X_n-c)+a_nO\left(\|X_n-c\|^2\right)\\
&= \frac{\partial g}{\partial x'}(c)\,a_n(X_n-c)+O\left(a_n\|X_n-c\|^2\right)\\
&= \frac{\partial g}{\partial x'}(c)\,a_n(X_n-c)+O\left(O_p(1)\cdot o_p(1)\right)\\
&= \frac{\partial g}{\partial x'}(c)\,a_n(X_n-c)+o_p(1)\\
&\xrightarrow{D} \frac{\partial g}{\partial x'}(c)\,Y
\end{aligned}
$$

因而

$$
a_n[g(X_n)-g(c)] \xrightarrow{D} \frac{\partial g}{\partial x'}(c)\,Y
$$

delta方法

- 特别的，如果$Y\sim \mathrm{N}\left(0,\Sigma\right)$，那么：

$$a_n\left[g\left(X_n\right)-g\left(c\right)\right]\overset{a}{\sim}\mathrm{N}\left(0,\frac{\partial g}{\partial x'}\left(c\right)\Sigma\frac{\partial g}{\partial x}\left(c\right)\right)$$

- 通常，$a_N=\sqrt{N}$，而Y一般为根据中心极限定理得到的正态分布及其衍生分布。

delta方法

delta方法示例

令$\{x_i\}$为概率空间$(\Omega, \mathscr{F}, \mathscr{P})$上$i.i.d$的随机变量序列，且$\mathbb{E}\left(x_i\right)=\mu, \mathbb{V}\left(x_i\right)=\sigma^2$，那么根据中心极限定理：

$$\sqrt{N}\left(\bar{x}-\mu\right) \overset{a}{\sim} \mathrm{N}\left(0, \sigma^2\right)$$

delta方法

delta方法示例

如果我们关心$Y=\exp(\bar{x})$的分布，那么可以对其进行泰勒展开：

$$
\begin{aligned}
\sqrt{N}Y-\sqrt{N}\exp(\mu) =&\sqrt{N}\exp(\bar{x})-\sqrt{N}\exp(\mu)\\
=&\sqrt{N}\exp(\mu)(\bar{x}-\mu)+\frac{1}{2}\sqrt{N}\exp(\mu)(\bar{x}-\mu)^2+\cdots\\
=&\sqrt{N}\exp(\mu)(\bar{x}-\mu)+\frac{1}{2}\exp(\mu)O_p(1)\,o_p(1)\\
=&\sqrt{N}\exp(\mu)(\bar{x}-\mu)+o_p(1)\\
\xrightarrow{D}&\sqrt{N}\exp(\mu)(\bar{x}-\mu)\\
\overset{a}{\sim}&\mathrm{N}\left(0,\exp(2\mu)\sigma^2\right)
\end{aligned}
$$

因而$Y\overset{a}{\sim}\mathrm{N}\left(e^{\mu},\frac{\exp(2\mu)\sigma^2}{N}\right)$。

样本方差

样本方差的大样本分布

对于一个随机样本$x_i \sim (\mu, \sigma^2)$ $i.i.d$，样本方差：

$$s^2 = \frac{N}{N-1}\left(\overline{x^2} - \bar{x}^2\right)$$

假设$\mathbb{E}\left(x_i^k\right) = m_k$为原点矩，由于样本减去常数计算样本方差完全等价，因而我们可以对每个样本进行中心化操作，即令

$$y_i = x_i - m_1$$

那么样本方差应该等于

$$s^2 = \frac{N}{N-1}\left(\overline{y^2} - \bar{y}^2\right) \triangleq \frac{N}{N-1}\phi\left(\bar{y}, \overline{y^2}\right)$$

样本方差

样本方差的大样本分布

$$\begin{aligned}\frac{N}{N-1}\phi\left(\bar{y},\overline{y^2}\right) \triangleq \frac{N}{N-1}\left(\overline{y^2}-\bar{y}^2\right) &= \frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})^2\right] \\ &= \frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}(x_i-m_1-\bar{x}+m_1)^2\right] \\ &= \frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2\right] \\ &= s^2\end{aligned}$$

样本方差

样本方差的大样本分布

记$\mu_k=\mathbb{E}\left(y_i^k\right)=\mathbb{E}\left[\left(x_i-m_1\right)^k\right]$为中心距，由于

$$
\begin{aligned}
\mathbb{V}\left(\begin{bmatrix} y_i \\ y_i^2 \end{bmatrix}\right) &= \mathbb{E}\left\{\left(\begin{bmatrix} y_i \\ y_i^2 \end{bmatrix}-\begin{bmatrix} 0 \\ \mu_2 \end{bmatrix}\right)\left(\begin{bmatrix} y_i \\ y_i^2 \end{bmatrix}-\begin{bmatrix} 0 \\ \mu_2 \end{bmatrix}\right)'\right\} \\
&= \mathbb{E}\left(\begin{bmatrix} y_i \\ y_i^2-\mu_2 \end{bmatrix}\begin{bmatrix} y_i \\ y_i^2-\mu_2 \end{bmatrix}'\right) \\
&= \mathbb{E}\left(\begin{bmatrix} y_i^2 & y_i^3-\mu_2 y_i \\ y_i^3-\mu_2 y_i & y_i^4+\mu_2^2-2\mu_2 y_i^2 \end{bmatrix}\right) \\
&= \begin{bmatrix} \mu_2 & \mu_3 \\ \mu_3 & \mu_4-\mu_2^2 \end{bmatrix}
\end{aligned}
$$

样本方差

样本方差的大样本分布

根据中心极限定理:

$$\sqrt{N}\left[\begin{pmatrix} \bar{y} \\ \overline{y^2} \end{pmatrix} - \mathbb{E}\begin{pmatrix} \bar{y} \\ \overline{y^2} \end{pmatrix}\right] = \sqrt{N}\left[\begin{pmatrix} \bar{y} \\ \overline{y^2} \end{pmatrix} - \begin{pmatrix} 0 \\ \mu_2 \end{pmatrix}\right]$$

$$\stackrel{a}{\sim} \mathrm{N}\left(0, \begin{bmatrix} \mu_2 & \mu_3 \\ \mu_3 & \mu_4 - \mu_2^2 \end{bmatrix}\right)$$

样本方差

样本方差的大样本分布

进行泰勒展开，得到：

$$
\begin{aligned}
\sqrt{N}\left(\phi\left(\bar{y},\overline{y^2}\right)-\phi\left(0,\mu_2\right)\right) &= \sqrt{N}\begin{bmatrix} -2\mu_1 & 1 \end{bmatrix}\begin{bmatrix} \bar{y} \\ \overline{y^2}-\mu_2 \end{bmatrix}+o_p\left(1\right) \\
&\xrightarrow{p} \begin{bmatrix} 0 & 1 \end{bmatrix}\sqrt{N}\left[\begin{pmatrix} \bar{y} \\ \overline{y^2} \end{pmatrix}-\begin{pmatrix} 0 \\ \mu_2 \end{pmatrix}\right] \\
&\overset{a}{\sim} \mathrm{N}\left(0,\mu_4-\mu_2^2\right)
\end{aligned}
$$

从而我们得到：

$$
\sqrt{N}\left(s^2-\mu_2\right)\overset{a}{\sim}\mathrm{N}\left(0,\mu_4-\mu_2^2\right)
$$

delta方法

- 值得注意的是，尽管在该方法中，通常泰勒展开第一阶就可以得到需要的渐近分布，但是也有很多时候展开第一阶是不够的
- 此时，我们可能需要进一步展开到第二阶甚至更高阶进而得到所需的渐近分布。

高阶展开

delta方法的高阶展开

如果$x_i \sim \mathrm{N}\left(0, \sigma^2\right)$ $i.i.d$，为了计算$\cos \bar{x}$的渐近分布，我们使用delta方法：

$$\sqrt{N}\left(\cos \bar{x} - \cos 0\right) = \sqrt{N}\left(\cos \bar{x} - 1\right) = \sqrt{N} \times 0 \times \left(\bar{x} - 0\right) + \sqrt{N} O\left(\bar{x}^2\right) = o_p\left(1\right)$$

因而泰勒展开第一项为0，因而乘以$\sqrt{N}$收敛速度还是太慢

高阶展开

delta方法的高阶展开

我们需要更进一步展开:

$$
\begin{aligned}
N\left(\cos\bar{x}-1\right) &= \frac{1}{2}\times(-1)\times N\left(\bar{x}-0\right)^2+N\cdot O\left(\left|\bar{x}-0\right|^3\right)\\
&= -\frac{1}{2}\times\sqrt{N}\left(\bar{x}-0\right)\times\sqrt{N}\left(\bar{x}-0\right)+o_p\left(1\right)\\
&\xrightarrow{p} -\frac{1}{2}\left[\sqrt{N}\left(\bar{x}-0\right)\right]^2
\end{aligned}
$$

由于$\sqrt{N}\left(\bar{x}-0\right)\overset{a}{\sim}\mathrm{N}\left(0,\sigma^2\right)\sim\sigma Z$，其中$Z\sim\mathrm{N}\left(0,1\right)$，从而:

$$
\frac{N\left(\cos\bar{x}-1\right)}{-\frac{1}{2}\sigma^2}=-\frac{2N\left(\cos\bar{x}-1\right)}{\sigma^2}\overset{a}{\sim}\chi^2\left(1\right)
$$

作业

- 7.4
- 7.5
- 7.6
- 7.7
- 7.8