高等统计学_第六周

3.5 Large Deviations

All the limit theorems and expansions derived so far deal with absolute errors. 绝对误差

Although they are useful for moderate values of x, they are less meaningful for large x.

> 适用于中等大小的 x ，但是对于比较大的 x 不适用

For instance, CLT states that

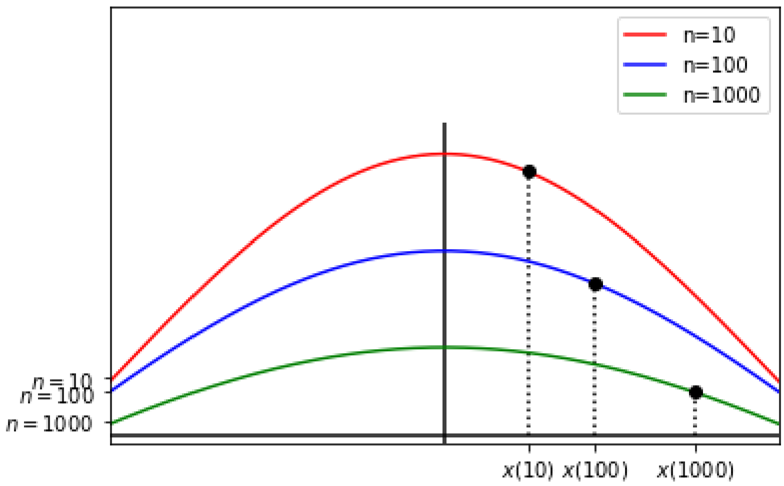
$$\sup_{x\in\mathbb{R}}|F_n(x)-\Phi(x)|\to 0$$

However, for large $|x|$, both $F_n(x)$ and $\Phi(x)$ are either close to 1 or 0, therefore, the statement of CLT becomes empty.

In this section, we will look at the tail probability $1-F_n(x)=P(\sqrt{n}\overline{X}>x)$ as $x=:x_n\to\infty$

$$P(\overline{x}>a)=p(\sqrt{n}\overline{x}>\sqrt{n}a)=1-F_n(\sqrt{n}a)$$

其中，$\sqrt{n}\overline{x}$ 近似标准正态，而 $\sqrt{n}a$ 跟 n 有关系，所以 $1-F_n(x)$ 会随着n 改变而改变（增大）



$$P(\overline{x}>a)=P(\sqrt{n}\overline{x}>\sqrt{n}a)=1-F_n(\sqrt{n}a)$$

n 增大，$\sqrt{n}a$ 增大，$1-f_n(\sqrt{n}a)$ 变小

For simplicity, we shall assume that Cramer condition holds.

3.5.1 Cramer condition

Let $X_1, X_2, \ldots, X_n$ be $i.i.d\ r.v.'s$

Let the following $Cramer's condition$ hold:

$$Ee^{tX} < \infty \qquad \text{in } |t| < H \quad \text{for some constant } H > 0$$

Cramer's condition simply means that the moment generating function exists near the origin, and implies that moments of all orders exist.

Several equivalent forms are given below.

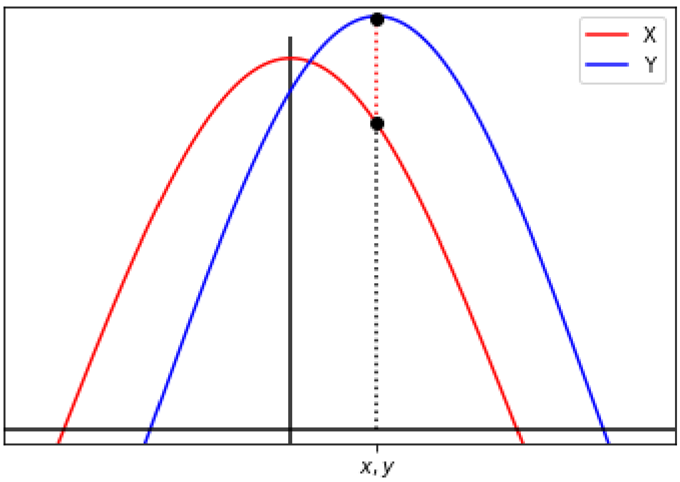
3.5.2 Conjugate (or associated) distributions

思想：移动期望

因为随着n 的增大，$1 - F_n(x)$ 会变小，所以在计算的时候只能取 x 为小范围，如果 x 的取值偏大，就会出现较大的偏差。

如果能将 X 的期望移动，令 Y 的期望刚好等于所求的那个值，则可以转化为求 Y 大于0，避免了 n 的增大而导致的误差

$$P(\overline{x} > a) =: P(\hat{y} > 0) = 1 - F(0)$$



此时求的概率即 Y 大于0的概率。最后可以通过 X 和 Y 的关系来反求 X 的概率

One of the key tools in large deviation theory is the so-called $conjugate$ (or associated) distribution.

共轭分布（随机过程破产概率）

Let $X_1, X_2, \ldots, X_n$ be $i.i.d\ r.v.'s$ with a common $d.f.\ F$. 独立同分布的随机变量，有共同的分布函数 F

Denote the **moment generating function** ($m.g.f.$) and **cumulant generating function** ($c.g.f.$) of X by

$$M_X(t) = Ee^{tX} = \int_{-\infty}^{\infty} e^{tX} dF(x) \qquad K_X(t) = \ln M_X(t)$$

- moment generating function ($m.g.f.$) 矩母函数
- cumulant generating function ($c.g.f.$) 累积生成函数

Definition 3.5.1

Given a $d.f.\ F$, define its conjugate (or associated) distribution by

$$G(y) = \frac{\int_{-\infty}^{y} e^{sx} dF(x)}{\int_{-\infty}^{\infty} e^{sx} dF(x)} = \frac{\int_{-\infty}^{y} e^{sx} dF(x)}{M_X(s)} = \int_{-\infty}^{y} e^{sy-K_X(s)} dF(x)$$

or equivalently

$$dG(x) = \frac{e^{sx} dF(x)}{\int_{-\infty}^{\infty} e^{sx} dF(x)} = \frac{e^{sx} dF(x)}{M_X(s)} = e^{sx-K_X(s)} dF(x)$$

We shall provide several useful lemmas(引理)

随机变量 X 与 Y 无关，但其分布有关系，满足上面的等式

Lemma 3.5.1

It is easy to see that $m.g.f.'s$ and $c.g.f.'s$ of X and Y are related by

$$M_Y(t)=\frac{M_X(t+s)}{M_X(s)} \qquad \rightarrow^{\log} \qquad K_Y(t)=K_X(t+s)-K_X(s)$$

```
Proof.
```

$$\begin{aligned} M_Y(t) &= Ee^{tY} \\ &= \int e^{ty}dG(y) \\ &= \int e^{ty}e^{sy-K_X(s)}dF(y) \\ &= e^{-K_X(s)}\int e^{(t+s)y}dF(y) \\ &= \frac{M_X(t+s)}{M_X(s)} \end{aligned}$$

In particular

$$EY=K'_Y(t)|_{t=0}=K'_X(s) \qquad Var(Y)=K''_Y(t)|_{t=0}=K''_X(s)$$

现在有了关于随机变量 X 与 Y 的期望的求法，则可以通过令 $EY=K'_X(s)=a$ 来反求 s

如果 s 求出来，则其共轭分布也就定下来了，即知道了 $dG(x)$

那么可以反求 $F(x)$ ，然后可以求出 $P(\overline{x}>a)$

$$\begin{array}{cc} EY=K'_X(s)=a & \\ \Downarrow & \\ s & \\ \Downarrow & \\ dG(x) & =e^{sx-K_X(s)}dF(x)?? \\ & \text{仅仅知道s就可以知道}dG(x)\text{吗？} \\ \Downarrow & \\ F(x) & \\ \Downarrow & \\ P(X>a) & \end{array}$$

Remark 3.5.1

After a change of measure from F to G, the mean of X is changed from $E_FX=0$ under F to $E_GX=K'_X(s)$ under G

By choosing different s, we can freely change the mean to any (allowable) location.

In particular, in estimating $P(\overline{X}>x)$ for some large x, we can choose appropriate s so that the point x becomes the center of $d.f.$ rather than the tail. 选择合适的 s 使得点 x 变成分布函数的中心，而不是尾部的概率

The reason for doing this is that one can apply many nice results such as CLT, Edgewoth expansions, etc, which behave very nicely near the center of the $d.f.$

知道了分布之间的关系，再来求统计量之间的关系

Lemma 3.5.2

Let $Y, Y_1, \ldots, Y_n$ be $i.i.d.\ r.v.'s$ with a common $d.f.\ G$. Define

$$F_n(x) = P\left(\sum_{i=1}^n X_i \le x\right) \qquad G_n(y) = P\left(\sum_{i=1}^n Y_i \le y\right)$$

Then

$$G_n(y) = \int_{-\infty}^{y} e^{sx-nK_X(s)} dF_n(x) = \frac{\int_{-\infty}^{y} e^{sx} dF_n(x)}{(e^{K_X(s)})^n} = \frac{\int_{-\infty}^{y} e^{sx} dF_n(x)}{M_X^n(s)} = \frac{\int_{-\infty}^{y} e^{sx} dF_n(x)}{\int_{-\infty}^{\infty} e^{sx} dF_n(x)}$$

or equivalently

$$dG_n(x) = e^{sx-nK_X(s)} dF_n(x) = \frac{e^{sx} dF_n(x)}{(e^{K_X(s)})^n} = \frac{e^{sx} dF_n(x)}{M_X^n(s)} = \frac{e^{sx} dF_n(x)}{\int_{-\infty}^{\infty} e^{sx} dF_n(x)}$$

证明思路：证明其特征函数或矩母函数一样

Proof.

The LHS has $m.g.f.$

$$M_{\sum_{i=1}^n Y_i}(t) = Ee^{t\sum_{i=1}^n Y_i} = M_Y^n(t) = \left(\frac{M_X(t+s)}{M_X(s)}\right)^n = \frac{M_{\sum_{i=1}^n X_i}(t+s)}{M_{\sum_{i=1}^n X_i}(s)} \qquad i.i.d$$

The RHS has $m.g.f.$

$$\int_{-\infty}^{\infty} e^{tx} \frac{e^{sx} dF_n(x)}{\int_{-\infty}^{\infty} e^{sx} dF_n(x)} = \frac{\int_{-\infty}^{\infty} e^{(t+s)x} dF_n(x)}{\int_{-\infty}^{\infty} e^{sx} dF_n(x)} = \frac{M_{\sum_{i=1}^n X_i}(t+s)}{M_{\sum_{i=1}^n X_i}(s)} \qquad \text{定义}$$

By the uniqueness theorem, we hence proved the lemma.

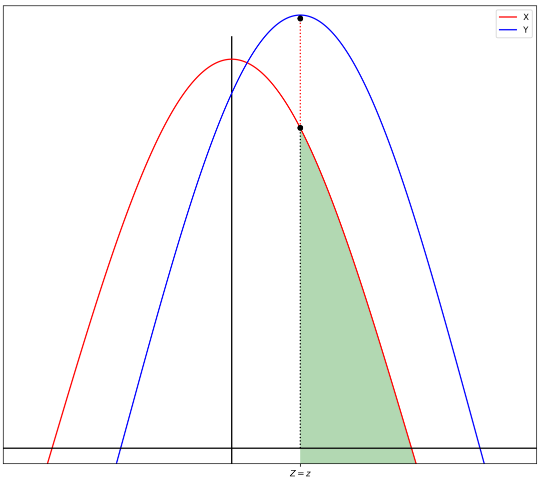
Lemma 3.5.3

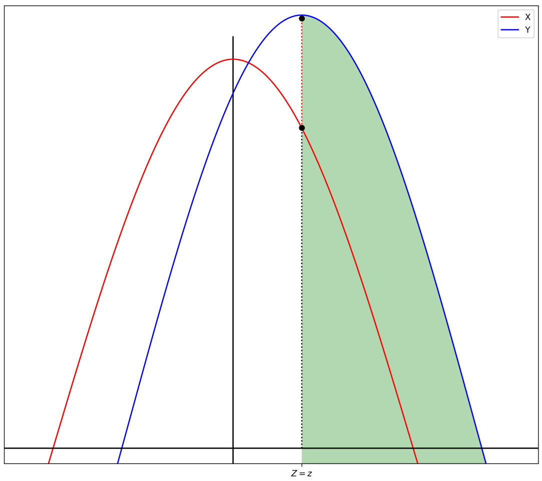
If we choose τ such that $K'_X(\tau) = z$, then we have

$$
\begin{aligned}
P(\overline{X} > z) &= e^{n[K_X(\tau)-\tau K'_X(\tau)]} \int_0^\infty e^{-\tau y\sqrt{nK''_x(\tau)}} dP(T_n \le Y) \\
&= e^{-n[z\tau - K_x(\tau)]} \int_0^\infty e^{-\tau y\sqrt{nK''_x(\tau)}} dP(T_n \le y) \\
&= e^{-n \sup_{t>0}[zt - K_X(t)]} \int_0^\infty e^{-\tau y\sqrt{nK''_x(\tau)}} dP(T_n \le y)
\end{aligned}
$$

where

$$
T_n = \frac{\sum_{i=1}^n Y_i - nEY}{\sqrt{nVar(Y)}} = \frac{\sum_{i=1}^n Y_i - nEY}{\sqrt{nK''_X(s)}}
$$





Proof.

$$
\begin{aligned}
P(\overline{X}>z) &= P\left(\sum_{i=1}^{n} X_i > nz\right) \\
&= \int_{nz}^{\infty} dP\left(\sum_{i=1}^{n} X_i \le x\right) \\
&= \int_{nz}^{\infty} e^{-sx+nK_X(s)} dP\left(\sum_{i=1}^{n} Y_i \le x\right) \\
&= e^{nK_X(s)} \int_{nz}^{\infty} e^{-sx} dP\left(\frac{\sum_{i=1}^{n} Y_i - nEY}{\sqrt{nVar(Y)}} \le \frac{x-nK_x'(\tau)}{\sqrt{nK_X''(\tau)}}\right) \\
&= e^{nK_X(s)} \int_{\frac{\sqrt{n}[z-K_X'(s)]}{\sqrt{K_x''(s)}}}^{\infty} e^{-s\left(nK_X'(s)+y\sqrt{nK_X''(s)}\right)} dP(T_n \le y) \\
&= e^{n[K_X(s)-sK_X'(s)]} \int_{\frac{\sqrt{n}[z-K_X'(s)]}{\sqrt{K_x''(s)}}}^{\infty} e^{sy\sqrt{nK_X''(s)}} dP(T_n \le y)
\end{aligned}
$$

where

$$
T_n = \frac{\sum_{i=1}^{n} Y_i - nEY}{\sqrt{nVar(Y)}} \qquad y = \frac{x-nK_X'(s)}{\sqrt{nK_X''(s)}}
$$

$\int_{\frac{\sqrt{n}[z-K_X'(s)]}{\sqrt{K_x''(s)}}}^{\infty}$ 是一个很小的数，而 $P(T_n \le y)$ 可以看作是一个标准正态分布，从0到∞积分，是 $\sum_{i=1}^{n} Y_i$ 的统计量标准化之后大于0的概率

If we choose τ such that $K_X'(\tau)=z$, then we have

$$
P(\overline{X}>z) = e^{n[L_X(\tau)-sK_X'(\tau)]} \int_0^{\infty} e^{-\tau y\sqrt{nK_X''(\tau)}} dP(T_n \le y)
$$

3.5.3 Large Deviations

From lemma ??, under Crammer's condition, $P(S_n > na)$ converges to 0 exponentially fast.

More careful analysis can give a precise convergence rate.

Theorem 3.5.1

Let

- $X_1, X_2, \cdots$ be $i.i.d.$ with mean 0
- $m.g.f.\ M(t) = Ee^{tx} < \infty$ for $|t| < \delta$
- $K(t) = \log M(t)$

Then

$$\lim_n P(S_n > na)^{\frac{1}{n}} = lim_n P(\overline{X} > a)^{\frac{1}{n}} = e^{-\sup_{t>0}[at-K(t)]} =: e^{-\eta(a)}$$

or equivalently

$$-lim_n \frac{1}{n} \ln P(S_n > na) = \sup_{t>0}[at - K(t)] =: \eta(a)$$

where $\eta(a) = \sup_{t>0}[at - K(t)]$

> $\sup[at-K(t)]$ 可以通过求导求得最大值
>
> $\eta(a)$ 是一个常数

Remark 3.5.2

From the theorem, the bound is interesting if $\eta(a) > 0$.

It suffices to show that $at - K(t) > 0$ for some $t > 0$.

To prove this, we note

$$\begin{aligned} at - K(t) &= at - [K(0) + K'(0)t + \frac{1}{2}K''(0)t^2 + \cdots] \\ &= at - \frac{1}{2}\sigma^2 t^2 + o(t^2) \\ &= at\left(1 - \frac{1}{2}\sigma^2 t + o(t)\right) \\ &> 0 \end{aligned}$$

when $t > 0$ is chosen to be small enough.

Proof. (对于凹函数，有 $Ef(x) \geq f(E(x))$)

We first give an upper bound.

For any $t > 0$, by Chebyshev inequality, we have

$$P(S_n > na)^{\frac{1}{n}} = P(e^{tS_n} > e^{nat})^{\frac{1}{n}} \leq \left(\frac{Ee^{tS_n}}{e^{nat}}\right)^{\frac{1}{n}} = \left(\frac{(Ee^{tX_1})^n}{e^{nat}}\right)^{\frac{1}{n}} = \frac{M(t)}{e^{at}} = e^{K(t)-at}$$

Take $\inf_{t>0}$ on both sides, we get

$$P(S_n > na)^{\frac{1}{n}} \leq \inf_{t>0} e^{K(t)-at} = e^{\inf_{t>0}[K(t)-at]} = e^{-\sup_{t>0}[at-K(t)]}$$

Next we will give a lower bound.

From the last lemma, we have

$$\begin{aligned} P\left(\overline{X} > a\right)^{\frac{1}{n}} &= e^{[K_X(\tau)-sK'_X(\tau)]}\left(\int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}} dP(T_n \leq y)\right)^{\frac{1}{n}} \\ &= -e^{\eta(a)}\left(\int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}} dP(T_n \leq y)\right)^{\frac{1}{n}} \\ &=: e^{-\eta(a)} A_n^{1/n} \end{aligned}$$

It suffices to show that $A_n^{1/n} \geq 1$ as $n \to \infty$.

To do this, note that

$$\begin{aligned} &=: \int_0^\infty e^{-\sqrt{n}\tau_0 y} dP(T_n \leq y) \qquad \text{where } \tau_0 = \tau\sqrt{K''_X(\tau)} \\ &=: \frac{\int_0^\infty e^{-\sqrt{n}\tau_0 y} dP(T_n \leq y)}{\int_0^\infty dP(T_n \leq y)} \int_0^\infty dP(T_n \leq y) \\ &= E\left(E^{-\sqrt{n}\tau_0 T_n} | T_n > 0\right) P(T_n > 0) \\ &\geq \exp(\{-\sqrt{n} tao_0 E[T_n | T_n > 0]\}) P(T_n > 0) \qquad \text{by Jensen's inequality} \end{aligned}$$

Now

$$E[T_n|T_n>0]=\frac{E[T_nI(T_n>0)]}{P(T_n>0)}\le\frac{E|T_n|}{P(T_n>0)}\le\frac{(ET_n^2)1/2}{P(T_n>0)}\le\frac{1}{P(T_n>0)}\le\frac{1}{0.5-\epsilon}$$

and therefore

$$A_n^{1/n}\ge\exp\left(\frac{-\sqrt{n}\tau_0}{nP(T_n>0)}\right)P^{1/n}(T_n>0)\to 1$$

Since $P(T_n>0)\to 1/2$ and $P^{1/n}(T_n>0)\to 1$

若不用大偏差，强行用 CLT 来算（假设 $EX=0, VarX=1$ ）

$$\begin{aligned}P(\overline{X}>a)&=P(\sqrt{n}\overline{x}>\sqrt{n}a)\\&=P(\delta>\sqrt{n}a)\\&\approx 1-\Phi(\sqrt{n}a)\\&\approx\frac{1}{\sqrt{n}a}\frac{1}{\sqrt{2\pi}}e^{-\frac{na^2}{2}}\qquad 1-\Phi(x)\text{等价于}\frac{1}{x}\phi(x)\quad(x\to\infty)\\&=c\times\frac{1}{\sqrt{n}}e^{-\frac{a^2}{2}n}\\-\frac{1}{n}\ln P(\overline{x}>a)&=c_1\times\frac{a^2}{2}\end{aligned}$$

用大偏差 $-lim_n\frac{1}{n}\ln P(S_n>na)=\sup_{t>0}[at-K(t)]=:\eta(a)$

则 $\eta(a)$ 不一定等于 $c_1\times\frac{a^2}{2}$ ，但比较靠近

3.5.4 Cramer-type large deviations

中偏差理论

下面这个表格不大确定，没听清楚

$(0, c)$	$(c, n^{1/4})$	$(n^{1/2}, \infty)$
小偏差	中偏差	大偏差

We indicated at the beginning of this section that normal approximation is of limited use when we look into the far tail of the $d.f.$ of standardized sums.

The CLT only works well not too far away from the center of the distribution.

The question is then how far can we actually safely use the CLT as we move away from the center of the distribution.

Since $1-F_n(x)$ and $1-\Phi(x)$ are both close to 0 as $x \to \infty$, using $|F_n(x)-\Phi(x)$ as a measure of closeness of the two $d.f.s$ may not be very helpful to us.

A more useful measure in this case is to estimate the relative error in approximating $1-F_n(x)$ by $1-\Phi(x)$ when $x \to \infty$, or approximating $F_n(x)$ by $\Phi(x)$ when $x \to \infty$.

In other words, we should like to have

$$\frac{1-F_n(x)}{1-\Phi(x)} \to 1 \qquad \frac{F_n(-x)}{\Phi(-x)} \to 1$$

when both x and n tend to ∞

However, the statement can not be true in general.

For instance, if $X_1, X_2, \ldots, X_n$ are $i.i.d$ with $P(X_1=0)=PX_2=1)=1/2$, then $P(S_n>n)=1-F_n(\sqrt{n})=0$.

Thus, for $|x|>\sqrt{n}$, the ration are 0.

But as we shall see next, the statement is true if x varies with x such that $x=x_n \to \infty$ at a certain rate.

Theorem 3.5.2

$$\frac{1-F_n(x)}{1-\Phi(x)}=1+O\left(\frac{x^3}{\sqrt{n}}\right)$$

> 若 $x<n^{1/6}$ ，则 $x^3<n^{1/2}$ ，则 $\frac{1-F_n(x)}{1-\Phi(x)}=1+O(1)=1+o(1)$
>
> 算尾部概率 P ，用 $\Phi(x)$ 近似，看 $\frac{1-F_n(x)}{1-\Phi(x)}-1$ 是不是很小，若 $x \in (0, o(n^{1/6}))$ 之间，则相对误差较小

CLT和大偏差是$|F_n(x)-\Phi(x)|$，是绝对误差

Proof. 看不懂，不写了-----------