数理统计
第二章
抽样分布及若干预备知识

2026 年 3 月 25 日

目录

2.5.1 定义

在许多情形下，统计量的精确分布很难求出，因此要研究统计量的极限分布。首先给出下列定义。

在许多情形下，统计量的精确分布很难求出，因此要研究统计量的极限分布。首先给出下列定义。

定义 (2.5.1（大样本分布）)

当样本容量$n \to \infty$时，统计量的分布趋于一确定分布，则后者的分布称为统计量的极限分布，也常称为大样本分布。

在许多情形下，统计量的精确分布很难求出，因此要研究统计量的极限分布。首先给出下列定义。

定义 (2.5.1（大样本分布）)

当样本容量$n \to \infty$时，统计量的分布趋于一确定分布，则后者的分布称为统计量的极限分布，也常称为大样本分布。

当样本容量n充分大时，极限分布可作为统计量的近似分布。

2.5.1 定义

研究统计量的极限分布有下列意义：

2.5.1 定义

研究统计量的极限分布有下列意义：

（1）为了获得统计推断方法的优良性，常常要知道统计量的分布。但统计量的精确分布一般很难求得，建立统计量的极限分布，提供了一种近似方法，总比什么方法都没有好。

2.5.1 定义

（2）有时统计量的精确分布虽可求出，但表达式过于复杂，使用不方便。若极限分布较简单，宁可使用极限分布。

2.5.1 定义

（2）有时统计量的精确分布虽可求出，但表达式过于复杂，使用不方便。若极限分布较简单，宁可使用极限分布。

（3）有些统计推断方法的优良性本身就是研究其极限性质，如相合性、渐近正态性等。

定义 (2.5.2（大样本和小样本性质）)

当样本容量$n \to \infty$时，一个统计量或统计推断方法的性质称为**大样本性质**（large sample properties）。

2.5.1 定义

定义 (2.5.2（大样本和小样本性质）)

当样本容量$n \to \infty$时，一个统计量或统计推断方法的性质称为**大样本性质**（large sample properties）。

当样本大小固定时，统计量或统计推断方法的性质称为**小样本性质**。

定义 (2.5.2（大样本和小样本性质）)

当样本容量$n \to \infty$时，一个统计量或统计推断方法的性质称为大样本性质（large sample properties）。

当样本大小固定时，统计量或统计推断方法的性质称为小样本性质。

大样本性质和小样本性质的差别不在于样本容量的多少，而是在于所讨论的问题是在样本容量$n \to \infty$时去考虑，还是在样本容量n固定时去研究。

2.5.2 几个例子

引理 (强大数定律 A.11（回顾）)

设$X_1, \ldots, X_n$为独立同分布的随机变量序列，若$\mu = E|X_1| < \infty$，则

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu.$$

引理 (强大数定律 A.11（回顾）)

设$X_1, \ldots, X_n$为独立同分布的随机变量序列，若$\mu = E|X_1| < \infty$，则

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu.$$

引理 (中心极限定理 A.12（回顾）)

设$\{X_i\}$为一列独立同分布随机变量序列，有共同的均值μ和方差$0 < \sigma^2 < \infty$。则有$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{\mathcal{L}} N(0,1)$，其中$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

2.5.2 几个例子

例 (2.5.1)

设$X_1, X_2, \ldots, X_n \ i.i.d. \sim F$，其中总体$F$有均值$a_F$和方差$\sigma_F^2$。

设$0 < \sigma_F^2 < \infty$，$\bar{X}_n = \sum_{i=1}^{n} \frac{X_i}{n}$为样本均值，试讨论$\bar{X}_n$的大样本性质和小样本性质。

解（1）由柯尔莫哥洛夫（Kolmogorov）强大数律有

解（1）由柯尔莫哥洛夫（Kolmogorov）强大数律有

$$P(\lim_{n\to\infty}\bar{X}_n = a_F) = 1 \Longleftrightarrow \bar{X}_n \xrightarrow{a.s.} a_F,\ \text{当} n \to \infty \text{时}.$$

解（1）由柯尔莫哥洛夫（Kolmogorov）强大数律有

$$P(\lim_{n\to\infty}\bar{X}_n = a_F) = 1 \Longleftrightarrow \bar{X}_n \xrightarrow{a.s.} a_F,\ \text{当}n\to\infty\text{时}.$$

上式表示当样本容量$n\to\infty$时，估计量$\bar{X}_n$以概率1任意接近估计值a_F（以概率1收敛于a_F，又称几乎处处收敛于a_F）。

解（1）由柯尔莫哥洛夫（Kolmogorov）强大数律有

$$P(\lim_{n\to\infty}\bar{X}_n = a_F) = 1 \Longleftrightarrow \bar{X}_n \xrightarrow{a.s.} a_F，\text{当}n \to \infty\text{时}.$$

上式表示当样本容量$n \to \infty$时，估计量$\bar{X}_n$以概率1任意接近估计值a_F（以概率1收敛于a_F，又称几乎处处收敛于a_F）。

这个性质称为$\bar{X}$的强相合性，是一个大样本性质。因为只有在$n \to \infty$时这个性质才有意义。

（2）按Lindeberg中心极限定理，当$n \to \infty$时有

$$\sqrt{n}(\bar{X}_n - a_F)/\sigma_F \xrightarrow{\mathcal{L}} N(0,1), \tag{1}$$

（2）按Lindeberg中心极限定理，当$n \to \infty$时有

$$\sqrt{n}(\bar{X}_n - a_F)/\sigma_F \xrightarrow{\mathcal{L}} N(0,1), \tag{1}$$

其中$\xrightarrow{\mathcal{L}}$表示依分布收敛。式子（1）刻画了$\bar{X}_n$的另一个大样本性质——渐近正态性。

（2）按Lindeberg中心极限定理，当$n \to \infty$时有

$$\sqrt{n}(\bar{X}_n - a_F)/\sigma_F \xrightarrow{\mathcal{L}} N(0,1), \tag{1}$$

其中$\xrightarrow{\mathcal{L}}$表示依分布收敛。式子（1）刻画了$\bar{X}_n$的另一个大样本性质——渐近正态性。

$\bar{X}_n$作为总体均值的估计与a_F的偏差超过c的概率$P(|\bar{X}_n - a_F| > c)$，可以作为衡量这一估计量优良性的一项指标。

若F为正态分布$N(a_F,\sigma_F^2)$，则这一概率可以精确地算出（当σ_F^2已知时，可通过查附表1得出），若F的分布类型根本不知道，这一概率无法计算。

若F为正态分布$N(a_F,\sigma_F^2)$，则这一概率可以精确地算出（当σ_F^2已知时，可通过查附表1得出），若F的分布类型根本不知道，这一概率无法计算。

但有了式（1）后，至少在样本容量n较大时，可用正态分布求得这一概率的近似值。

（3）关于$\bar{X}$的小样本性质有$E(\bar{X}_n) = a_F$，即估计量$\bar{X}_n$的期望值等于被估计的未知参数a_F。

（3）关于$\bar{X}$的小样本性质有$E(\bar{X}_n) = a_F$，即估计量$\bar{X}_n$的期望值等于被估计的未知参数a_F。

这个性质称为$\bar{X}_n$的无偏性（即$\bar{X}_n$为a_F的无偏估计）。这是一个小样本性质。因为这个性质的意义是在样本大小n固定时去理解的。

引理 ((Slutsky 引理)(A.9)（回顾）)

令$\{X_n\}$和$\{Y_n\}$是两个随机变量的序列，满足$n \to \infty$时$X_n \xrightarrow{\mathcal{L}} X$，$Y_n \xrightarrow{p} c$，$-\infty < c < \infty$为常数，则有

引理 ((Slutsky 引理)(A.9)（回顾）)

令$\{X_n\}$和$\{Y_n\}$是两个随机变量的序列，满足$n \to \infty$时$X_n \xrightarrow{\mathcal{L}} X$，$Y_n \xrightarrow{p} c$，$-\infty < c < \infty$为常数，则有

(1) $X_n \pm Y_n \xrightarrow{\mathcal{L}} X \pm c$，

引理 ((Slutsky 引理)(A.9)（回顾）)

令$\{X_n\}$和$\{Y_n\}$是两个随机变量的序列，满足$n \to \infty$时$X_n \xrightarrow{\mathcal{L}} X$，$Y_n \xrightarrow{p} c$，$-\infty < c < \infty$为常数，则有

(1) $X_n \pm Y_n \xrightarrow{\mathcal{L}} X \pm c$，

(2) $X_n Y_n \xrightarrow{\mathcal{L}} cX$，

引理 ((Slutsky 引理)(A.9)（回顾）)

令$\{X_n\}$和$\{Y_n\}$是两个随机变量的序列，满足$n \to \infty$时$X_n \xrightarrow{\mathcal{L}} X$，$Y_n \xrightarrow{p} c$，$-\infty < c < \infty$为常数，则有

(1) $X_n \pm Y_n \xrightarrow{\mathcal{L}} X \pm c$，

(2) $X_n Y_n \xrightarrow{\mathcal{L}} cX$，

(3) $X_n / Y_n \xrightarrow{\mathcal{L}} X/c \ (c \neq 0)$。

例 (2.5.2)

设$X \sim B(1,p)$，令$X_1,\cdots,X_n$为自总体X中抽取的简单样本。试证$\sqrt{n}(\bar{X}-p)/\sqrt{\bar{X}(1-\bar{X})}$的极限分布为$N(0,1)$。

例 (2.5.2)

设$X \sim B(1,p)$，令$X_1,\cdots,X_n$为自总体X中抽取的简单样本。试证$\sqrt{n}(\bar{X}-p)/\sqrt{\bar{X}(1-\bar{X})}$的极限分布为$N(0,1)$。

解 令$\hat{p}=\bar{X}=\sum_{i=1}^{n} X_i/n$，由中心极限定理和大数定率可知，当$n\to\infty$时有

例 (2.5.2)

设$X \sim B(1,p)$，令$X_1, \cdots, X_n$为自总体X中抽取的简单样本。试证$\sqrt{n}(\bar{X}-p)/\sqrt{\bar{X}(1-\bar{X})}$的极限分布为$N(0,1)$。

解 令$\hat{p}=\bar{X}=\sum_{i=1}^{n} X_i/n$，由中心极限定理和大数定率可知，当$n \to \infty$时有

$$\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \xrightarrow{\mathcal{L}} N(0,1), \quad \frac{p(1-p)}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{p} 1.$$

故由 Slutsky 引理可知，当$n \to \infty$时有

故由 Slutsky 引理可知，当$n \to \infty$时有

$$\frac{\sqrt{n}(\bar{X}-p)}{\sqrt{\bar{X}(1-\bar{X})}} = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \cdot \frac{\sqrt{p(1-p)}}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{\mathcal{L}} N(0,1).$$

$\square$

例 (2.5.3*)

设总体X有密度f，$X_1,\ldots,X_n$，$i.i.d\sim X$，令总体X的中位数为$\xi_{1/2}$，即$\xi_{1/2}$满足$\int_{-\infty}^{\xi_{1/2}}f(x)dx=1/2$，且假定$f$在$\xi_{1/2}$点连续非0（由此可知$\xi_{1/2}$为$f$唯一的中位数）。设$m_{1/2}$为样本中位数，则当$n\longrightarrow\infty$时有

$$2\sqrt{n}f(\xi_{1/2})(m_{1/2}-\xi_{1/2})\xrightarrow{\mathcal{L}}N(0,1),$$

例 (2.5.3*)

设总体X有密度f，$X_1,\ldots,X_n$, $i.i.d\sim X$，令总体X的中位数为$\xi_{1/2}$，即$\xi_{1/2}$满足$\int_{-\infty}^{\xi_{1/2}} f(x)dx=1/2$，且假定$f$在$\xi_{1/2}$点连续非0（由此可知$\xi_{1/2}$为$f$唯一的中位数）。设$m_{1/2}$为样本中位数，则当$n\longrightarrow\infty$时有

$$2\sqrt{n}f(\xi_{1/2})(m_{1/2}-\xi_{1/2})\xrightarrow{\mathcal{L}} N(0,1),$$

此处样本中位数$m_{1/2}$的定义为

$$m_{\frac{1}{2}}=\begin{cases} X_{((n+1)/2)}, & n\text{为奇数}, \\ \frac{1}{2}[X_{(n/2)}+X_{(n/2+1)}], & n\text{为偶数}. \end{cases}$$

例2.5.3* 是下面定理2.5.1的特例。

定理 (2.5.1)

设$0 < p < 1$，$\lim_{n\to\infty}(m-np)/\sqrt{n}=0$，$\xi_p$满足$F(\xi_p)=\int_{-\infty}^{\xi_p}f(x)dx=p$，且假定密度$f$在点$\xi_p$处连续非0（由此知$\xi_p$为$f$唯一的$p$分位数）。则当$n\to\infty$时有

$$\sqrt{n}f(\xi_p)(X_{(m)}-\xi_p)/\sqrt{p(1-p)} \xrightarrow{\mathcal{L}} N(0,1),$$

其中$X_{(m)}$是$X_1,\ldots,X_n$的一个次序统计量。

分位数-分位数图

分位数-分位数图

分位数-分位数图（quantile-quantile plot, Q-Q plot），也称为概率图，是一种检查观测数据是否来自指定的理论分布的图形化工具。

分位数-分位数图

分位数-分位数图（quantile-quantile plot, Q-Q plot），也称为概率图，是一种检查观测数据是否来自指定的理论分布的图形化工具。

与分位数-分位数图类似的另外一种图形工具是概率-概率图（probability-probability plot，P-P plot），其根据变量的经验累积概率对应于所指定的理论分布累积概率而绘制的散点图。

设$X_1,\dots,X_n$ $i.i.d. \sim F$，相应的次序统计量为$(X_{(1)},X_{(2)},\dots,X_{(n)})$。

如果样本确实来自分布F，则根据定理2.5.1 我们可以期待点

设$X_1,\ldots,X_n$ $i.i.d.\sim F$，相应的次序统计量为$(X_{(1)},X_{(2)},\ldots,X_{(n)})$。如果样本确实来自分布$F$，则根据定理2.5.1 我们可以期待点

$$\left(x_{(i)},F^{-1}\left(\frac{i}{n+1}\right)\right),\ i=1,2,\ldots,n$$

设$X_1,\dots,X_n$ $i.i.d.\sim F$，相应的次序统计量为$(X_{(1)},X_{(2)},\dots,X_{(n)})$。如果样本确实来自分布$F$，则根据定理2.5.1 我们可以期待点

$$\left(x_{(i)},F^{-1}\left(\frac{i}{n+1}\right)\right),\ i=1,2,\dots,n$$

接近对角线$y=x$；相反地，偏离对角线越远就表明样本不是来自该分布的证据越强烈。

特别当F为正态分布$N(\mu, \sigma^2)$时则称为正态分位数-分位数图。此时成立

2.5.2 几个例子

特别当F为正态分布$N(\mu,\sigma^2)$时则称为正态分位数-分位数图。此时成立

$$F(X_{(i)}) = \Phi\left(\frac{X_{(i)}-\mu}{\sigma}\right) = \frac{i}{n+1}$$
$$\Longrightarrow x_{(i)} \approx \mu + \sigma\Phi^{-1}\left(\frac{i}{n+1}\right).$$