绪 论

苏勤亮

中山大学计算机学院

超算中心503M

suqliang@mail.sysu.edu.cn

教材与参考书

- 教材
  - 茆诗松、吕晓玲，《数理统计学》（第2版），北京：中国人民大学出版社，2016，ISBN：978-7-300-22410-7
  - 韦来生，《数理统计》，北京：科学出版社，2015，ISBN：978-7-03-046573-3
- 参考书
  - 陈希孺、倪国熙，《数理统计学教材》，合肥：中国科学技术大学出版社，2009，ISBN：987-7-312-02282-1

课程支持

- 课程群

提纲

- 什么是数理统计
- 数理统计的若干概念
- 统计量

数理统计的定义

- 自然界的现象大致可以分为两大类, 一类称为确定性现象, 另一类称为非确定性现象，亦称为随机现象

- 数理统计是研究随机现象的统计规律性的数学学科
  - 统计学是研究怎样有效地收集和分析带有随机性影响的数据，从而对所考虑的问题作出一定结论的方法和理论

- 定义：数理统计是数学的一个分支，它是研究如何有效地收集和分析带有随机性数据的一门学科

有效的收集数据——抽样调查

- 收集数据的方法有:
  1) 全面观察(或普查)
  2) 抽样调查
  3) 试验设计

例：人口普查和抽样调查. 我国每十年进行一次人口普查. 如果普查的数据准确无误, 无随机性, 则不需用数理统计方法

- 人口普查虽是全面调查, 但数据并不很可靠
- 国家统计局还需派出专业人员对全国人口进行抽样调查, 根据抽样调查结果, 对人口普查的数字进行适当的修正

- 如何安排抽样调查, 是有效收集数据的一个重要问题, 这构成数理统计学的一个重要分支——《抽样调查方法》

例：考察某地区10000户农户的经济状况, 从中挑选100户做抽样调查. 若该地区分成平原和山区两部分, 平原较富, 占该地区农户的70%, 而占30%的山区农户较穷

- 我们的抽样方案规定在抽取的100户中, 从平原地区抽70户, 山区抽30户, 在各自范围内用随机化方法抽取
- 本例是通过合理地设计抽样方案来实现数据有效收集的

有效的收集数据——试验设计

例：某化工产品的得率与温度、压力和原料配方有关. 为提高得率，通过试验寻找最佳生产条件

实验因素和水平如下：

因素 \ 水平	1	2	3	4
温度	800℃	1000℃	1200℃	1400℃
压力	10	20	30	40
配方	A	B	C	D

- 3个因素, 每个因素4个水平，共要做$4^3 = 64$次试验. 做这么多试验, 人力、物力、财力都不可能

- 如何通过尽可能少的试验获得尽可能多的信息?
- 比如，可以采用正交实验设计的方法从全面实验中挑选出部分有代表性的点进行实验

- 如何安排试验方案和分析试验结果，这构成数理统计的另一分支——《试验的设计和分析》

有效的使用数据—建立统计模型

- 获取数据后，需要用有效的方法提取数据中的有关信息，以对所研究问题作出一定的结论，这在统计上称为“推断”

- 为了有效地使用数据进行统计推断，需要对数据建立一个统计模型（概率分布）

有效的使用数据——比较不同的统计方法

- 依据不同准则，评判不同统计推断方法的优劣

  例如，为估计一个物体的重量a，把它在天平上称 5 次获得数据 $x_1, x_2, \cdots, x_5$, 估计a的大小采用下列三种不同方法：

  1）用5个数的算术平均值$\bar{x}=\frac{1}{5}(x_1+\cdots+x_5)$去估计$a$；

  2）将$x_1, x_2, \cdots, x_5$按大小排列为 $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(5)}$，取中间一个值$x_{(3)}$去评估$a$

  3）用$W=\frac{1}{2}(x_{(1)}+x_{(5)})$去估计 a

- $\bar{x}$优于$x_{(3)}$, 而$x_{(3)}$优于W，这样的判断对不对？为什么？在什么条件下才对？这些正是数理统计学所研究的问题.

例：某地有100户农户，调查此村农户是否脱贫. 脱贫的标准是每户年均收入超过1万元. 经调查此村90户农户年收入5000元, 10户农户年收入10万元，问此村农户是否脱贫?

1） 用算术平均值: $x = (90 \times 0.5 + 10 \times 10)/100 = 1.45$（万）.

得出结论：该村农户已脱贫. 事实上多数农户并未脱贫.

2） 用样本中位数：设 $x_{(1)} \le x_{(2)} \le \cdots \le x_{(100)}$，则

$$样本中位数 = (x_{(50)} + x_{(51)})/2 = 0.5（万）$$

得出结论：该村农户尚未脱贫，这与实际情况相符.

- 采用合适的统计方法也是有效使用数据的一个重要方面，如在本例中使用样本中位数更能反映真实情况

数理统计学与其它学科间的关系

- 数理统计方法所处理的只是在各种专门学科中带普遍性(共性)且受随机性影响的数据收集、整理和推断问题，而不上涉及各种专门学科中的具体问题

- 这种带共性的问题既然从专门领域中提炼出来，就可以用数学的方法去研究，研究成果又可以用于指导专门学科中的应用

- 统计方法的重要特点：统计方法只是从事物的外在数量上的表现上推断该事物可能的规律．统计方法本身不能说明何以会有这个规律性, 这是各个专门学科的任务

数理统计学的归纳性质

- 数理统计与其它数学学科的推理方法不一样
  - 统计方法的推理本质是归纳式的
  - 其它数学学科的推理则是演绎式的

例如，统计学家通过大量的观察资料发现吸烟与某种呼吸系统的疾病有关，但并不能说明解释为什么会有关系

例如，统计学 vs 几何学

- 在几何学中要证明“等腰三角形两底角相等”，可以运用几何公理，一步步地推出结论，属于演绎推理
- 一个习惯于统计方法的人，就可能作出很多大小形状不一的等腰三角形，实际测量它的底角，看看可否作出底角相等的结论，这一方法属于归纳推理.

- 归纳推理所依据的数据具有随机性，归纳推理做出的推断不是100%可靠，但它的可靠程度是可以通过概率来度量

统计学的应用

- 国家行政机关和职能机构，如国家统计局，经常需要收集有关的数据和资料，以了解情况并做出相应的决策
- 在工农业生产中我们常常要利用试验设计和方差分析的方法寻找最佳生产条件
- 产品质量控制、抽样调查和工业产品寿命的可靠性问题——统计质量管理
- 在经济和金融领域也有广泛的应用，在经济学中定量分析的应用比其他社会科学部门更早更深入
- 在生物、医学和遗传学中有广泛的应用

- 在气象预报、水文、地震、地质等领域有广泛应用

- 数理统计方法在科学研究中也具有重要作用

- **大数据、人工智能**

提纲

- 什么是数理统计
- 数理统计的若干概念
- 统计量

总体

例：假定一批产品有10000 件, 其中有正品也有废品. 为估计废品率, 我们往往从中抽取一部分，如100件进行检查. 此时这批10000件产品称为总体，其中的每件产品称为个体，而从中抽取的100件产品称为样本. 样本中个体的数目称为样本大小, 也称为样本容量. 而抽取样本的行为称为抽样

- 总体是由与我们所研究的问题有关的所有个体组成的集合, 而样本是从总体中抽取的一部分个体

- 总体可以看成由所有个体上的某种数量指标构成的集合, 因此, 它是数的集合.
- 总体可以用一个随机变量及其分布来描述

定义：一个统计问题所研究的对象的全体称为总体. 在数理统计学中，总体可以用一个随机变量及其概率分布来描述

总体的表示

- 总体可用*r.v.* X来表示, 或用其分布函数$F(x)$来表示. 若$F(x)$有密度, 记为$f(x)$, 则此总体也可用密度函数$f(x)$来表示
- 有时也根据总体分布的类型来称呼总体的名称，如正态总体、二项分布总体、0 - 1分布总体
- 若总体分布函数记为$F(x)$, 当从该总体中抽取相互独立同分布(*i.i.d*)的大小为n的样本$X_1, \cdots, X_n$, 则常记为

$$X_1, \cdots, X_n \ i.i.d. \sim F(x)$$

或

$$X_1, \cdots, X_n \ i.i.d. \sim f(x)$$

样本空间的定义

- 定义：样本 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 可能取值的全体, 构成样本空间，记为$\mathcal{X}$

例如，在前述物体重量例子中，样本空间为

$$\mathcal{X} = \{(x_1, \cdots, x_5):\ 0 < x_i < \infty, i = 1, 2, \cdots 5\}$$

例：打靶试验, 每次打三发, 考察中靶的环数. 此时样本空间为

$$\mathcal{X} = \{(x_1, x_2, x_3):\ x_i = 0, 1, \cdots 10, i = 1, 2, 3\}$$

样本的两重性

- 样本的两重性：样本既可看成具体的数，又可以看成随机变量(或随机向量)．在实施抽样后，它是具体的数；在实施抽样前，它被看成随机变量

- 今后用大写的英文字母表示随机变量或随机向量，用小写字母表示它们的具体观察值

简单随机样本

- 定义：设$X_1, \cdots, X_n$为从总体F中抽取的容量为n的样本，若

  1）$X_1, \cdots, X_n$相且独立；

  2）$X_1, \cdots, X_n$相同分布，即同有分布F，

  则称$X_1, \cdots, X_n$为简单随机样本，有时简称为简单样本或随机样本，常简记为*i.i.d.*样本

简单随机样本的联合分布

- 设$X_1, \cdots, X_n$ $i.i.d. \sim F$, 则 $X_1, \cdots, X_n$的联合分布为

$$F(x_1) \cdot F(x_2) \cdots F(x_n) = \prod_{i=1}^{n} F(x_i);$$

- 若F有密度f, 则其联合密度为

$$f(x_1) \cdot f(x_2) \cdots f(x_n) = \prod_{i=1}^{n} f(x_i);$$

- 显然，有放回抽样获得的样本是简单随机样本. 当总体中包含的个体数较大或所抽样本在总体中所占比例较小时，可以把无放回抽样获得的样本当成简单随机样本

样本分布的定义及例子

- 样本既然是随机变量，就有一定的概率分布，这个概率分布就叫作样本分布

例：一大批产品共有N件，其中废品M件，N已知, 而M未知. 现在从中抽出n个检验其中废品的件数，用以估计M或废品率$p = M/N$. 抽样方式为：不放回抽样, 一次抽一个，依次抽取，直到抽完n个为止. 求样本分布

- 由于不放回抽样获得的样本$X_1, \cdots, X_n$不是相且独立的, 样本分布是利用乘法公式，通过条件概率计算出来的

例：如果将上述例子中的抽样方式改为有放回抽样，即每次抽样后记下结果，然后将其放回上，再抽第二个，直到抽完n个为止, 求样本分布

- 本例中样本$X_1, \cdots, X_n$是独立同分布的, 计算样本分布比较容易

- 在上述两个例子中，由于抽样方式不同，获得的样本分布也就不同．这两个例子说明一个重要事实，对于同一个问题，样本分布的结果依赖于不同的抽样方式

例：为估计一物件的重量a, 用一架天平将它重复称n次，结果记为 $X_1, \cdots, X_n$, 求样本$X_1, \cdots, X_n$的联合分布

- 本例中要确定样本的分布，就没有前面两个例子那种简单的算法，需作一些假定：
  1) 假定各次称重是独立进行的；
  2) 假定各次称重是在“相同条件”下进行的，即认为在此条件下获得的样本是同分布的

- 除了上述两个假定外，还要假设随机误差近似服从正态分布，从而导出样本分布

例：某工厂生产一种电子元件，如晶体管．由于大批量生产受随机因素的干扰, 生产出的晶体管的寿命不同. 我们从中抽取n个做寿命试验，用以估计其平均寿命，求样本的分布

- 本例中要确定样本的分布，像上例一样要作一些假定，即样本是相且独立、相同分布的

- 除了做上述假定外，还需要有进一步的假定，如：假设

  1） 寿命具有无后效性；

  2） 设当元件在时刻t尚未失效, 它在$[t, t + \Delta t]$的时间段内失效的概率有$\lambda \Delta t + o(\Delta t)$的形式

统计模型

- 统计模型：指一个问题所对应的抽样样本的分布, 也常称为概率模型或数学模型
  - 显然，只有明确了样本的产生方法, 并辅以必要的假定, 才能确定模型
- 许多不同的问题，在一定条件下可归入到同一模型下；对这一模型种种统计问题的研究反过来又可应用于解决这些性质各异的同一类问题
  - 同一模型下可提出很多不同的统计问题，如对模型中的未知参数提出点估计、区间估计和假设检验问题等

参数空间和样本分布族

- 统计学上把出现在样本分布中的未知的常数称为参数. 参数取值的范围称为参数空间. 如正态分布的参数空间为

$$\Theta = \{(\mu, \sigma^2): -\infty < \mu < +\infty, \sigma^2 > 0\}$$

- 样本分布包含未知参数, 当参数取不同值时就得到不同的分布, 因此这些样本分布就构成一个分布族. 在正态分布中，每个μ, σ^2可能的取值都对应一个具体的分布

- 因此，可以认为统计模型就是样本分布族. 样本分布族从总的方面确定了统计问题的范围

统计推断的概念

- 从总体中抽取一定大小的样本上推断总体的概率分布的方法称为统计推断
  - 统计推断是着手于样本，着眼于总体，其任务是用样本上推断总体

- 参数统计推断的任务是确定样本分布中未知参数的值

- 非参数统计推断的任务是通过样本对总体分布形式作出推断

- 统计推断的形式主要有参数估计和假设检验问题
- 统计推断包括下列三个方面内容:

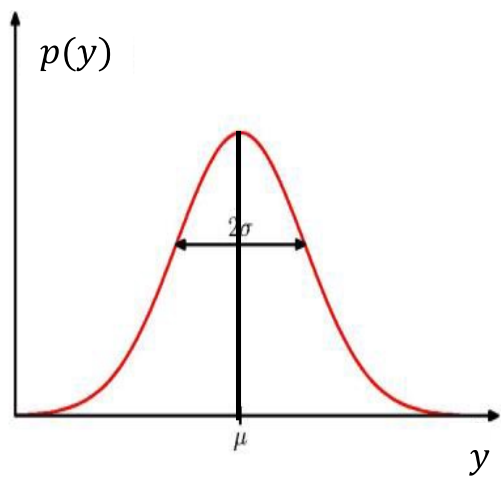
1) 提出种种统计推断的方法;
2) 计算衡量推断正确程度的数量指标;
3) 在一定的条件和优良性准则下寻找最优的统计推断方法

概念解释——线性回归

- Statistical modelling: Given a set of training samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$, we propose to model them with the statistical model

$$p(y_i|\boldsymbol{x}_i;\boldsymbol{w}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\frac{(y_i-\boldsymbol{x}_i\boldsymbol{w})^2}{\sigma^2}\right]$$

$$= \mathcal{N}(y_i;\boldsymbol{x}_i\boldsymbol{w},\sigma^2)$$



$$p(y_1, y_2, \cdots, y_n|\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n;\boldsymbol{w})$$

$$= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\frac{(y_i-\boldsymbol{x}_i\boldsymbol{w})^2}{\sigma^2}\right]$$

- Statistical inference: finding the parameter $\boldsymbol{w}$ that maximizes the log-likelihood, that is,

$$\max_{\boldsymbol{w}} \log p(y|\boldsymbol{x};\boldsymbol{w})$$

Log-likelihood function

- From the expression of $p(y|\boldsymbol{x};\boldsymbol{w})$, we obtain

$$\log p(y|\boldsymbol{x};\boldsymbol{w}) = -\frac{1}{2}\frac{(y-\boldsymbol{x}\boldsymbol{w})^2}{\sigma^2} + constant$$

Thus, maximizing the log-likelihood $\log p(y|\boldsymbol{x};\boldsymbol{w})$ is equivalent to

$$\min_{\boldsymbol{w}} \ (y-\boldsymbol{x}\boldsymbol{w})^2,$$

which is the same as the loss used in the regression

- For N training samples $\left(\boldsymbol{x}^{(i)}, y^{(i)}\right)$, by assuming they are *i.i.d.*, we can obtain their joint conditional pdf

$$p\left(y^{(1)}, \cdots, y^{(N)} \middle| \boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(N)}\right) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{\left(y^{(i)} - \boldsymbol{x}^{(i)}\boldsymbol{w}\right)^2}{\sigma^2}\right]$$

- The log-likelihood function is

$$\log p\left(y^{(1)}, \cdots, y^{(N)} \middle| \boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(N)}\right) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y^{(i)} - \boldsymbol{x}^{(i)}\boldsymbol{w}\right)^2 + constant$$

- Maximizing the log-likelihood $\log p\left(y^{(1)}, \cdots, y^{(N)} \middle| \boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(N)}\right)$ is equivalent to minimize

$$L(\boldsymbol{w}) = \sum_{i=1}^{N}\left(y^{(i)} - \boldsymbol{x}^{(i)}\boldsymbol{w}\right)^2,$$

which is the same as the loss used in the regression

概念解释——逻辑回归

- The Bernoulli distribution

$$p(z) = \begin{cases} \pi, & \text{if } z = 1 \\ 1 - \pi, & \text{if } z = 0 \end{cases}$$

with $\pi \in [0, 1]$ can be equivalently expressed as

$$p(z) = \pi^z \cdot (1 - \pi)^{1-z}$$

- Statistical modelling: Given a set of training samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$, we propose to model them with the statistical model

$$p(y_1, \cdots, y_n | \boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) = \prod_{i=1}^{n} \big(\sigma(\boldsymbol{x}_i \boldsymbol{w})\big)^{y_i} \cdot \big(1 - \sigma(\boldsymbol{x}_i \boldsymbol{w})\big)^{1-y_i}$$

where $y_i = 0$ or 1

- Statistical inference: maximize the log-likelihood function

$$\log p(y_1, \cdots, y_n | \boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$$

$$= \sum_{i=1}^{n} \left[y_i \log \sigma(\boldsymbol{x}_i \boldsymbol{w}) + (1 - y_i) \log\big(1 - \sigma(\boldsymbol{x}_i \boldsymbol{w})\big) \right]$$

Recall that the logistic regression minimizes

$$\textit{cross entropy} \triangleq -y \log \sigma(\boldsymbol{x}\boldsymbol{w}) - (1 - y) \log\big(1 - \sigma(\boldsymbol{x}\boldsymbol{w})\big)$$

Maximizing $\log p(y|\boldsymbol{x})$ is equivalent to minimize the cross entropy

- The well-known linear and logistic regression can be interpreted from the perspective of statistics

1) Statistical modeling: assume different conditional pdfs for the outputs y

    - Regression: *Gaussian distribution*

    - Logistic regression: *Bernoulli distribution*

2) Statistical inference: maximize the log-likelihood functions

概率论与数理统计的关系

- 数理统计学与数学的其它学科有密切的关系，但与数理统计关系最密切的是概率论
- 数理统计进行统计推断的第一步是利用概率论提供的种种模型对数据建模. 一旦确定了模型，就必须把数据看成来自具有一定概率模型的样本，概率论中关于这种模型的理论结果, 就可用于统计推断的目的
- 因此，在很大程度上可以说，概率论是数理统计学的基础, 数理统计学是概率论的一种应用. 但是, 它们是两个并列数学分支学科，并无从属关系

提纲

- 什么是数理统计
- 数理统计的若干概念
- 统计量

统计量的定义

定义：统计量是样本的函数（由样本算出的量）

- 统计量只与样本有关, 不能包含未知参数. 如$X_1,\cdots,X_n \sim N(a,\sigma^2)$, 则$\sum_{i=1}^{n} X_i$和$\sum_{i=1}^{n} X_i^2$ 都是统计量，当a和σ^2皆为未知参数时，$\sum_{i=1}^{n}(X_i - a)$和$\sum_{i=1}^{n} X_i^2/\sigma^2$都不是统计量

- 统计量是样本的函数, 由于样本具有两重性，因此统计量也具有两重性. 即：既可以看成具体的数, 也可以看成随机变量

- 在具体问题中选用什么统计量，要看问题的性质. 一般说来, 希望所提出的统计量应尽可能多的集中样本所含的信息

常用的统计量

设 $X_1,\cdots,X_n$是从某总体X中抽取的样本

- 样本均值：$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$. 它反映了总体均值的信息
- 样本方差：$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$. 它反映了总体方差的信息, 而$S$称为样本标准差, 它反映了总体标准差的信息
- 样本矩：$a_{n,k}=\frac{1}{n}\sum_{i=1}^{n}X_i^k$，为样本$k$阶原点矩

  $m_{n,k}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$, 为样本$k$阶中心矩

二维随机向量的样本矩：设$(X_1,Y_1),\cdots,(X_n,Y_n)$为从二维总体$F(x,y)$中抽取的样本，则

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \qquad S_X^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

$$\bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i \qquad S_Y^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$$

$$S_{XY}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$$

称为X和Y的样本均值、样本方差及X,Y的样本协方差

次序统计量及其相关统计量

- 设$X_1,\cdots,X_n$为从总体F中抽取的样本，将其按大小排列为$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$，则称 $\left(X_{(1)},X_{(2)},\cdots,X_{(n)}\right)$或其任一部分为次序统计量.
  - 样本中位数

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

  - 极值：$X_{(1)}$和$X_{(n)}$称为样本的极小值和极大值，它们统称为样本极值． 极值统计量在关于灾害问题和材料试验的统计分析中是常用的统计量

- 样本p分位数$(0 < p < 1)$：定义为$X_{[(n+1)p]}$，其中$[a]$表示a的整数部分

- 样本极差：定义为$R = X_{(n)} - X_{(1)}$称为样本极差，它是反映总体分布散布程度的信息

样本变异系数、样本偏度、样本峰度

- 样本变异系数：衡量总体分布散布程度的量，但这散布程度是以总体均值为单位来度量

$$\hat{v} = S_n/\bar{X}$$

- 样本偏度：反映总体分布的非对称性的一种度量

$$\hat{\beta}_1 = m_{n,3}/m_{n,2}^{3/2}$$

- 样本峰度：反映总体密度函数在众数(即密度函数的最大值点)附近“峰” 的尖峭程度的一种度量

$$\hat{\beta}_2 = m_{n,4}/m_{n,2}^2 - 3$$

经验分布函数

定义：设 $X_1,\cdots,X_n$为自总体$F(x)$中抽取的$i.i.d.$样本，将其按大小排列为$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$，对任意实数$x$，称下列函数

$$F_n(x) = \begin{cases} 0, & x \leq X_{(1)} \\ \dfrac{k}{n}, & X_{(k)} < x \leq X_{(k+1)} \\ 1, & X_{(n)} < x \end{cases}$$

为经验分布函数

- 由定义可知$F_n(x)$是仅依赖于样本 $X_1,X_2,\cdots,X_n$的函数, 因此, 它是统计量

- 显而易见，经验分布函数是单调、非降、左连续函数，具有分布函数的基本性质. 它在 $x = X_{(k)}$，$k = 2, 3, \cdots, n-1$ 处有间断, 它是在每个间断点跳跃的幅度为$1/n$的阶梯函数

- 格里汶科定理：在大样本情况下（$n \to +\infty$），经验分布函数 $F_n(x)$会收敛于总体分布函数，即：记$D_n = \sup\limits_{-\infty<x<+\infty} |F_n(x) - F(x)|$，那么

$$P\left(\lim_{n\to+\infty} D_n = 0\right) = 1$$