若总体分布未知，要用样本对总体分布进行非参数推断，常用方法是直方图和经验分布函数.

直方图 $\xrightarrow[\text{近似}]{n\to\infty}$ 总体的概率密度图;

经验分布函数 $\xrightarrow[\text{近似}]{n\to\infty}$ 总体的分布函数.

一、直方图

1.问题：设 $X_1, X_2, \cdots, X_n$ 是总体X的一个样本，又设总体X具有概率密度 $f(x)$，如何用样本来推断 $f(x)$?

2.解决方法——直方图:

理论依据：大数定律中频率近似于概率的原理.

(1) 找出 $X_{(1)} = \min\limits_{1\le i\le n} X_i$, $X_{(n)} = \max\limits_{1\le i\le n} X_i$. 取 a 略小于 $X_{(1)}$, b 略大于 $X_{(n)}$.

(2) 将$[a,b]$任意分成m(<n)个小区间，设分点为$a=t_0<t_1<\cdots<t_m=b$，

且记$\Delta t_j=t_j-t_{j-1}$，$j=1,2,\cdots,m$.

(3) 记$n_j=$落入$(t_{j-1},t_j]$中观察值的频数，计算频率$f_j=n_j/n$.

(4) 分别以$(t_{j-1},t_j]$为底边，以$f_j/\Delta t_j$为高作矩形，即得直方图，见图1-1.

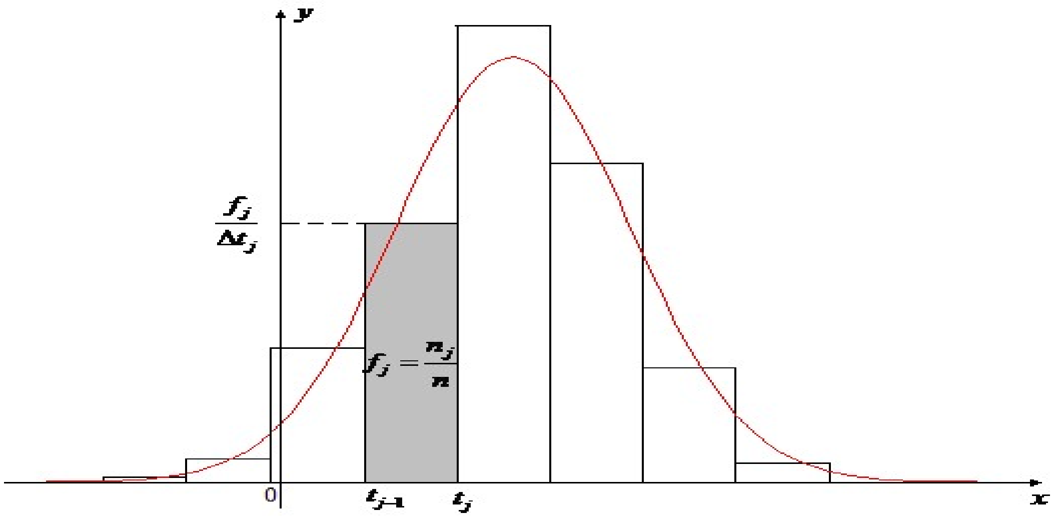


图 1-1

3.直方图的实质：用直方图对应的分段函数

$$\Phi_n(x) = f_j / \Delta t_j, \quad x \in (t_{j-1}, t_j], \quad j = 1,2,\cdots,m$$

来近似总体的概率密度函数 $f(x)$.

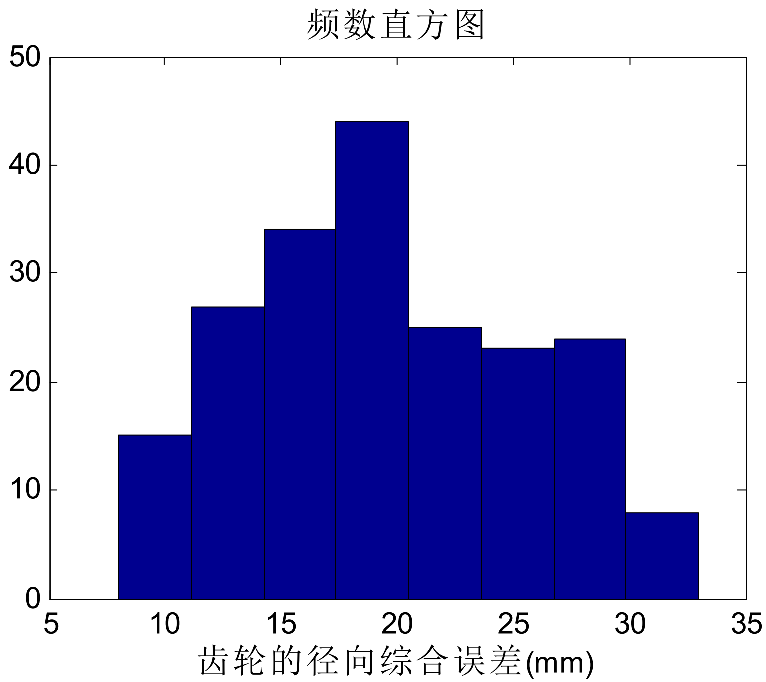
【★例1.6(P10)】在齿轮加工中，齿轮的径向综合误差 $\Delta F_i''$ 是个随机变量，今对200件同样的齿轮进行测量，测得 $\Delta F_i''$ 的数值 (mm) 如下，求作 $\Delta F_i''$ 的直方图.

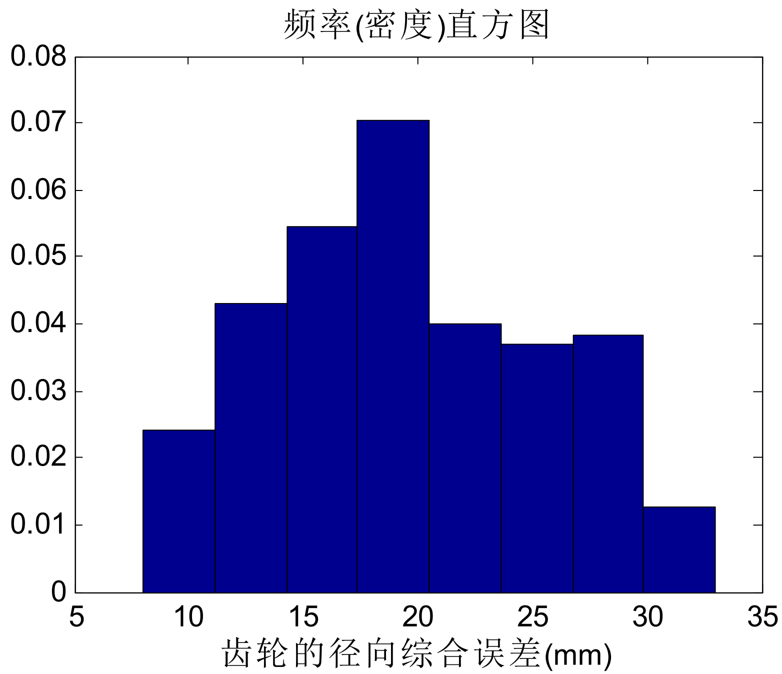
16 25 19 20 25 33 24 23 20 24 25 17 15 21 22 26 15 23 22 24
20 14 16 11 14 28 18 13 27 31 25 24 16 19 23 26 17 14 30 21
18 16 18 19 20 22 19 22 18 26 26 13 21 13 11 19 23 18 24 28
13 11 25 15 17 18 22 16 13 12 13 11 09 15 18 21 15 12 17 13
14 12 16 10 08 23 18 11 16 28 13 21 22 12 08 15 21 18 16 16
19 28 19 12 14 19 28 28 28 13 21 28 19 11 15 18 24 18 16 28
19 15 13 22 14 16 24 20 28 18 18 28 14 13 28 29 24 28 14 18

18 18 08 21 16 24 32 16 28 19 15 18 18 10 12 16 26 18 19 33

08 11 18 27 23 11 22 22 13 28 14 22 18 26 18 16 32 27 25 24

17 17 28 33 16 20 28 32 19 23 18 28 15 24 28 29 16 17 19 18





二、经验分布函数

1.定义1: 设$X_1, X_2, \cdots, X_n$是一样本，记$S(x)$表示落入$(-\infty, x)$中X_i的个数，

则

$$F_n(x) = \frac{S(x)}{n}, \quad -\infty < x < +\infty$$

称为总体X的经验分布函数.

2.定义2: 将样本$X_1, X_2, \cdots, X_n$从小到大排列(重复数据合并)

$$X_{(1)} < X_{(2)} < \cdots < X_{(m)}, \quad (1 \le m \le n)$$

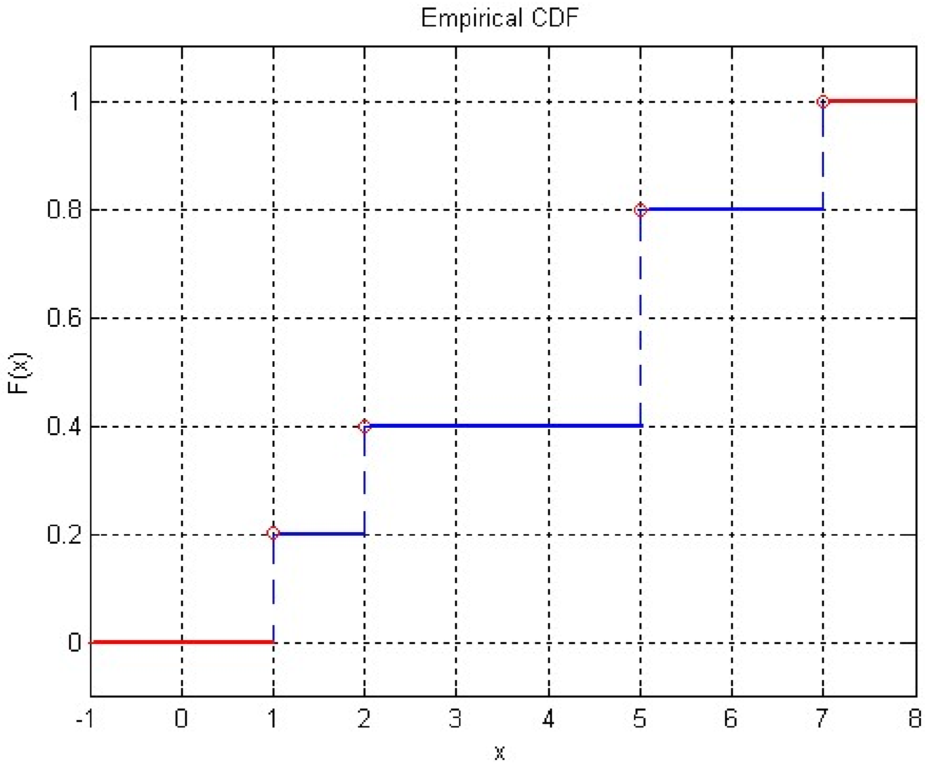
设n_i为$X_{(i)}$的频数，则

$$F_n(x) = \begin{cases} 0, & x \le X_{(1)} \\ \dfrac{n_1 + \cdots + n_i}{n}, & X_{(i)} < x \le X_{(i+1)}, \quad i = 1, 2, \cdots, m-1 \\ 1, & x > X_{(m)} \end{cases}$$

称为总体X的经验分布函数.

【补例1】设总体X的样本为 $X_1=2, X_2=5, X_3=1, X_4=7, X_5=5$，试写出X的经验分布函数并作图.

$$F_5(x)=\begin{cases}0, & x\le 1\\ 0.2, & 1<x\le 2\\ 0.4, & 2<x\le 5\\ 0.8, & 5<x\le 7\\ 1, & x>7\end{cases}$$



3.经验分布函数的性质:

(1) $F_n(x)$只在$x=X_{(i)}, i=1,2,\cdots,m$处出现跳跃间断点;

(2) 若观察值$X_{(i)}$出现了n_i次, 则$F_n(x)$在$X_{(i)}$处的跃度为n_i/n.

(3) 格里文科定理:

若$F(x)$为总体X的分布函数,

$F_n(x)$为X的经验分布函数,

记 $$D_n=\sup_{-\infty<x<+\infty}\left|F_n(x)-F(x)\right|,$$

则 $$\boxed{P\left\{\lim_{n\to\infty}D_n=0\right\}=1}.$$