实际问题

E3 测量某零件长度和直径所产生的误差X.

E4 随意在市场上买来一架电视机, 其使用寿命Y.

取值特点：可取某个区间或$(-\infty,+\infty)$中的一切值.

E3和E4随机变量取值的概率分布如何描述?

能否像离散型随机变量那样用分布律来描述?

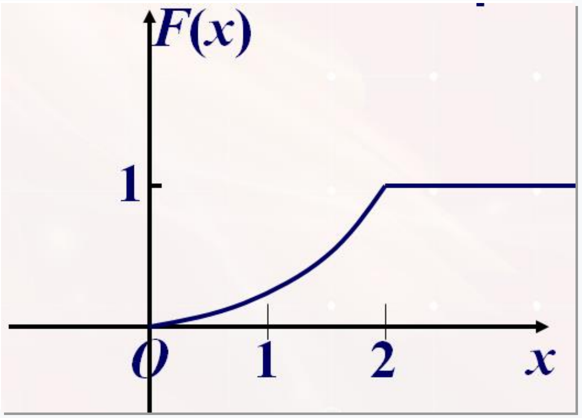
如: 在靶面上指定一个几何意义下(质点:只有位置,无任何向度)的点，则“射击时正好命中该点”的概率，只能取为0?

原因: 若指定一个值a, 则变量X恰好是a一丝不差, 事实上几乎不可能.

例1：一个靶子是半径为2米的圆盘，设击中靶上任一同心圆盘上的点的概率与该圆盘的面积成正比，射击均能中靶，用X 表示弹着点与圆心的距离.

解：X 的分布函数为

$$F(x)=P\{X\le x\}=\begin{cases} 0, & x<0; \\ \dfrac{x^2}{4}, & 0\le x<2; \\ 1, & x\ge 2. \end{cases}$$



$$\forall x\in(-\infty,+\infty), P\{X=x\}=F(x)-F(x-0)=0$$

原因: 若指定一个值a, 则变量X恰好是a一丝不差, 事实上几乎不可能.

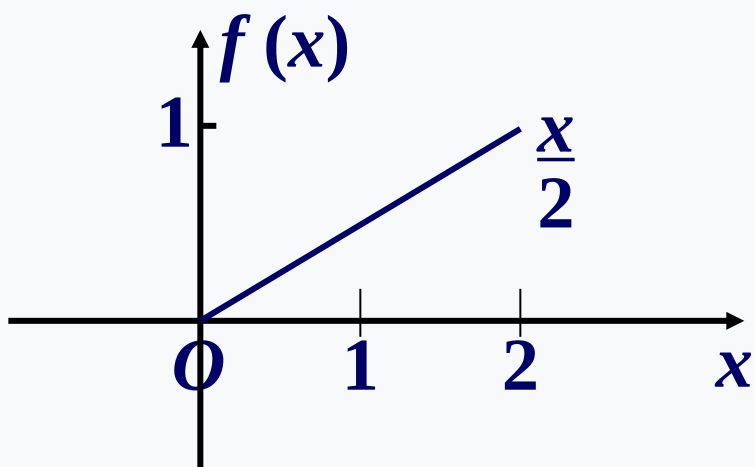
考虑函数 $f(x)=\begin{cases} x/2, & 0\le x<2; \\ 0, & \text{其它} \end{cases}$

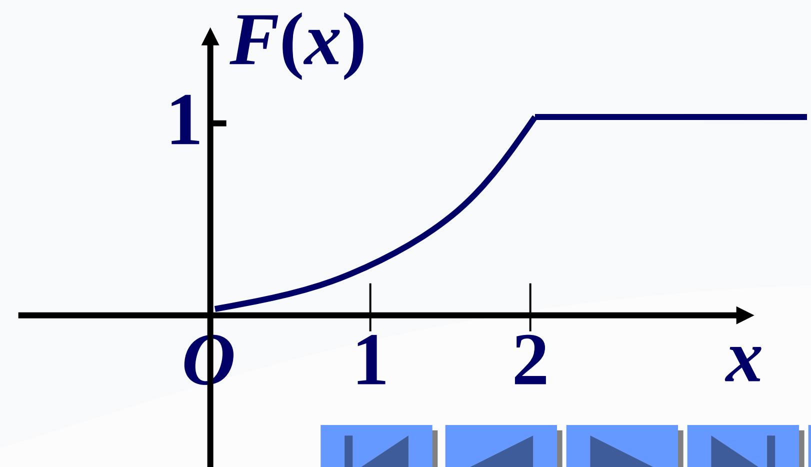
考虑函数 $f(x)=\begin{cases} x/2, & 0<x<2; \\ 0, & 其它 \end{cases}$

$f(x)$的变上限积分为

$$\int_{-\infty}^{x} f(t)dt = \begin{cases} \int_{-\infty}^{x} 0dt = 0, & x<0 \\ \int_{-\infty}^{0} 0dt + \int_{0}^{x} \frac{t}{2}dt = \frac{x^2}{4}, & 0 \le x < 2 \\ \int_{-\infty}^{0} 0dt + \int_{0}^{2} \frac{t}{2}dt + \int_{2}^{x} 0dt = 1, & 2 \le x \end{cases}$$

$= F(x)$





任何类型随机变量:

$$(1)P\{X=x\}=F(x)-F(x-0),\forall x\in(-\infty,+\infty)$$

$$(2)P\{X<x\}=F(x-0),\forall x\in(-\infty,+\infty)$$

(3)$P\{a<X\leq b\}=P\{X\leq b\}-P\{X\leq a\}=F(b)-F(a)$

证明:$(1)P\{X=x\}=\lim\limits_{\Delta x\to0+}P\{x-\Delta x<X\leq x\}$

$$=\lim_{\Delta x\to0+}[P\{X\leq x\}-P\{X\leq x-\Delta x\}]$$

$$=F(x)-F(x-0),\forall x\in(-\infty,+\infty)$$

$$(2)P\{X<x\}=\lim_{\Delta x\to0+}P\{X\leq x-\Delta x\}$$

$$=\lim_{\Delta x\to0+}F(x-\Delta x)=F(x-0),$$

$$\forall x\in(-\infty,+\infty)$$

第三节　连续型随机变量

一、概率密度函数

定义： 设随机变量X 的分布函数为$F(x)$,

若存在非负函数 $f(x)$, 对于任意实数 x , 均有

$$F(x) = P\{X \le x\} = \int_{-\infty}^{x} f(t)dt$$

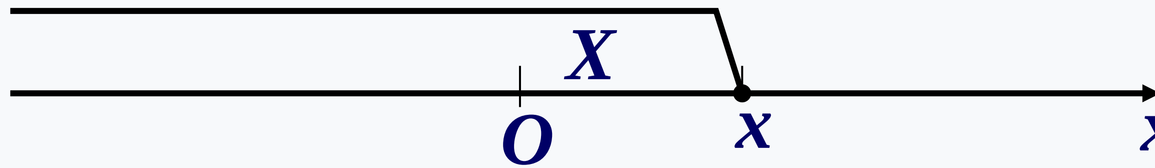
称随机变量X 是连续型随机变量;

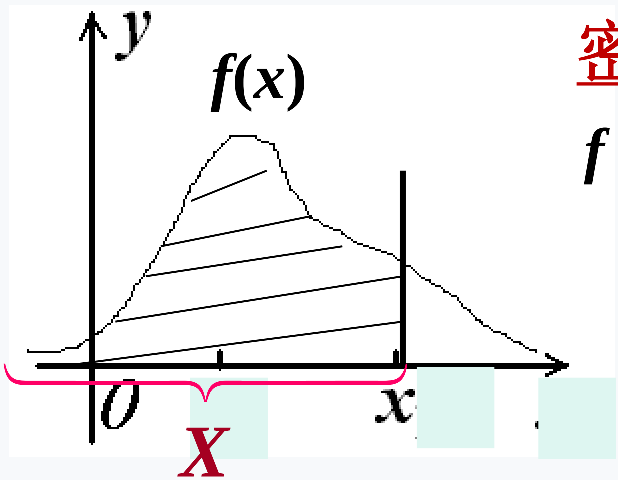
称函数 $f(x)$ 为X的概率密度函数(density function),

简称概率密度.

$$F(x)=P\{X\le x\}=\int_{-\infty}^{x} f(t)dt$$

思考：几何意义？





(概率曲线下面积)

<u>密度函数不是概率,</u> 但在$f(x)$的连续点x处,

$$f(x)\Delta x \approx \int_{x}^{x+\Delta x} f(t)dt$$

$$=\int_{-\infty}^{x+\Delta x} f(t)dt-\int_{-\infty}^{x} f(t)dt$$

$$=F(x+\Delta x)-F(x)=P\{x<X\le x+\Delta x\}$$

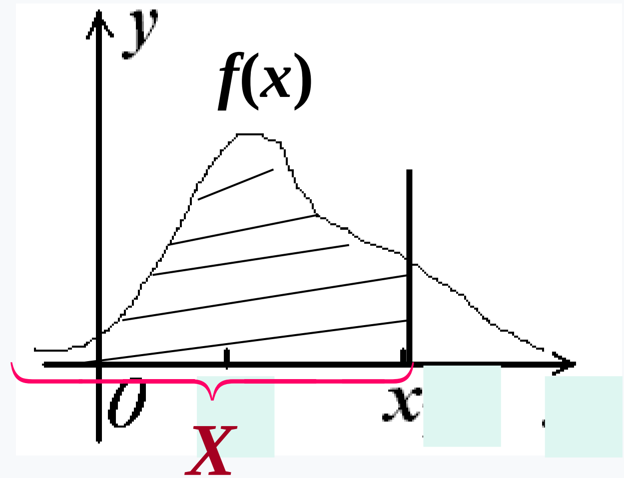
$$\approx P\{X=x\}$$

密度函数$f(x)$的数值反映了随机变量<u>取x的邻近值的概率的大小</u>,

所以用密度函数描述连续型随机变量的概率分布在某种意义上与离散型时用分布律描述, 又有相似之处.

思考1：密度函数$f(x)$是否唯一？

由于在若干个点上甚至一个零测集上改变被积函数$f(x)$的值，都不影响积分$F(x)$值.



(概率曲线下面积)

概率密度函数满足的条件：

（1）$f(x) \geq 0$；（非负性）

（2）$\int_{-\infty}^{+\infty} f(x)dt = 1$.（“归一性”）(概率曲线下面积为1)

若有函数$f(x)$满足上述(1)和(2)，则它必是某个随机变量的概率密度.

概率密度是连续型随机变量的标志.

思考2：密度函数名称的由来？

连续型随机变量相关的性质:

(1) 连续型随机变量X 的分布函数是连续函数.

证: 目的$\forall x \in R$, $F(x)=F(x+)=F(x-)$

由分布函数的性质可知, $F(x)$在x 处右连续.

对于$\Delta x > 0$,

$$0 \le F(x)-F(x-\Delta x)$$

$$=\int_{-\infty}^{x} f(t)dt-\int_{-\infty}^{x-\Delta x} f(t)dt$$

$$=\int_{x-\Delta x}^{x} f(t)\,dt \to 0, \quad \text{当}\Delta x \to 0^{+}$$

即$F(x)$在x 处左连续, 故$F(x)$在x 处连续.

（2）X 是连续型随机变量, 则对任意实数$x_0 \in R$,

$$P\{X = x_0\} = 0.$$

证明：当$\Delta x > 0$, 有$\{X = x_0\} \subset \{x_0 - \Delta x < X \le x_0\}$

从而有$0 \le P\{X = x_0\} \le P\{x_0 - \Delta x < X \le x_0\}$

$$= F(x_0) - F(x_0 - \Delta x)$$

令$\Delta x \to 0$，由$F(x)$的连续性可知有

$$0 \le P\{X = x_0\} = F(x_0) - F(x_0 - \Delta x) \to 0$$

故 $P\{X = x_0\} = 0.$

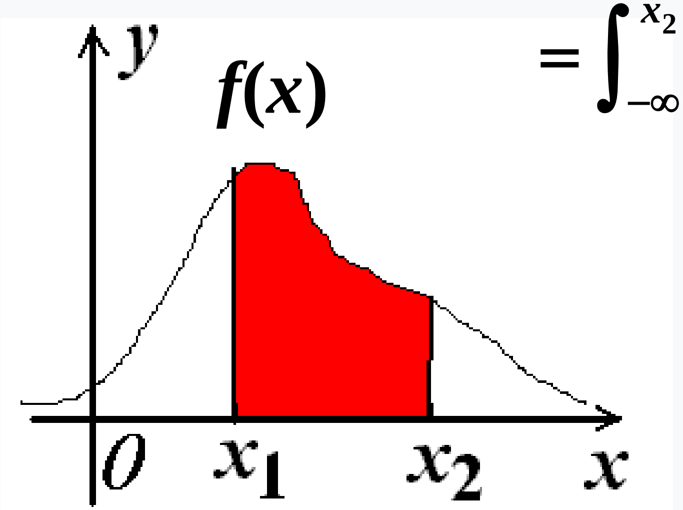
推论: $P(\phi) = 0$, 但是其逆不真,即一个事件的概率等于零, 这事件并不一定是不可能事件.

一个事件的概率等于1, 这事件并不一定是必然事件.

(3) $P\{x_1 < X \le x_2\} = P\{x_1 \le X \le x_2\}$

$= P\{x_1 \le X < x_2\} = P\{x_1 < X < x_2\} = \int_{x_1}^{x_2} f(x)dx$

证明：$P\{x_1 < X \le x_2\} = F(x_2) - F(x_1)$



$= \int_{-\infty}^{x_2} f(t)dt - \int_{-\infty}^{x_1} f(t)dt = \int_{x_1}^{x_2} f(x)dx$

$\{x_1 \le X \le x_2\}$

$= \{x_1 < X \le x_2\} \cup \{X = x_1\}$

及$P\{X = x_1\} = 0$即得上述结果.

注：由（3）知X落在$(x_1, x_2]$的概率$P(x_1 < X \le x_2)$等于区间$(x_1, x_2]$上，曲线$y=f(x)$之下曲边梯形的面积.

（4）若$f(x)$在点x处连续，则有

$$\boxed{F'(x)=f(x)}$$

证明：$F'(x)=\left[\int_{-\infty}^{x} f(t)dt\right]'_x = f(x)$

注：1. $\left[\int_{\varphi_2(x)}^{\varphi_1(x)} g(t)dt\right]' = \varphi_1'(x)g(\varphi_1(x)) - \varphi_2'(x)g(\varphi_2(x))$;

2. $f(x)=\begin{cases} F'(x), & F'(x)\text{存在区间}, \\ 0, & else. \end{cases}$

或者 $f(x)=F'(x),\ x\in(-\infty,+\infty)$

"密度函数"名称的由来

$$f(x)=F'(x)=\lim_{\Delta h\to 0+}\frac{F(x+\Delta h)-F(x)}{\Delta h}(P\{x_1<X\le x_2\})$$

$$1=\int_{-\infty}^{\infty} f(t)dt$$

在x点附近h这么长的区间$(x, x+h)$内，单位长所占有的概率;

这个比的极限，就是在x点处(无穷小区段内)单位长的概率，或者说，它反映了概率在x点处的"密集程度".

设想：一条极细的无穷长的金属杆，总质量为1，概率密度相当于杆上个点的质量密度(线密度).

(1) 连续型随机变量X 的分布函数是连续函数.

(2) X 是连续型随机变量, 则$x_0 \in R$, $P\{ X = x_0 \} = 0$.

(3) $P\{ x_1 < X \le x_2 \} = P\{ x_1 \le X \le x_2 \}$

$= P\{ x_1 \le X < x_2 \} = P\{ x_1 < X < x_2 \} = \int_{x_1}^{x_2} f(x) dx$

(4) 若$f(x)$在点x 处连续, 则有 $F'(x)=f(x)$.

注：以上四个性质可用于概率密度函数判定,

函数参数确定, 概率的计算.

性质的应用实例 ⇒ 函数参数确定 参数确定2

概率的计算1 概率的计算2 概率密度判定

例2：设随机变量X的概率密度函数为

$$f(x)=\begin{cases} k e^{-\frac{x}{\theta}}, & x>\alpha \\ 0, & x\le\alpha \end{cases}$$

试确定常数k.

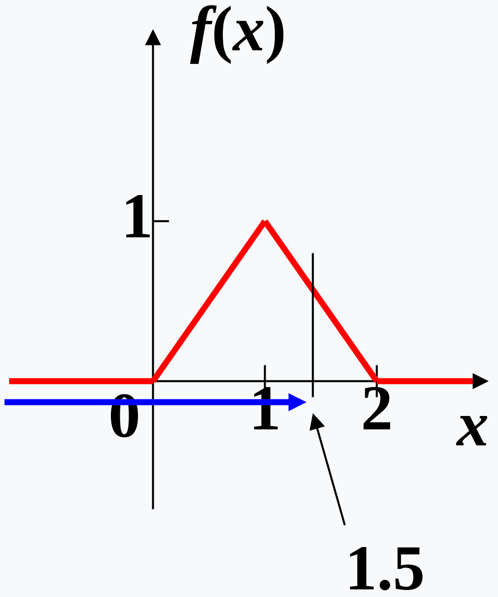
解：$1=\int_{-\infty}^{+\infty} f(x)\,dx=\int_{\alpha}^{+\infty} k e^{-\frac{x}{\theta}}\,dx$

$$=-\theta k\int_{\alpha}^{+\infty} e^{-\frac{x}{\theta}}\,d\left(-\frac{x}{\theta}\right)=-\theta k e^{-\frac{x}{\theta}}\Big|_{\alpha}^{+\infty}$$

$$=\theta k e^{-\frac{\alpha}{\theta}} \Longrightarrow k=\frac{1}{\theta}e^{\frac{\alpha}{\theta}}$$

例4.设连续性随机变量X的概率密度为：

$$f(x)=\begin{cases} x, & 0\le x<1 \\ 2-x, & 1\le x<2 \\ 0, & else \end{cases}$$



则$P(X\le 1.5)=($ $)$

(a) 0.875 (b) $\int_0^{1.5}(2-x)dx$

(c) $\int_0^{1.5} f(x)dx$ (d) $\int_{-\infty}^{1.5}(2-x)dx$

连续型随机变量X的密度函数$f(x)=\begin{cases} x, & 0\le x<1 \\ 2-x, & 1\le x<2 \\ 0, & else \end{cases}$

(a) 0.875　　(b) $\int_0^{1.5}(2-x)dx$

(c) $\int_0^{1.5} f(x)dx$　　(d) $\int_{-\infty}^{1.5}(2-x)dx$

$$P(X\le 1.5)=\int_{-\infty}^{1.5} f(x)dx$$

$$=\int_{-\infty}^{0} 0\,dx+\int_0^1 x\,dx+\int_1^{1.5}(2-x)\,dx =0.875$$

$$P(X\le 1.5)=\int_0^1 x\,dx+\int_1^{1.5}(2-x)\,dx$$

$$=\int_0^{1.5} f(x)\,dx$$

二、均匀分布

密度函数：概率在各处的"密集程度"一样，即概率均匀分布在这区间上(属于几何概率).

设随机变量X的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & else.\end{cases}$$

$$\therefore F(x)=\begin{cases}0, & x\le a,\\ \dfrac{x-a}{b-a}, & a<x\le b,\\ 1, & x>b.\end{cases}$$

称随机变量X在区间(a, b)上服从均匀分布(Uniform distribution)，记为$X \sim U(a, b)$.

$$F_Y(y)=\begin{cases}0, & y\le 0,\\ y, & 0<y<1,\\ 1, & y\ge 1\end{cases}$$

特别地：当$a=0,b=1$时，

$U(0, 1)$为标准的均匀分布.

4. 理由：因为$X \sim U(0,1)$，所以

$$F_X(x)=\begin{cases}0, & x\le 0\\ x, & 0<x\le 1\\ 1, & x>1\end{cases}$$

所以：

常见均匀分布：

（1）均匀分布：$X \sim U(a, b)$

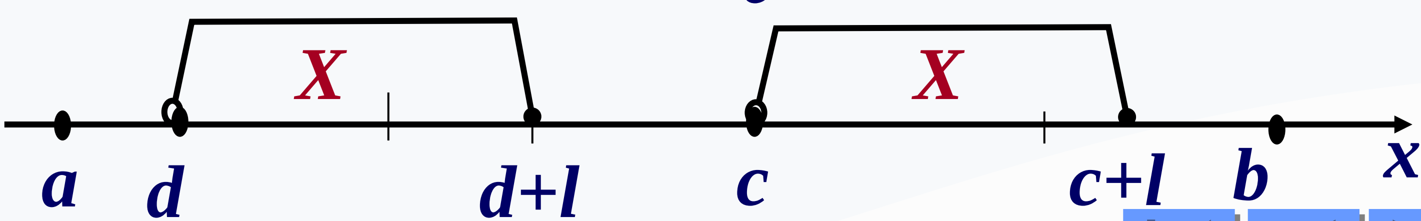
$$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & \text{其它}\end{cases}$$

密度函数：概率在各处的"密集程度"一样，即概率均匀分布在这区间上。

特点：随机变量落在(a, b)的子区间的概率与位置无关，仅与测度(即长度)成正比(几何概率). (与古典概率区别?)

即对于$(c,\ c+l)\subset(a,\ b)$，有

$$P\{c<X\le c+l\}=\int_c^{c+l}\frac{1}{b-a}dx=\frac{l}{b-a}$$



解方程

例6：设随机变量$X \sim U(0, 5)$，求方程$4r^2 + 4Xr + X + 2 = 0$有实根 的概率 p .

解：$p = P\{(4X)^2 - 4\times 4(X+2) \geq 0\}$

$= P\{X^2 - (X+2) \geq 0\}$

$= P\{(X-2)(X+1) \geq 0\}$

$= P(\{X \leq -1\} \cup \{X \geq 2\})$

$= P\{X \leq -1\} + P\{X \geq 2\}$

$= P\{2 \leq X \leq 5\} = \dfrac{5-2}{5} = \dfrac{3}{5}$

应用:

(1) 大量试验服从均匀分布;

例如："四舍五入"产生的误差:

小数点后第一位小数所引起的误差，则可看作是一个服在(-1/2, 1/2)上的均匀分布的随机变量.

(2) 是其它随机变量的计算机模拟的基础.(思考？)

例如：指数分布

方法：反函数法、取舍法、Box-Muller方法.

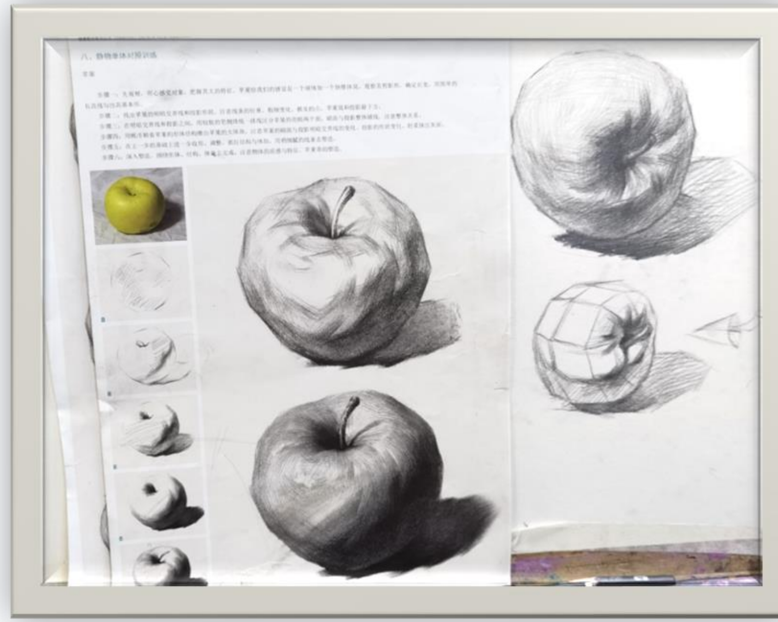
思考：

1. 在实际中如何判断一个随机变量是连续型随机变量?

2. 如何通过一个分布函数来判断随机变量的类型?

本次课的重点是：

连续型随机变量的密度函数与分布函数的判定及性质，均匀分布及相关的计算问题.



下次课内容：

指数分布及正态分布的密度函数与分布函数的性质、计算及应用，二维随机变量联合分布性质.