概率统计课件之二

第二章 随机变量及分布

主讲教师 邓小艳

随机变量及分布

§ **2-1** 随机变量

§ **2-2** 一维离散型***R.V.***及概率分布

§ **2-3** 一维连续型随机***R.V.***及概率分布

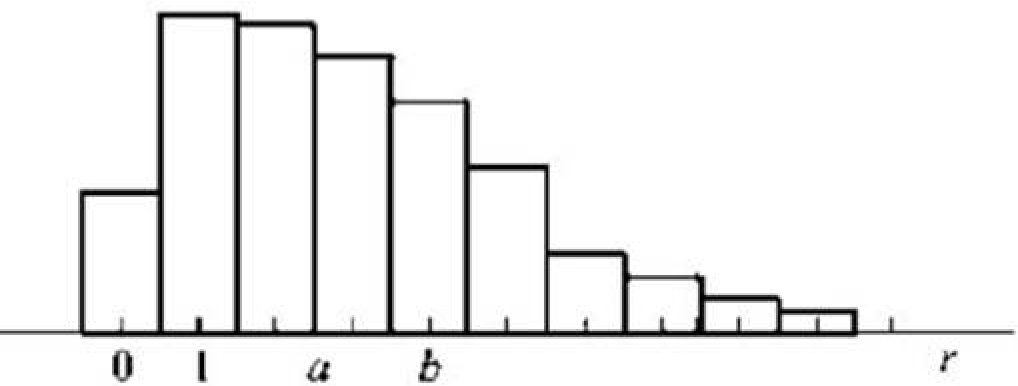
§ **2-5** 二维离散型及***R.V.***概率分布

§ **2-6** 二维连续型***R.V.***及概率分布

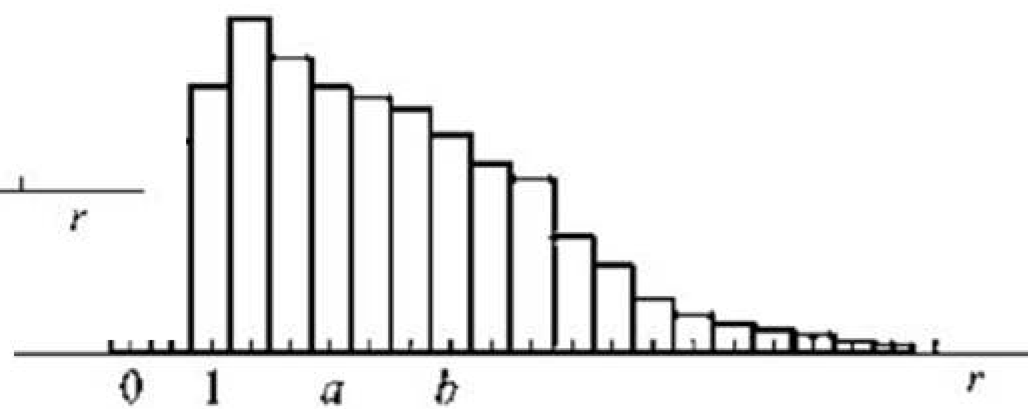
§ **2-8** 相互独立的随机变量

§ **2-9** 随机变量函数的分布

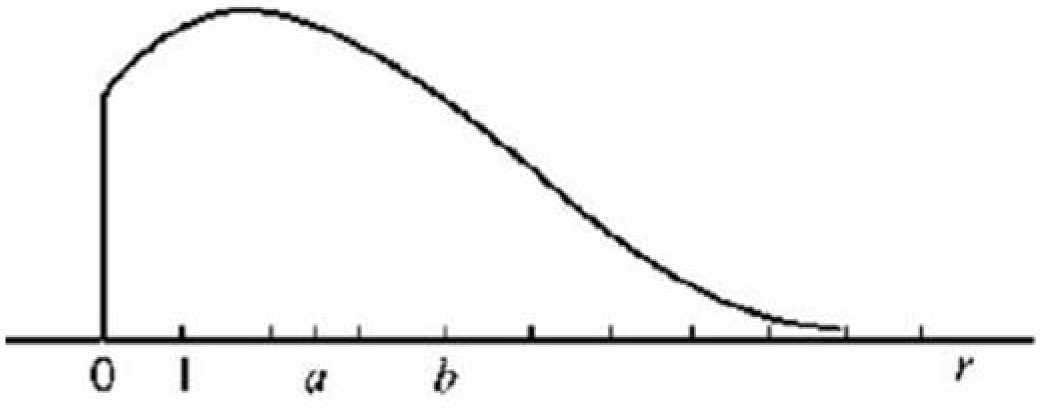
例1　一射击运动员进行射击，设靶是中心在原点，半径为 r 的圆盘. 又设射击不会脱靶. 以 X 记弹着点到靶心的距离，X 是一个随机变量，它可能取的值充满一个区间，显然它不是离散型随机变量. 我们先取 1 cm 作为度量距离的单位，X 取整数值. 这样就将 X 的值离散化，从而得到一个离散型随机变量. 据运动员以前射击的成绩，可以写出这一离散型随机变量的分布律，然后作出对应的概率直方图如图 2 - 3(a) 所示. 接着取 0. 5 cm 作为度量距离的单位，又得到直方图如图 2 - 3(b) 所示. 这样继续缩小度量距离的单位，作出一系列的概率直方图. 这些直方图顶部的台阶型曲线趋于一条光滑曲线 $C:y=f(x)$（如图 2 - 3(c)），



(a)



(b)



(c)

由曲线 C 的行成过程可知：

(1) 曲线 C 位于 ox 轴的上方

(2) 曲线 C 与 ox 轴所夹部分的面积等于1

(3) X 落入 $[a, b]$ 概率等于曲线 C，ox 轴及 $x=a$，$x=b$ 所成的曲边梯形的面积.

推广：将 $f(x)$ 的定义域扩充为 $(-\infty, \infty)$，得到概率密度的定义.

一. 连续型$R.V.$的概率密度

1. 定义： 对**R.V.** X，若存在 $f(x), x \in R$ 满足:

(1) $f(x) \geq 0$ ；

(2) $\int_{-\infty}^{+\infty} f(x)dx = 1$

(3)对$\forall a, b \in R(a \leq b)$(**a**可为$-\infty$，**b**可为$+\infty$)有

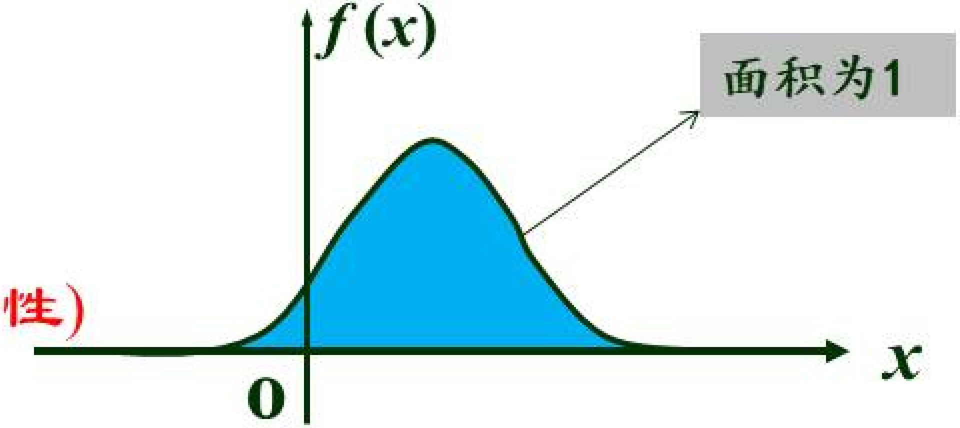
$$P\{a \leq X \leq b\} = \int_a^b f(x)dx$$

则称 **R.V.** X为连续型**R.V.**，称 $f(x)$为**X**的概率密度函数，简称概率密度。

2. 概率密度的性质

(1) $f(x) \geq 0$ (非负性)

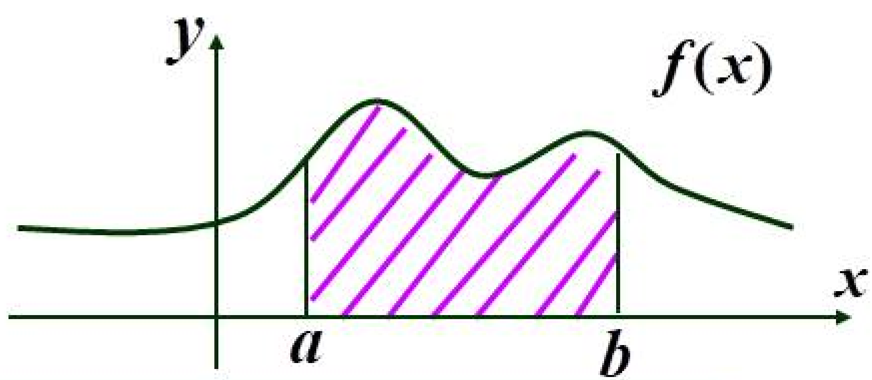
(2) $\int_{-\infty}^{\infty} f(x)dx = 1$ (归一性)



这两条是判定一个函数$f(x)$是否为某$R.V.\ X$ 的概率密度的充要条件

(3) 对于任意实数 $a, b, (a \leq b)$,有

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$



该条性质表明利用概率密度可确定随机点落入任意一个区间的概率

几点说明：

① $P\{X=a\}=0$，即连续型**R.V.**取任一单点值的概率为**0.**

$$P(X=a)=\int_a^a f(x)dx=0$$

② $P\{a\le X\le b\}=P\{a<X\le b\}=P\{a\le X<b\}$

$$=P\{a<X<b\}=\int_a^b f(x)dx$$

$$P\{X>a\}=\int_a^{+\infty} f(x)dx$$

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

③ 对连续型 $r.v$ X, 改变被积函数 $f(x)$ 在有限个点的函数值不影响积分 $\int_a^b f(x)dx$ 的值，即不影响概率 $P\{a \le X \le b\}$, 因此对于概率密度 $f(x)$ 来说，改变它在有限个点上的值，是被允许的.

④ 若 $f(x)$ 在点 x 处连续，则有

$$\lim_{\Delta x \to 0^+} \frac{P\left(x < X \le x + \Delta x\right)}{\Delta x} = \lim_{\Delta x \to 0^+} \frac{\int_x^{x+\Delta x} f(x)dx}{\Delta x}$$

$$= \lim_{\Delta x \to 0^+} \frac{f(\xi)\Delta x}{\Delta x} = f(x)$$

故 X的密度 $f(x)$ 在 x 这一点的值，恰好是 X落在区间$(x, x+\Delta x]$上的概率与区间长度 Δx 之比的极限. 这里，如果把概率理解为质量，$f(x)$相当于线密度.

例1：连续型随机变量X的概率密度函数为：

$$f(x)=\begin{cases} kx & 0\le x<3 \\ 2-\dfrac{x}{2} & 3\le x<4 \\ 0 & \text{其它} \end{cases}$$

求：(1) 系数k；

(2) $P\{X\le 1\}$； $P\{1\le X\le \frac{7}{2}\}$；

$P\{1<X<\frac{7}{2}\}$；$P\{X>3\}$；$P\{X=3.5\}$；

(3) X的分布函数；

解：(1) 由概率密度 $f(x)$ 的性质 $\int_{-\infty}^{+\infty} f(x)dx = 1$，得

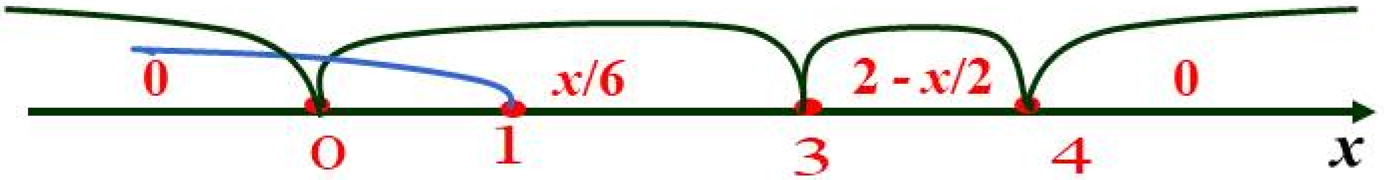
$$1 = \int_{-\infty}^{+\infty} f(x)dx$$

$$= \int_{-\infty}^{0} 0dx + \int_{0}^{3} kxdx + \int_{3}^{4} (2 - \frac{x}{2})dx + \int_{4}^{+\infty} 0dx$$

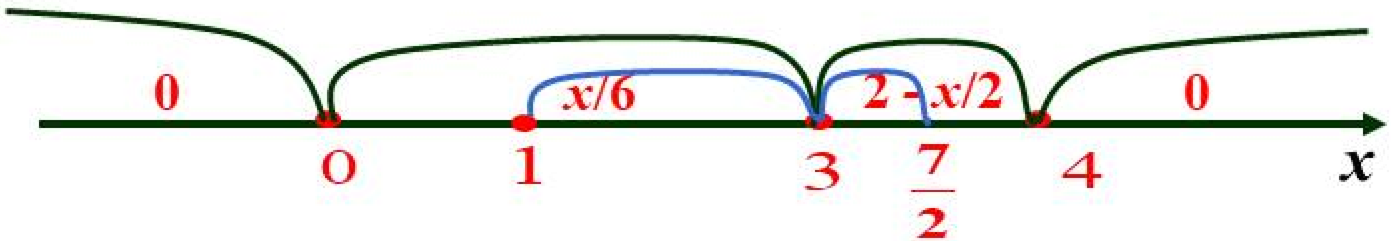
即 $\dfrac{9k}{2} + \dfrac{1}{4} = 1 \Rightarrow k = \dfrac{1}{6}$

∴ 概率密度为：

$$f(x) = \begin{cases} \frac{1}{6}x & 0 \le x < 3 \\ 2 - \frac{x}{2} & 3 \le x < 4 \\ 0 & \text{其它} \end{cases}$$



$$P\{X \le 1\} = \int_{-\infty}^{0} 0dx + \int_{0}^{1} \frac{x}{6} dx = \int_{0}^{1} \frac{1}{6} xdx = \frac{1}{12}$$

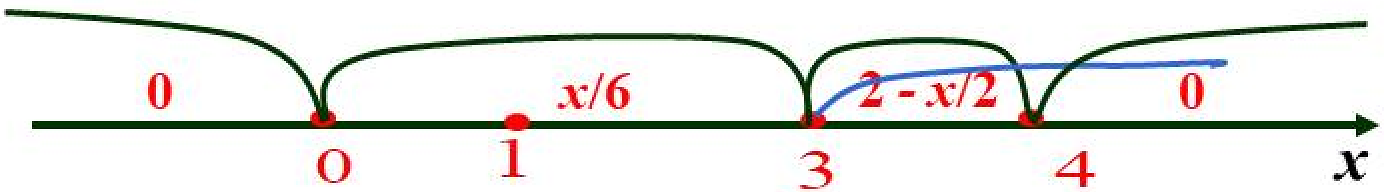


$$P\{1 \le X \le \frac{7}{2}\} = \int_{1}^{\frac{7}{2}} f(x)dx$$

$$= \int_{1}^{3} \frac{1}{6} xdx + \int_{3}^{\frac{7}{2}} (2 - \frac{x}{2})dx = \frac{41}{48}$$

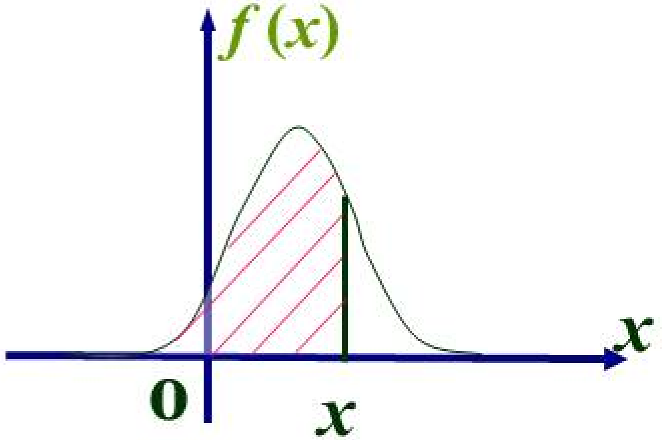
$$P\{1 < X < \frac{7}{2}\} = P\{1 \le X \le \frac{7}{2}\} = \frac{41}{48}$$

$$P\{\mathbf{X}>3\}=\int_3^{+\infty} f(x)dx=\int_3^4 (2-\frac{x}{2})dx+\int_4^{+\infty} 0dx=\frac{1}{4}$$



$$P\{\mathbf{X}=3.5\}=0$$

二. 连续型R.V.的分布函数



1. 定义：连续型R.V.的分布函数

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

$F(x)$在点x的函数值，等于曲线$f(x)$之下，Ox轴的区间$(-\infty,x]$之上的曲边梯形的面积.

2、性质

(1) $F(x)$ 是单调不减函数

(2) $F(x)$ 是连续函数

(3) 若$f(x)$在x处连续，则 $F'(x)=\left[\int_{-\infty}^{x} f(t)dt\right]'=f(x)$

(4) $P\{a<X\le b\}=P\{a\le X\le b\}=P\{a<X\le b\}$

$=P\{a<X<b\}=F(b)-F(a)$ （$\forall a,b$）

例1：连续型随机变量X的概率密度函数为：

$$f(x)=\begin{cases} kx & 0\le x<3 \\ 2-\dfrac{x}{2} & 3\le x<4 \\ 0 & \text{其它} \end{cases}$$

求：(1) 系数k；

(2) $P\{X\le 1\}$；　$P\{1\le X\le \frac{7}{2}\}$；

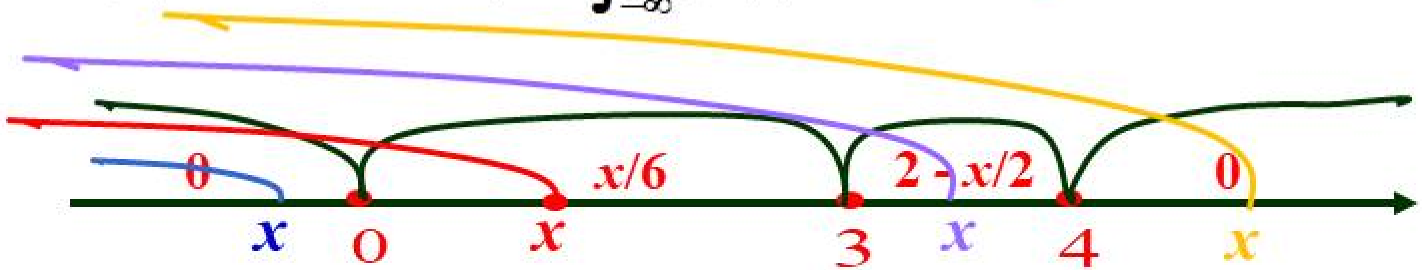
$P\{1<X<\frac{7}{2}\}$；$P\{X>3\}$；$P\{X=3.5\}$；

(3) X的分布函数；

(3) X的分布函数为：

步骤：①定义②分区间（根据f(x)的取值区间划分）③对于不同区间确定积分区间和被积函数④计算积分得分布函数

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



$$=\begin{cases} \int_{-\infty}^{x} 0dt & x<0 \\ \int_{-\infty}^{0} 0dt+\int_{0}^{x}\frac{1}{6}tdt & 0\le x<3 \\ \int_{-\infty}^{0} 0dt+\int_{0}^{3}\frac{1}{6}tdt0+\int_{3}^{x}(2-\frac{t}{2})dt & 3\le x<4 \\ \int_{-\infty}^{0} 0dt+\int_{0}^{3}\frac{1}{6}tdt0+\int_{3}^{4}(2-\frac{t}{2})dt+\int_{4}^{+\infty}0dt & x\ge 4 \end{cases}$$

$$
=\begin{cases}
0 & x<0 \\
\dfrac{x^2}{12} & 0\le x<3 \\
-3+2x-\dfrac{x^2}{4} & 3\le x<4 \\
1 & x\ge 4
\end{cases}
$$

(2) 另解：

$$P\{x\le 1\}=F(1)=\frac{1}{12}$$

$$P\{1<x<\frac{7}{2}\}=P\{1\le x\le \frac{7}{2}\}=F(\frac{7}{2})-F(1)=\frac{41}{48}$$

$$P\{x>3\}=1-F(3)=1-(-3+2\times 3-\frac{3^2}{4})=\frac{1}{4}$$

例2：连续型*R.V.* *X*的概率密度为：

$$f(x) = Ae^{-|x|} \quad -\infty < x < \infty$$

求：(1) 常数A；

(2) X的分布函数

(3) $P\{X = 1/2\}$；$P\{X^2 \le 1\}$；

例2 解：(1) 由 $\int_{-\infty}^{+\infty} f(x)\,dx = 1$ 得，

$$\int_{-\infty}^{0} Ae^{x}dx + \int_{0}^{+\infty} Ae^{-x}dx = A + A = 2A = 1$$

$\therefore A = \frac{1}{2}$

$$\therefore f(x) = \begin{cases} \frac{1}{2}e^{x} & x < 0 \\ \frac{1}{2}e^{-x} & x \geq 0 \end{cases}$$

(2) X 的分布函数

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

$$= \begin{cases} \frac{1}{2}e^{x} & x > 0 \\ 1 - \frac{1}{2}e^{-x} & x \leq 0 \end{cases}$$

(3) $P\{X^2 \leq 1\} = P\{-1 \leq X \leq 1\} = \int_{-1}^{1} f(x)\,dx = F(1) - F(-1) = 1 - e^{-1}$

$P\{X = \frac{1}{2}\} = 0$

例3:设连续型随机变量X的分布函数为

$$F(x)=\begin{cases}0, & x\le -a\\ A+B\arcsin\dfrac{x}{a}, & -a<x\le a\\ 1, & x>a\end{cases}$$

求：(1) 系数A,B的值；

(2) $P\{-a<X<\dfrac{a}{2}\}$;

(3) X的概率密度函数.

解:(1) $\because$ X是连续型随机变量

$\therefore\ F(x)$ 连续

故有 $F(-a)=\lim\limits_{x\to -a}F(x),\ F(a)=\lim\limits_{x\to a}F(x),$

即
$$\begin{cases} A+B\arcsin\left(\dfrac{-a}{a}\right)=A-\dfrac{\pi}{2}B=\mathbf{0} \\ A+B\arcsin\left(\dfrac{a}{a}\right)=A+\dfrac{\pi}{2}B=\mathbf{1} \end{cases}$$

解得: $\boldsymbol{A}=\dfrac{\mathbf{1}}{\mathbf{2}},\qquad \boldsymbol{B}=\dfrac{\mathbf{1}}{\boldsymbol{\pi}}$

∴ **X的分布函数为**

$$F(x)=\begin{cases}0, & x\le -a\\ \dfrac{1}{2}+\dfrac{1}{\pi}\arcsin\dfrac{x}{a}, & -a<x\le a\\ 1, & x>a\end{cases}$$

$$(2)\ P\{-a<X<\frac{a}{2}\}=F(\frac{a}{2})-F(-a)$$

$$=\frac{1}{2}+\frac{1}{\pi}\arcsin(\frac{a}{2a})-0=\frac{1}{2}+\frac{1}{\pi}\times\frac{\pi}{6}=\frac{2}{3}.$$

(3) $\because$ 在 $f(x)$ 连续点处有：$f(x)=F'(x)$

令 $f(-a)=f(a)=0$

$\therefore$ 机变量X的密度函数为:

$$f(x)=F'(x)=\begin{cases} 1\big/\pi\sqrt{a^2-x^2}, & -a<x<a \\ 0, & \text{其它}. \end{cases}$$

例4：连续型*R.V. X*的分布函数为：

$$F(x)=\begin{cases}0 & x<1\\ \ln x & 1\le x<e\\ 1 & x\ge e\end{cases}$$

求：(1) $P\{X<2\}$； $P\{0<X\le 3\}$；

$P\{2<X<\frac{5}{2}\}$；

(2) X的概率密度函数；

三、几个常用的连续型分布

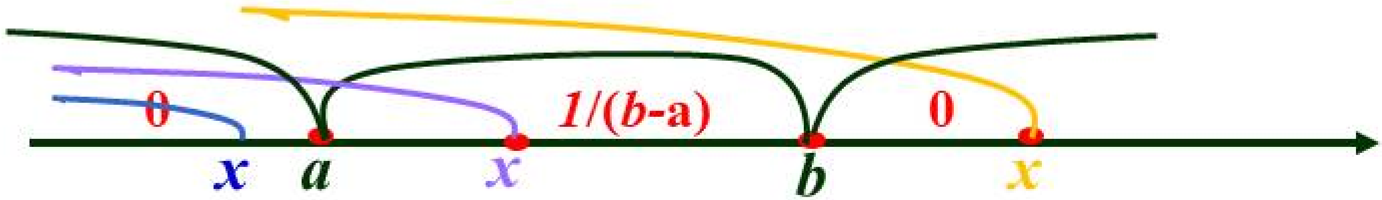
(一) 均匀分布

1、定义：若连续型R.V.X的概率密度为：

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

则称X在(a，b)上服从均匀分布，记作：$X \sim U(a,b)$

2、分布函数



$$F(x)=\int_{-\infty}^{x} f(t)dt=\begin{cases}\int_{-\infty}^{x} 0dt & x<a\\ \int_{-\infty}^{a} 0dt+\int_{a}^{x}\frac{1}{b-a}dt & \mathrm{a}\le x<b\\ \int_{-\infty}^{a} 0dt+\int_{a}^{b}\frac{1}{b-a}dt+\int_{b}^{x} 0dt & \mathrm{b}\le x\end{cases}$$

$$=\begin{cases}0 & x<a\\ \dfrac{x-a}{b-a} & a\le x<b\\ 1 & b\le x\end{cases}$$

3、性质：若$X\sim U(a,b)$，则X在(a,b)中等长子区间内取值的概率相等。即X落入任何子区间的概率仅与子区间的长度成正比，而与子区间的位置无关。

$$P\{x<X<x+l\}=\int_x^{x+l}\frac{1}{b-a}dt$$

$$=F(x+l)-F(x)=\frac{x+l}{b-a}-\frac{x}{b-a}$$

$$=\frac{l}{b-a}$$

例4：(候车问题)公共汽车每10分钟按时通过一车站，一乘客在随机选择的时间到达车站，以X记他等车时间(以分钟计)，则X是一个R.V.。求：(1) X的概率密度函数；(2) 他等候时间少于3分钟的概率；(3) 他等候时间在3~6分钟的概率；

解：(1) X的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{10} & 0<x<10\\ 0 & \text{其它}\end{cases}$$

(2) 他等候时间少于3分钟的概率：$P(X<3)=\dfrac{3-0}{10-0}=\dfrac{3}{10}$

(3) 他等候时间在3~6分钟的概率：$P(3\le X\le 6)=\dfrac{3}{10}$

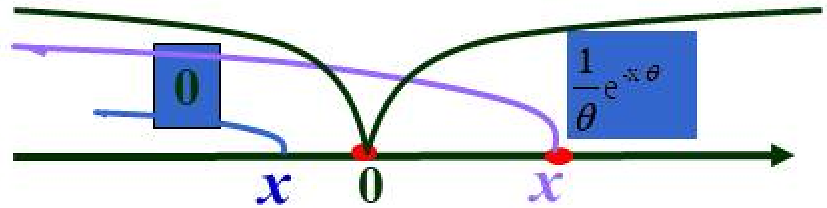
(二) 指数分布

1、定义： 若连续型**R.V.X**的概率密度为：

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-x/\theta} & x>0\\ 0 & \text{其它}\end{cases}$$

则称**X**服从参数为 $\theta(\theta>0)$ 指数分布。

2、分布函数



$$F(x)=\int_{-\infty}^{x} f(t)dt=\begin{cases}\int_{-\infty}^{x}0dt & x\le 0\\ \int_{-\infty}^{0}0dt+\int_{0}^{x}\frac{1}{\theta}e^{x/\theta}dt & x>0\end{cases}$$

$$=\begin{cases}0 & x\le 0\\ 1-e^{-x/\theta} & x>0\end{cases}$$

例5：设某种电子元件寿命X(以年记)服从参数 $\theta = 3$ 的指数分布，求：

(1) 寿命超过2年的概率；

(2) 设已经正常使用了S年，求至少还能使用t年的概率；

解：X的概率密度为：$f(x)=\begin{cases}\frac{1}{3}e^{-x/3} & x>0\\ 0 & 其它\end{cases}$

X的分布函数为：$F(x)=\begin{cases}1-e^{-x/3} & x>0\\ 0 & 其它\end{cases}$

(1) 寿命超过2年的概率为：

$$P\{X>2\}=1-F(2)=1-(1-e^{-2/3})=e^{-2/3}$$

(2) 设已经正常使用了S年，求至少还能使用t年的概率：

$$P\{X>s+t \mid X>s\}=\frac{P\{(X>s+t)\cap(X>s)\}}{P\{X>S\}}$$

$$=\frac{P\{X>s+t)\}}{P\{X>S\}}=\frac{1\text{-}F(s+t)}{1\text{-}F(s)}=\frac{e^{-(s+t)/3}}{e^{-s/3}}=e^{-t/3}$$

$$P\{X>t\}=1-F(t)=e^{-t/3}$$

$$\therefore\ P\{X\geq s+t \mid X>s\}=P\{X\geq t\}$$

3、性质

$$P\{X\geq s+t \mid X>s\}=P\{X\geq t\}$$ ——无记忆性

例6： 设某种电子元件寿命X(以年记)服从指数分布，其概率密度为：

$$f(x)=\begin{cases}\dfrac{1}{100}\mathrm{e}^{-x/100} & x>0\\ 0 & \text{其它}\end{cases}$$

(1) 元件寿命至少为200小时的概率；

(2) 将3只这种元件联接成一个系统，设系统工作的 方式是至少2只元件失效时系统失效，又设3只元件工作相互独立。求系统的寿命至少为200小时的概率；

分析：

(1) 元件寿命至少为200小时的概率

即是求 $P\{X \geq 200\}$

(2) 将3只这种元件联接成一个系统，设系统工作的 方式是至少2只元件失效时系统失效，又设3只元件工作相互独立。求系统的寿命至少为200小时的概率；

由于系统工作的方式是至少2只元件失效时系统失效，即是说至少2只元件正常工作时系统正常工作。故求系统的寿命至少为200小时的概率，即是求至少2只元件的寿命为至少200小时的概率。

解: (1) 元件寿命至少为200小时的概率为

$$P\{X \geq 200\} = \int_{200}^{+\infty} f(x)dx = \int_{200}^{\infty} \frac{1}{100} e^{-\frac{x}{100}} dx$$

$$= -e^{-x/100}\Big|_{200}^{\infty} = e^{-2} \text{ (小时)}$$

(2) 令Y: “3只元件中寿命至少为200小时的元件的只数.

∵各元件的工作相互独立，又由(1)知一元件寿命至少为200小时的概率为 e^{-2}

$$\therefore\ Y \sim B(3, e^{-2})$$

$$\therefore \quad P\{Y \geq 2\} = P\{Y=2\} + P\{Y=3\}$$

$$= C_3^2 (e^{-2})^2 (1-e^{-2})^{3-2} + C_3^3 (e^{-2})^3$$

$$= 3e^{-4} - 2e^{-6} = 0.05$$

故系统的寿命至少为200小时的概率为0.05

(三) 正态分布

1、定义： 若连续型R.V.X的概率密度为：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\quad(-\infty<x<+\infty)$$

则称X服从参数为 μ,σ^2 的正态分布或高斯分布，记作：$X\sim N(\mu,\sigma^2)$

2、标准正态分布

当 $\mu=0,\sigma^2=1$时，称X服从标准正态分布，记作：$X\sim N(0,1)$，其概率密度为：

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\quad(-\infty<x<+\infty)$$

小　　结

- 连续型随机变量
  - 定义
  - 概率密度
    - 定义
    - 性质
  - 几种常见的连续型随机变量的分布
    - 均匀分布
    - 指数分布
    - 正态分布
  - 4.分布函数
    - 定义
    - 性质