

复习

随机变量 $X: S \rightarrow R^1$

$$\omega \rightarrow X(\omega)$$

离散型随机变量 分布列

$$P(X=x_i)=p_i, i=1, 2, \dots$$

几种重要的离散型随机变量



§ 3 随机变量的分布函数



一 分布函数的概念与性质

对于随机变量 X ，我们不仅要知道 X 取哪些值，还要知道 X 取这些值的概率；而且更重要的是想知道 X 在任意有限区间内取值的概率



无论是离散型还是连续型，以至其它类型的
随机变量

对事件 $\{X \in (a, b]\}$ 的概率，都有

$$\begin{aligned} P(X \in (a, b]) &= P(a < X \leq b) \\ &= P(X \leq b) - P(X \leq a) \end{aligned}$$

为了对不同类型的随机变量给出一种
统一的描述方法，我们引进分布函数的概念

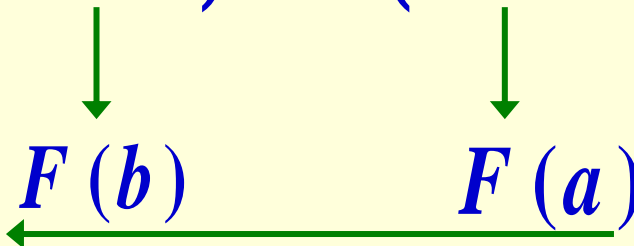


事实上，如果我们定义

$$F(x) = P(X \leq x)$$

则上述概率

$$\begin{aligned} P(X \in (a, b]) &= P(a < X \leq b) \\ &= P(X \leq b) - P(X \leq a) \end{aligned}$$



分布
函数

即 $P(X \in (a, b]) = F(b) - F(a)$

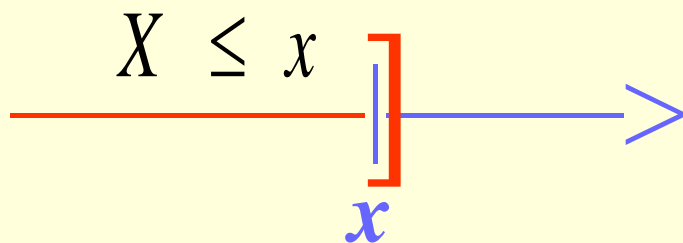


分布函数

设 X 是一随机变量，对任意的实数 x ,令

$$F(x) = P(X \leq x)$$

则称 $F(x)$ 为 X 的分布函数。



$$F(x) = P(X \leq x), \quad -\infty < x < \infty$$

a 在上式中, X 是随机变量, x 是自变量.

$F(x)$ 是 $r.v$ X 取值不大于 x 的概率.

b 对任意实数 $x_1 < x_2$,

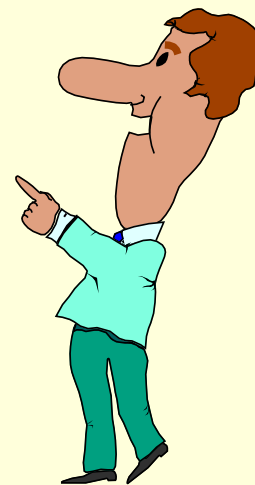
$$\begin{aligned} P\{x_1 < X \leq x_2\} &= P\{X \leq x_2\} - P\{X \leq x_1\} \\ &= F(x_2) - F(x_1) \end{aligned}$$

因此, 只要知道了随机变量 X 的分布函数, 它的统计特性就可以得到全面的描述.



$$F(x) = P(X \leq x), \quad -\infty < x < \infty$$

c 分布函数是一个普通的函数，正是通过它，我们可以用数学分析的工具来研究 随机变量.



离散型随机变量的分布函数

设 $P(X=x_i)=p_i, i=1, 2, \dots$

$$F(x) = P(X \leq x) = P\left(\bigcup_{x_k \leq x} (X = x_k)\right)$$

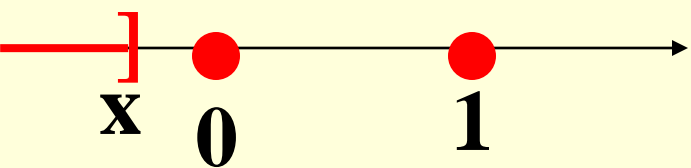
$$= \sum_{x_k \leq x} P(X = x_k) = \sum_{x_k \leq x} p_k$$



例1: 设 X 服从参数为 p 的二点分布, 即:

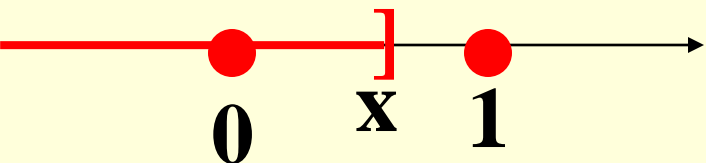
$$P(X = k) = p^k q^{1-k} \quad k = 0, 1$$

其中 $0 < p < 1, q = 1 - p$ 。求 X 的分布函数 $F(x)$ 。



解: $F(x) = P(X \leq x)$

当 $x < 0$ 时, $F(x) = 0$



当 $0 \leq x < 1$ 时,

$$F(x) = P(X=0) = q$$

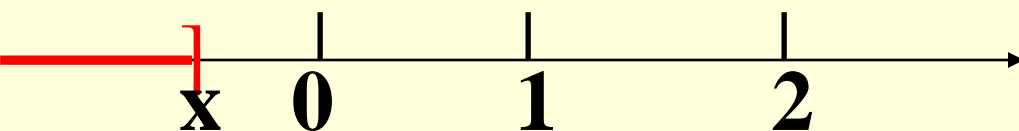


当 $x \geq 1$ 时,

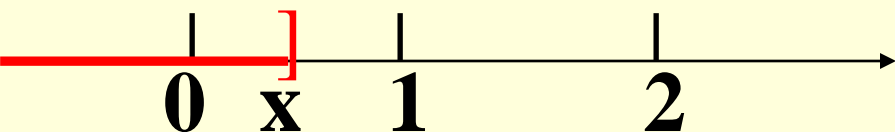
$$F(x) = P(X=0) + P(X=1) = 1$$

例2 $X \sim \begin{Bmatrix} 0 & 1 & 2 \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{Bmatrix}$, 求 $F(x)$.

解: $F(x) = P(X \leq x)$



当 $x < 0$ 时, $F(x) = 0$



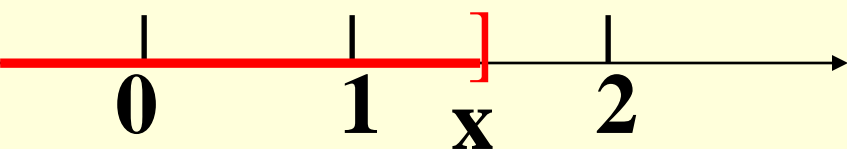
当 $0 \leq x < 1$ 时,

$$F(x) = P(X=0) = 1/3$$



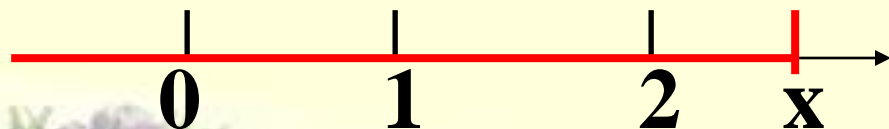
例2 $X \sim \left\{ \begin{array}{ccc} 0 & 1 & 2 \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{array} \right\}$, 求 $F(x)$.

解: $F(x) = P(X \leq x)$



当 $1 \leq x < 2$ 时,

$$F(x) = P(X=0) + P(X=1) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}$$



当 $x \geq 2$ 时,

$$F(x) = P(X=0) + P(X=1) + P(X=2) = 1$$

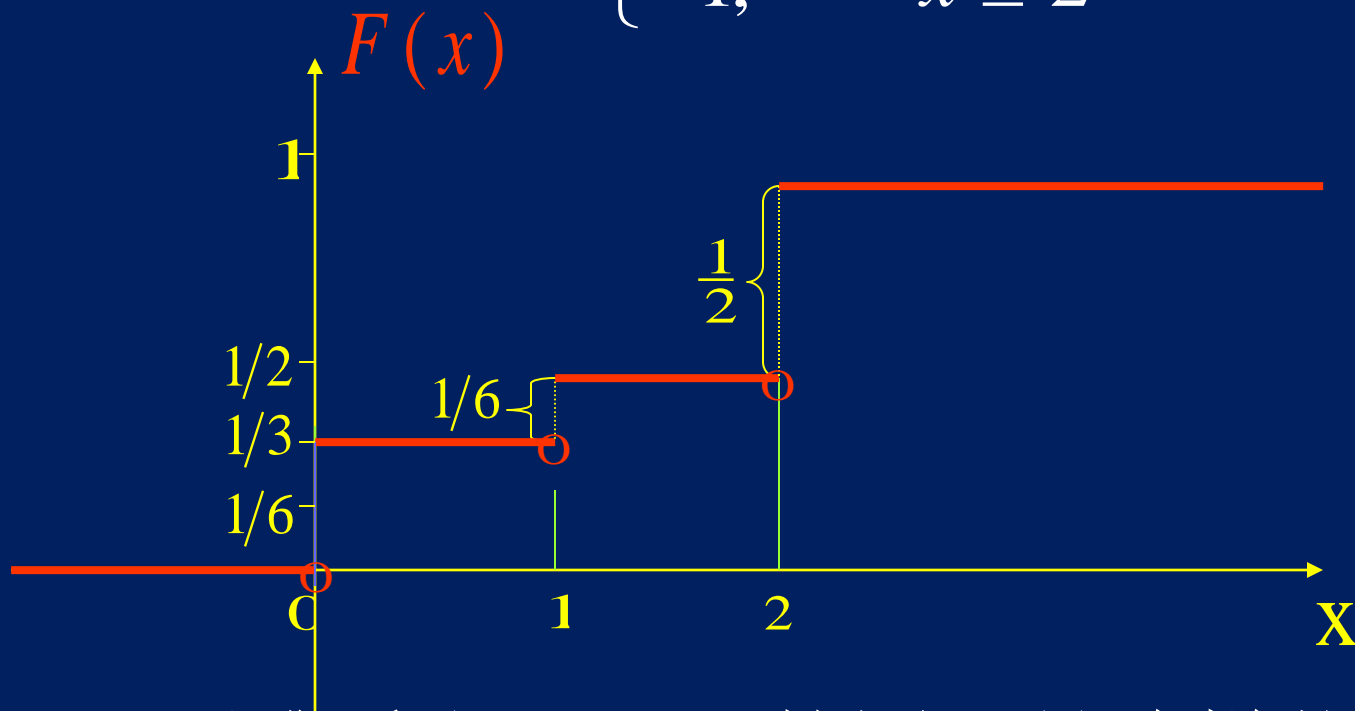
故

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{3}, & 0 \leq x < 1 \\ \frac{1}{2}, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

下面我们从图形上来看一下.



$$F(x) = \begin{cases} 0, & x < 0 \\ 1/3, & 0 \leq x < 1 \\ 1/2, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$



不难看出, $F(x)$ 的图形是阶梯状的图形, 在 $x=0, 1, 2$ 处有跳跃, 其跃度分别等于 $P(X=0), P(X=1), P(X=2)$.

分布律与分布函数的关系图示如下

分布律

$$P(X = x_k) = p_k$$

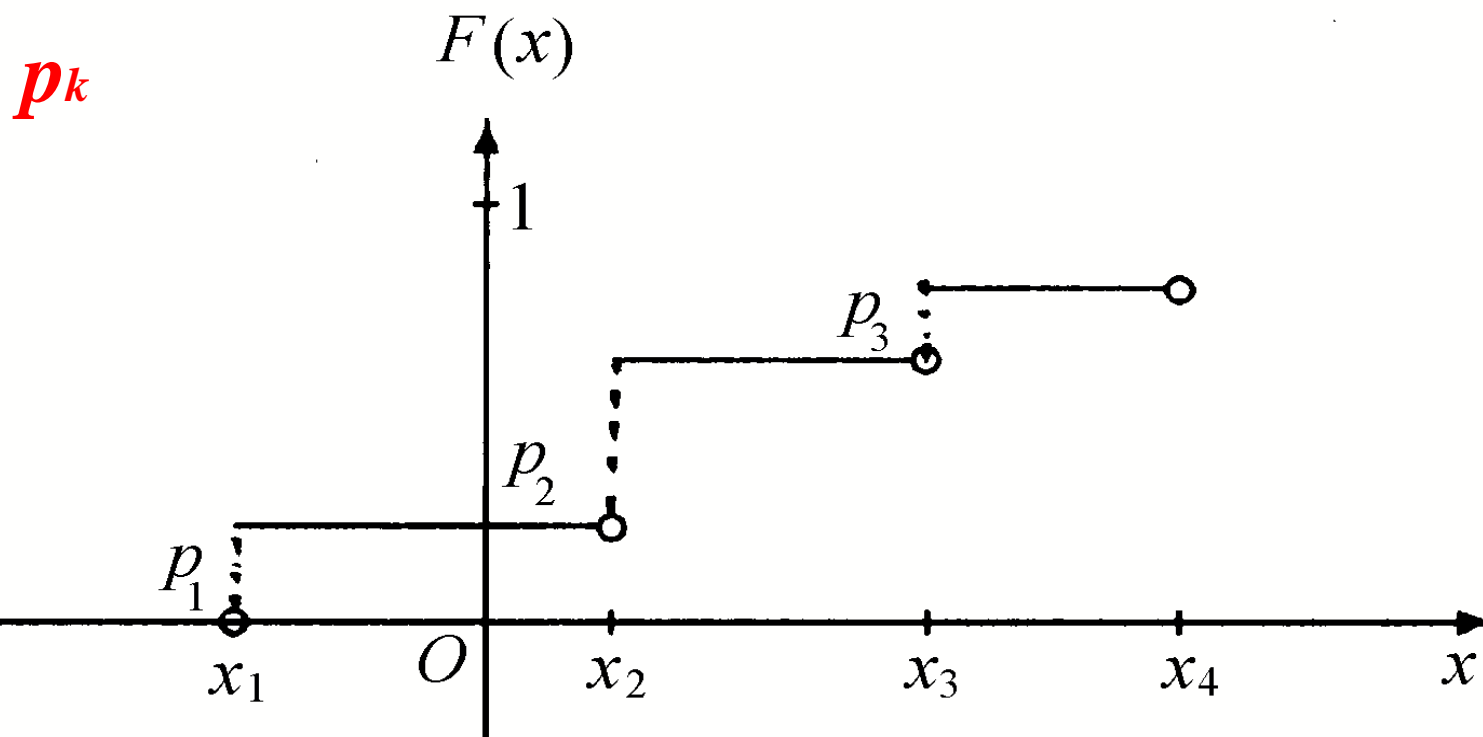


分布函数

$$F(x) = P(X \leq x) = \sum_{x_k \leq x} p_k$$



分布函数 $F(x)$ 的图形是阶梯形、右连续的曲线；在 $X = x_k$ ， $k = 1, 2, 3, \dots$ 处，有跳跃，其跳跃值恰好等于 $P(X = x_k) =$



随机变量分布函数 $F(x)$ 的性质

(1) 单调性 若 $x_1 < x_2$, 则 $F(x_1) \leq F(x_2)$

(2) 规范性 对任意的实数 x , 均有

$$\lim_{x \rightarrow -\infty} F(x) = 0$$
$$\lim_{x \rightarrow +\infty} F(x) = 1$$

(3) 右连续性 对任意的实数 x_0 , 有

$$\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$$



例3 设

$$F(x) = \begin{cases} \sin x, & 0 \leq x \leq \pi \\ 0, & \text{其它} \end{cases}$$

试说明 $F(x)$ 能否是某个随机变量的分布函数

解：注意到函数 $F(x)$ 在 $[\pi/2, \pi]$ 上下降
不满足性质(1)，故 $F(x)$ 不能是分布函数

或者 $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 0$

不满足性质(2)，可见 $F(x)$ 也不能是
分布函数.



例4 设随机变量 X 的分布函数为

$$F(x) = A + B \arctg x, \quad -\infty < x < +\infty$$

试求：

(1) 常数 A 与 B

(2) $P(-1 < X^3 \leq 8)$



§ 4 连续型随机变量

连续型随机变量及其概率密度

设随机变量 X 的分布函数为 $F(x)$ ，如果存在一个非负可积函数 $f(x)$ ，使对任意的实数 x ，均有

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

则称 X 是连续型随机变量，称 $f(x)$ 是 X 的概率密度或密度函数，简称密度。

连续型随机变量 X 的分布函数 $F(x)$ 和密度函数 $f(x)$ 统称为 X 的概率分布，简称 X 的分布。



概率密度的意义

若 x 是 $f(x)$ 的连续点, 则

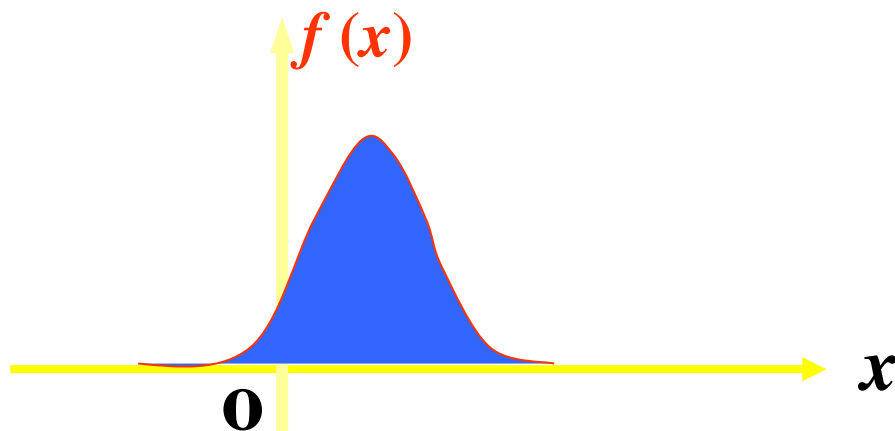
$$\lim_{\Delta x \rightarrow 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x}$$

$$= \lim_{\Delta x \rightarrow 0} \frac{\int_x^{x+\Delta x} f(t) dt}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \theta \Delta x) \Delta x}{\Delta x}$$

$$= f(x)$$

$$f(x) \approx \frac{P(x < X \leq x + \Delta x)}{\Delta x}$$

f(x)反映了 X 在 x 附近取值的概率。



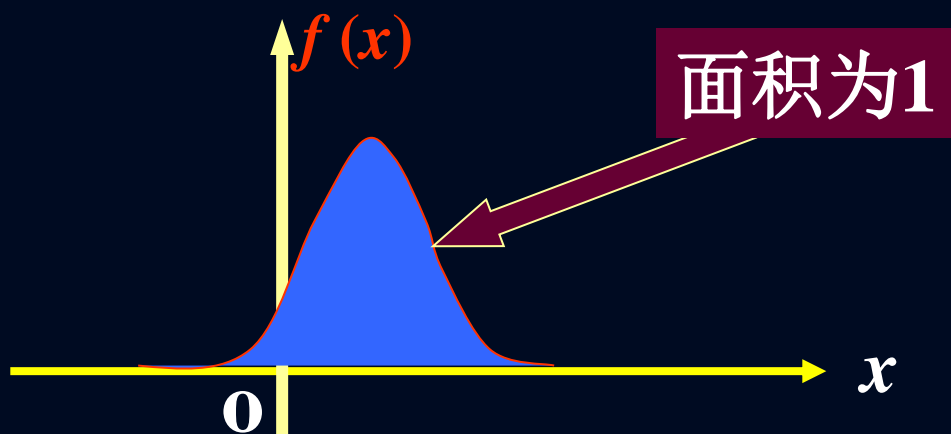
要注意的是，密度函数 $f(x)$ 在某点处 a 的高度，并不反映 X 取值的概率. 但是，这个高度越大，则 X 取 a 附近的值的概率就越大. 也可以说，在某点密度曲线的高度反映了概率在该点的密集程度.

概率密度函数的性质

(1) $f(x) \geq 0$

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

这两条性质是判定一个函数 $f(x)$ 是否为某 $r.v X$ 的概率密度函数的充要条件.



(3)对于连续型随机变量 X

$$P\{X \in G\} = \int_G f(x)dx$$

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

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

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



需要指出的是：

连续型*r.v*取任一指定值的概率为0.

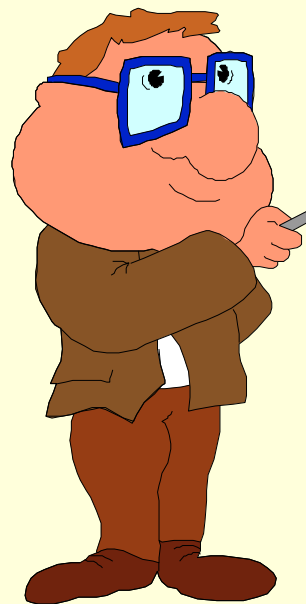
即： $P(X = a) = 0$, a 为任一指定值

1) 对连续型 *r.v* X ,有

$$\begin{aligned} P(a \leq X \leq b) &= P(a < X \leq b) \\ &= P(a \leq X < b) \\ &= P(a < X < b) \end{aligned}$$



需要指出的是：



连续型r.v取任一指定值的概率为0.

即： $P(X = a) = 0$, a 为任一指定值

2) 由 $P(A)=0$, 不能推出 $A=\phi$

真的
吗



例1 某汽车加油站每周补充汽油一次，已知此加油站每周销售量 X （单位： kL ）是以

$$f(x) = \begin{cases} C(1-x^2), & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

为概率密度的随机变量，其中 C 为待定常数。如果要求在一周内加油站的油售完的概率不得大于0.01，那么此加油站储油库的容油体积应至少为多少升。



$$f(x) = \begin{cases} C(1-x^2), & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

解： 由密度函数的性质得

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^1 C(1-x^2) dx = 1$$
$$C = 3/2$$

设此加油站储油库的容油体积为 l 千升($0 < l < 1$)

l 应满足 $P(X \geq l) \leq 0.01$

又有 $P(X \geq l) = \int_l^{\infty} f(x) dx = \int_l^1 \frac{3}{2}(1-x^2) dx = \frac{3}{2} \left[1-l - \frac{1}{3}(1-l)^3 \right]$

$$l \geq 0.92$$



几种常见的分布



均匀分布



指数分布

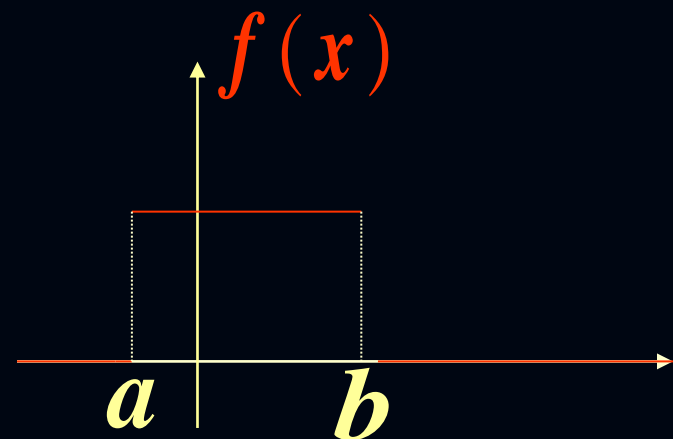


正态分布

均匀分布 (Uniform)

若随机变量 X 的概率密度为

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{其他} \end{cases}$$



则称 X 在 $[a, b]$ 上服从均匀分布, 记为

$$X \sim U[a, b]$$

一个性质

$$\forall (c, c+l) \subset (a, b)$$

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

若 X 取值在区间 $[a, b]$ 上, 并且它在 $[a, b]$ 中任意小区间内取值的概率与这个小区间的长度成正比. 则 $X \sim U[a, b]$



应用场合

均匀分布常见于下列情形：

如在数值计算中，由于四舍五入，小数点后某一位小数引入的误差；

公交线路路上两辆公共汽车前后通过某汽车车站的时间，即乘客的候车时间等。

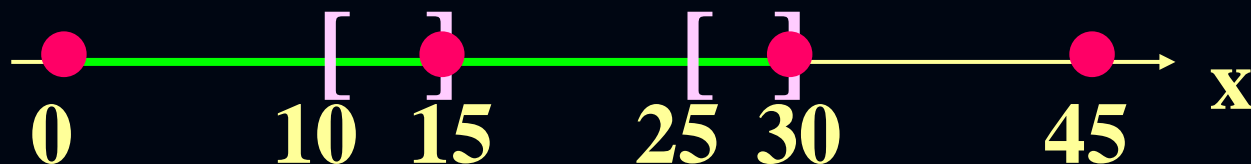
例2 某公共汽车站从上午7时起，每15分钟来一班车，即 7:00, 7:15, 7:30, 7:45 等时刻有汽车到达此站，如果乘客到达此站时间 X 是7:00 到 7:30 之间的均匀随机变量，试求他候车时间少于5 分钟的概率.

解： 以7:00为起点0，以分为单位

依题意， $X \sim U(0, 30)$ $f(x) = \begin{cases} \frac{1}{30}, & 0 < x < 30 \\ 0, & \text{其它} \end{cases}$

$$P\{10 < X < 15\} + P\{25 < X < 30\}$$

$$= \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = \frac{1}{3}$$



一般地，设D是轴上一些不相交的区间之和，若X的概率密度为

$$f(x) = \begin{cases} \frac{1}{D \text{ 的长度}} & , x \in D \\ 0 & , x \notin D \end{cases}$$

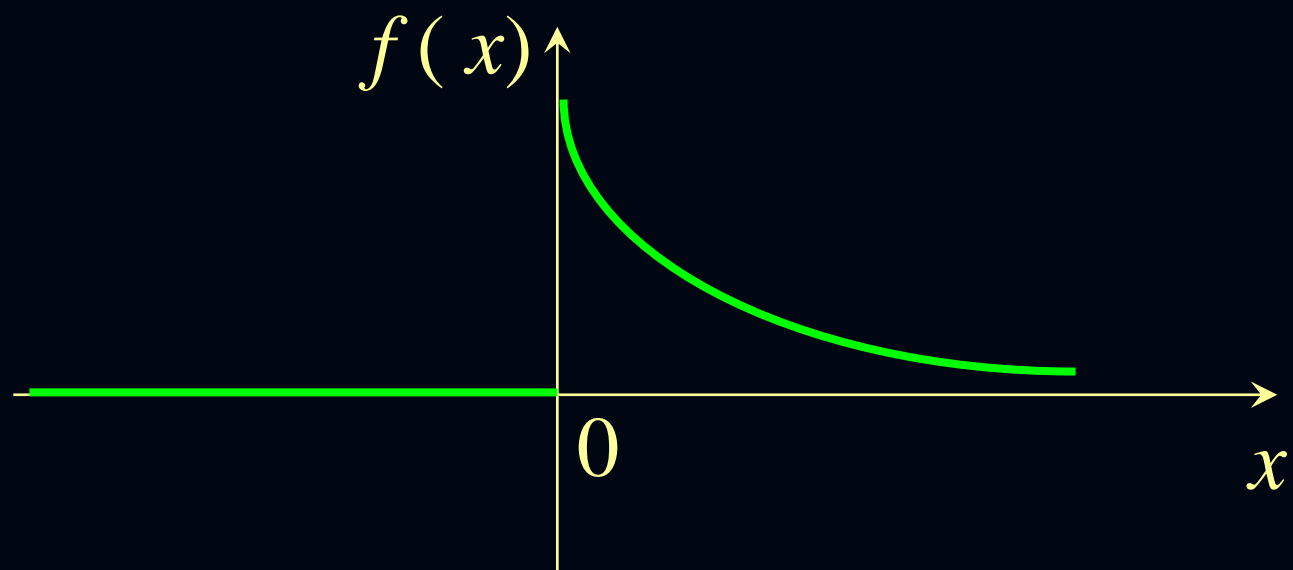
则称X在D上服从均匀分布。

指数分布

若随机变量 X 的概率密度为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

其中常数 $\lambda > 0$ ，则称 X 服从参数为 λ 的指数分布，记为 $X \sim E(\lambda)$ 。



分布函数

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

一个性质

无

记

忆

性

$$\forall \quad s > 0, \quad t > 0$$

$$P(X > s + t \mid X > s) = P(X > t)$$

$$P\{X > s+t \mid X > s\} = \frac{P\{X > s+t, X > s\}}{P\{X > s\}}$$

$$= \frac{P\{X > s+t\}}{P\{X > s\}}$$

$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - F(t) = P\{X > t\}$$

如果 X 表示某仪器的工作寿命，无后效性的解释是：当仪器工作了 s 小时后再能继续工作 t 小时的概率等于该仪器刚开始就能工作 t 小时的概率。

应用场合

用指数分布描述的实例有：

随机服务系统中的服务时间

电话问题中的通话时间

无线电元件的寿命

动物的寿命

} 指数分布常作为各种
“寿命”分布的近似

例 3

设打一次电话所用的时间 X (单位: 分钟) 是以 $\lambda=1/10$ 为参数的指数随机变量. 如果某人刚好在你前面走进公用电话间, 求你需等待10分钟到20分钟之间的概率.

解:

X 的密度函数为

$$f(x) = \begin{cases} \frac{1}{10} e^{-\frac{x}{10}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$



$$f(x) = \begin{cases} \frac{1}{10} e^{-\frac{x}{10}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$P\{10 \leq X \leq 20\}$$

$$= \int_{10}^{20} f(x) dx = \int_{10}^{20} \frac{1}{10} e^{-\frac{x}{10}} dx = -\frac{1}{10} e^{-\frac{x}{10}} \Big|_{10}^{20}$$

$$= (e^{-1} - e^{-2}) / 10 = \mathbf{0.02325}$$



正态分布

1 定义

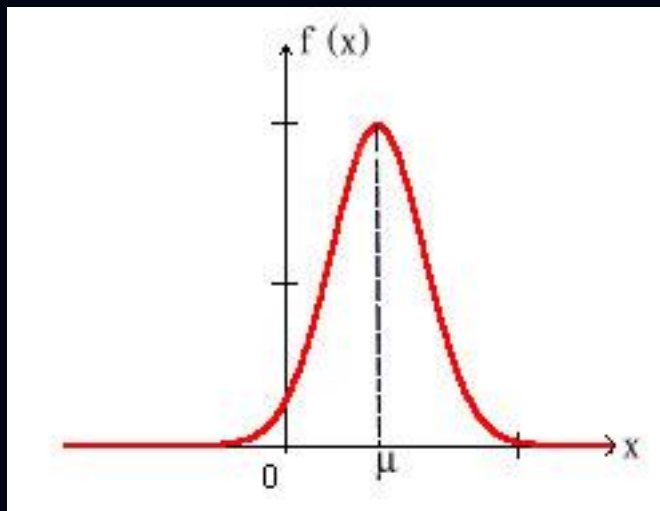
若r.v X 的概率密度为

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

其中 μ 和 σ^2 都是常数, μ 任意, $\sigma > 0$,
则称 X 服从参数为 μ 和 σ^2 的正态分布.

记作 $X \sim N(\mu, \sigma^2)$

2 密度函数 $f(x)$ 的性质

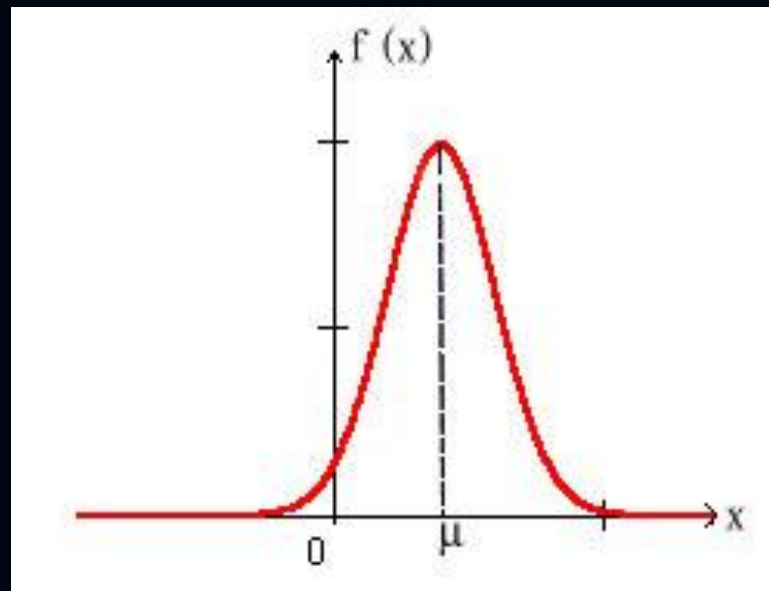


- a. 正态分布的密度曲线是一条关于 μ 对称的钟形曲线. $f(\mu+c)=f(\mu-c)$

特点是“两头小，中间大，左右对称”。

在 $x=\mu$ 处达到最大值: $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$

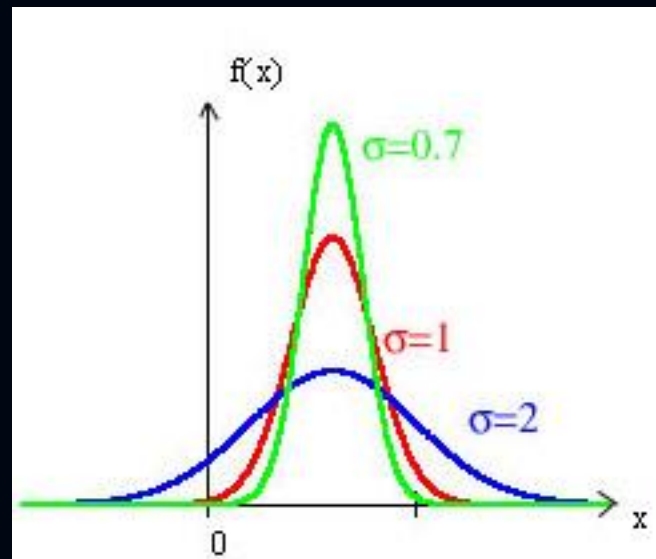
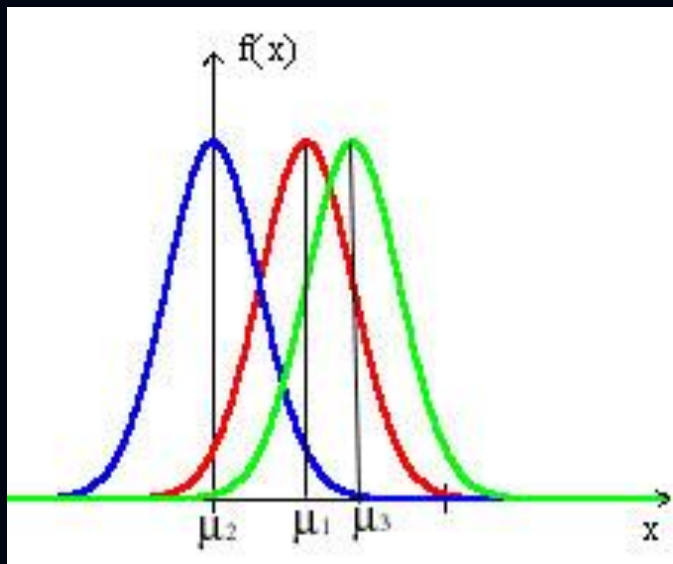
当 $x \rightarrow \pm\infty$ 时, $f(x) \rightarrow 0$,



这说明曲线 $f(x)$ 向左右伸展时, 越来越贴近 x 轴。即 $f(x)$ 以 x 轴为渐近线。

$$x = \mu \pm \sigma$$

为 $f(x)$ 的两个拐点的横坐标。



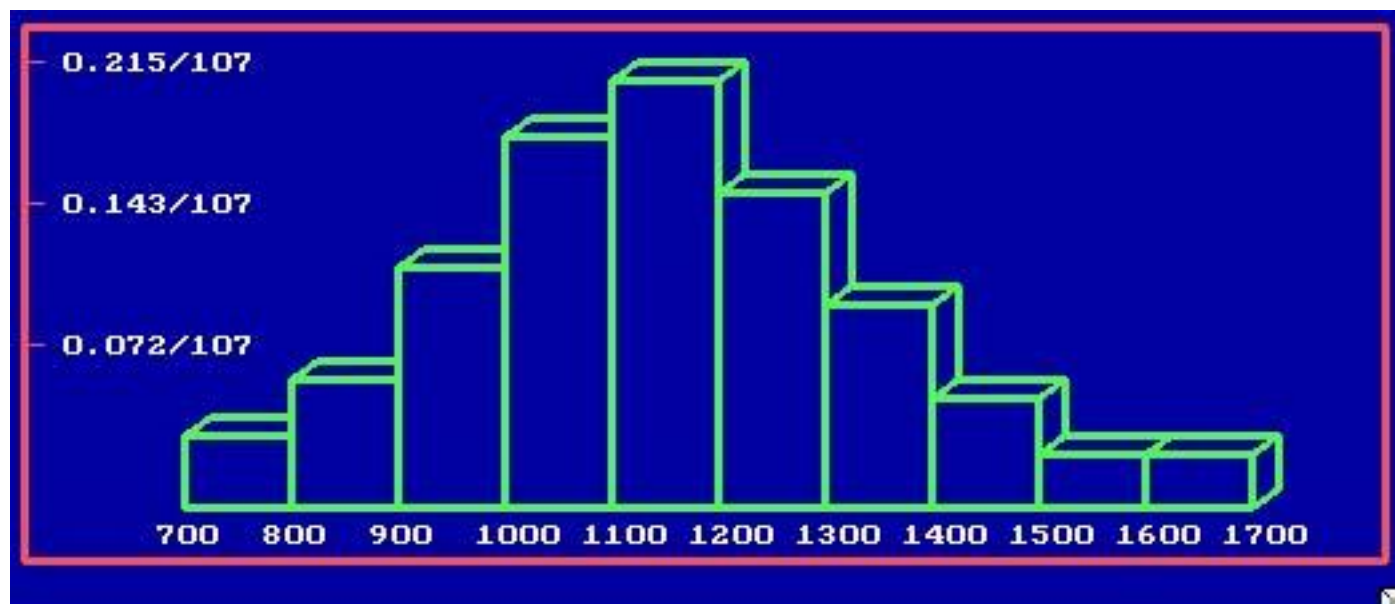
b. μ 决定了图形的中心位置， σ 决定了图形中峰的陡峭程度.

3 应用场合

年降雨量、同龄人身高、在正常条件下各种产品的质量指标——如零件的尺寸；纤维的强度和张力、农作物的产量，小麦的穗长、株高、测量误差、射击目标的水平或垂直偏差、信号噪声等等，都服从或近似服从正态分布。

年降雨量问题

这是用上海1999年年降雨量的数据画出的频率直方图



从直方图可以初步看出，年降雨量近似服从正态分布

身高问题

此外，人的身高高低不等，但中等身材的占大多数，特高和特矮的只是少数，而且较高和较矮的人数大致相近，这从一个方面反映了服从正态分布的随机变量的特点



说 明

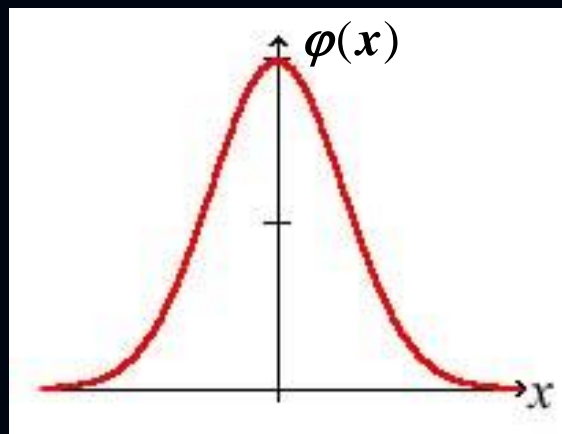
- (1). 正态分布是自然界及工程技术中最常见的分布之一，大量的随机现象都是服从或近似服从正态分布的。可以证明，如果一个随机指标受到诸多因素的影响，但其中任何一个因素都不起决定性作用，则该随机指标一定服从或近似服从正态分布
- (2). 正态分布有许多良好的性质，这些性质是其它许多分布所不具备的
- (3). 正态分布可以作为许多分布的近似分布

4 标准正态分布

$\mu = 0, \sigma = 1$ 的正态分布称为标准正态分布.

其密度函数常用 $\varphi(x)$ 表示:

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

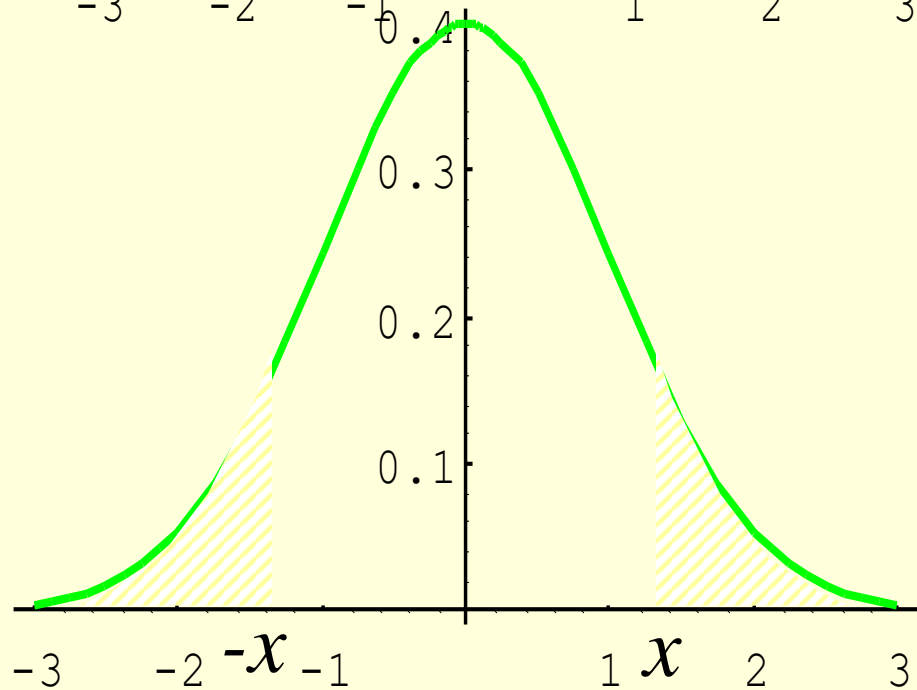
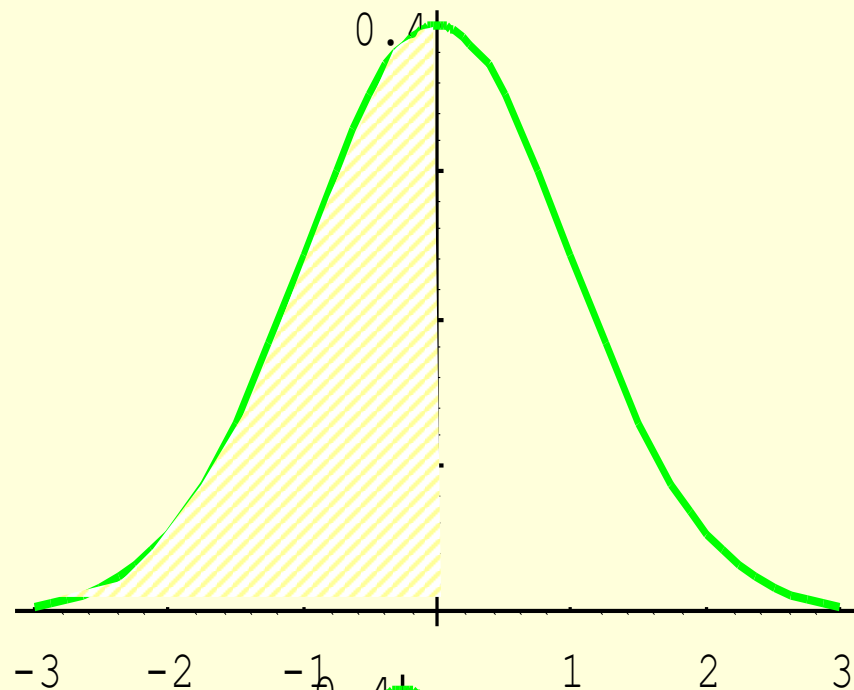


$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

注意：

$$\Phi(0)=0.5$$

$$\Phi(-x)=1-\Phi(x)$$



5 正态分布的计算

对一般的正态分布： $X \sim N(\mu, \sigma^2)$

$$\text{若 } X \sim N(\mu, \sigma^2), \quad \frac{X - \mu}{\sigma} \sim N(0, 1)$$

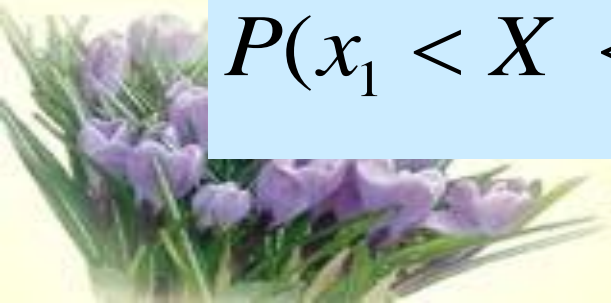


对任意的实数 x_1, x_2 ($x_1 < x_2$), 有

$$P(X \leq x_1) = \Phi\left(\frac{x_1 - \mu}{\sigma}\right)$$

$$P(X > x_1) = 1 - \Phi\left(\frac{x_1 - \mu}{\sigma}\right)$$

$$P(x_1 < X < x_2) = \Phi\left(\frac{x_2 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right)$$



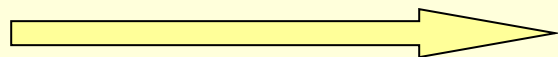
例4 设 $X \sim N(1,4)$, 求 $P(0 \leq X \leq 1.6)$

解
$$P(0 \leq X \leq 1.6) = P\left(\frac{0-1}{2} \leq \frac{X-1}{2} \leq \frac{1.6-1}{2}\right)$$

$$= \Phi(0.3) - \Phi(-0.5)$$

$$= \Phi(0.3) - [1 - \Phi(0.5)]$$

附录



$$= 0.6179 - [1 - 0.6915]$$

$$= 0.3094$$



例5 公共汽车车门的高度是按男子与车门顶头碰头机会在0.01以下来设计的. 设男子身高 $X \sim N(170, 6^2)$, 问车门高度应如何确定?

解: 设车门高度为 h cm, 按设计要求

$$P(X \geq h) \leq 0.01$$

或 $P(X < h) \geq 0.99,$



下面我们来求满足上式的最小的 h .



求满足 $P(X < h) \geq 0.99$ 的最小的 h .

因为 $X \sim N(170, 6^2)$, $\frac{X - 170}{6} \sim N(0, 1)$

故 $P(X < h) = \Phi\left(\frac{h - 170}{6}\right) \geq 0.99$

查表得 $\Phi(2.33) = 0.9901 > 0.99$

所以 $\frac{h - 170}{6} = 2.33$,

即 $h = 170 + 13.98 \approx 184$

设计车门高度为
184厘米时，可使
男子与车门碰头
机会不超过0.01.



例6 一桥长60m, 以桥的中点为原点, 沿着桥的方向引入坐标轴. 一架飞机沿着坐标轴俯冲投弹轰炸此桥, 假定弹着点的坐标 $X \sim N(0, 100^2)$.

(1) 求投掷一枚炸弹, 命中此桥的概率;

(1) 解:

$$P(|X| \leq 30)$$

$$= P(-30 \leq X \leq 30) = P\left(-0.3 \leq \frac{X}{100} \leq 0.3\right)$$

$$= 2\Phi(0.3) - 1 = 0.2358$$



一桥长60m, 以桥的中点为原点, 沿着桥的方向引入坐标轴. 一架飞机沿着坐标轴俯冲投弹轰炸此桥, 假定弹着点的坐标 $X \sim N(0, 100^2)$.

(1) 求投掷一枚炸弹, 命中此桥的概率;

(2) 问独立投掷多少枚炸弹, 才能使至少有一枚弹命中此桥的概率大于0.9.

(2) 分析: 设独立投掷 k 枚炸弹

$A = \{\text{至少有一枚弹命中此桥}\}$

k 需满足 $P(A) > 0.9$

或 $P(\bar{A}) \leq 0.1$



一桥长60m, 以桥的中点为原点, 沿着桥的方向引入坐标轴. 一架飞机沿着坐标轴俯冲投弹轰炸此桥, 假定弹着点的坐标 $X \sim N(0, 100^2)$.

(1) 求投掷一枚炸弹, 命中此桥的概率;

(2) 问独立投掷多少枚炸弹, 才能使至少有一枚弹命中此桥的概率大于0.9.

(2) 解:

设独立投掷 k 枚炸弹,

A : 至少一枚炸弹命中此桥

k 需满足 $P(A) > 0.9$

即 k 满足 $P(\bar{A}) \leq 0.1$

$$P(\bar{A}) = 0.7642^k < 0.1$$

$$k > 8.5622$$



