第 2 章-随机变量及其分布

李孟棠, 朱彬

中山大学-智能工程学院

{limt29,zhub26}@mail.sysu.edu.cn

December 18, 2020

章节概述

知识点:

1. 随机变量
2. 离散型随机变量及其分布律
3. 随机变量的分布函数
4. 连续型随机变量及其概率密度
5. 随机变量的函数的分布

2.1 随机变量

回顾:
有些随机试验的结果可以用数字来表示，如某同学的考试排名 $S_1 = \{1, 2, 3, ...\}$，或今天的气温 $S_1 = \{x | T_{min} \leq x \leq T_{max}\}$; 有些随机试验的结果不可以用数字来表示，如连续抛掷一枚硬币两次 $S_3 = \{HH, HT, TH, TT\}$.

问题:
非数字的结果，难以描述和研究.

解决方法:
引入一个法则，将随机试验的每一个结果 e，与实数 x 对应起来.

例 (1)

将一枚硬币连续抛掷三次，观察出现正面和反面的情况，样本空间为

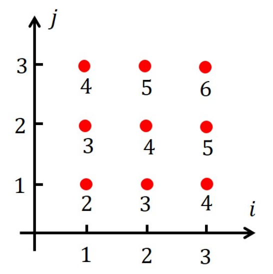
$$S = HHH, HHT, HTH, THH, HTT, THT, TTH, TTT$$

以 X 计三次抛掷得到正面 H 的综述，那么对于样本空间 $S = \{e\}$ 中的每一个样本点 e,X 都有一个数与之对应.X 是定义在样本空间 S 上的一个**实值单值**函数，定义域为 S，值域为 $\{0, 1, 2, 3\}$:

$$X = X(e) = \begin{cases} 3 & , e = HHH, \\ 2 & , e = HHT, HTH, THH \\ 1 & , e = HTT, THT, TTH \\ 0 & , e = TTT \end{cases}$$

例 (2)

一袋中装有编号分别为 1,2,3 的 3 只球, 现从中任取一只球, 放回, 再取一只, 记录球的号码. 样本空间为 $S=\{e\}=\{(i,j)|i,j=1,2,3\}$,$i$,$j$ 分别为第 1 次第二次取到的球的号码. 以 X 记两球号码之和.



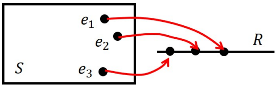
对于试验的每一个结果 $e=(i,j)\in S$,X 都有一个指定的值 $i+j$ 与之对应 (如图所示).X 是定义在样本空间 S 上的一个实值单值函数, 定义域为 S, 值域为 $\{2,3,4,5,6\}$:

$$X=X(e)=X((i,j))=i+j, i,j=1,2,3$$

定义

设随机试验的样本空间为 $S=\{e\}$.$X=X(e)$ 是定义在样本空间 S 上的实值单值函数. 称 $X=X(e)$ 为随机变量.

下图给出了样本点 e 与实数 $X=X(e)$ 对应的示意图:



有一些随机试验, 结果本身就是一个实数, 即有随机变量 $X=X(e)=e$. 例如:Y 记某车间一天的缺勤人数,W 记某地第一季度的降雨量,Z 记某工厂一天的耗电量.

注意: 通常, 我们使用大写字母 $X,Y,Z,\ldots$ 表示随机变量, 小写字母 $x,y,z,\ldots$ 表示实数.

2.2 离散型随机变量及其分布

0-1 分布
伯努利试验、二项分布
泊松分布、泊松定理

定义

随机变量的全部可能取值若为有限个，或可列无限多个，则称这种随机变量为离散型随机变量.

例如，抛掷一枚硬币三次，其中正面向上的次数只可能为 0,1,2,3 四个值，因此它是一个离散型随机变量.

某城市的 120 急救台一昼夜收到的呼唤次数可能是 0,1,2,3,...., 虽然是无限个，但是可列出，因此它也是一个离散型随机变量.

某地的某季度的降雨量，它所可能取得值充满某个区间，无法一一列举出来，因此它是一个非离散型 (实际为连续性) 随机变量.

容易得知，要掌握一个离散型随机变量 X 的统计规律，必须且只需要知道 X 的所有可能取值以及取每一个可能值的概率即可.
设离散型随机变量 X 所有可能取值 $x_k(k=1,2,3,...)$,X 取各个可能值的概率 (即事件 $\{X=x_k\}$) 为

$$P\{X=x_k\}=p_k, k=1,2,3,... \tag{1}$$

由概率的定义可知,p_k 满足如下两个条件:

1. $p_k \geq 0, k=1,2,3,...$
2. $\sum_{k=1}^{\infty} p_k = 1$

其中性质②:
因为 $\{X=x_1\}\cup\{X=x_2\}\cup ...$ 是必然事件
并且 $\{X=x_i\}\cap\{X=x_j\}=\emptyset, i\neq j$
所以 $1=P[\bigcup_{k=1}^{\infty}\{X=x_k\}]=\sum_{k=1}^{\infty}P\{X=x_k\}$，即 $\sum_{k=1}^{\infty}p_k=1$

我们称公式 (1) 为离散型随机变量 X 的**分布律**, 可用表格形式表示:

X	x_1	x_2	$\cdots$	x_n	$\cdots$
p_k	p_1	p_2	$\cdots$	p_n	$\cdots$

例 (1)

设一汽车在开往目的地的道路上需经过四组信号灯, 每组信号灯以 0.5 的概率允许或禁止汽车通行. 以 X 表示汽车首次停下时, 它已经通过的信号灯的组数 (设各组信号灯的工作是相互独立的), 求 X 的分布律.

解 以 p 表示魅族信号灯禁止汽车通过的概率, 易知 X 的分布律

X	0	1	2	3	4
p_k	p	$(1-p)p$	$(1-p)^2p$	$(1-p)^3p$	$(1-p)^4$
p_k	0.5	0.25	0.125	0.0625	0.0625

或写成

$$P\{X=k\}=(1-p)^k p^{4-k}, k=0,1,2,3,4$$

一、0-1 分布

定义

设随机变量 X 只可能取 0 与 1 两个值，它的分布律为

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

则称 X 服从以 p 为参数的0-1 分布或两点分布，其分布律为

X	0	1
p_k	$1-p$	p

对于一个随机试验，如果它的样本空间只包含两个元素例如 $S=\{e_1,e_2\}$，我们总能在 S 上定义一个服从 0-1 分布的随机变量

$$X=X(e)=\begin{cases} 0 & ,\text{ if } e=e_1 \\ 1 & ,\text{ if } e=e_2 \end{cases}$$

二、伯努利试验、二项分布

定义

设试验 E 只有两个可能结果:A 和 $\bar{A}$, 则称 E 为**伯努利 (Bernoulli) 试验**. 设试验中 $P(A)=p(0\le p\le 1)$, 此时 $P(\bar{A})=1-p$. 将 E**独立重复地**进行 n 次, 则称这一串重复的独立试验为**n 重伯努利试验**.

注意: 这里的" **重复**"指每次实验中都有 $P(A)=p$ 保持不变;" **独立**"指各此试验结果互不影响. 若以 C_i 记第 i 次试验结果 ($C_i=A or \bar{A}$), 则

$$P(C_1C_2...C_n)=P(C_1)P(C_2)...P(C_n) \tag{2}$$

例如,①E 为抛一枚硬币观察得到正面或反面.A 表示得正面, 这是一个伯努利试验. 如抛 n 次, 则为 n 重伯努利试验.
②抛一枚骰子,A 表示得到 1 点,$\bar{A}$ 表示得到非 1 点. 将骰子抛 n 次, 则为 n 重伯努利试验.
③一个袋中装有 a 个白球 b 个黑球, 试验 E 是在袋中任取一个球, 观察颜色.A 表示取得白球,$P(A)=a/(a+b)$. 若连续取球 n 次, 作放回抽样, 则为 n 重伯努利试验; 若作不放回抽样, 则各次试验不在互相独立, 因此不是 n 重伯努利试验.

以随机变量 X 表示 n 重伯努利试验中事件 A 发生的次数,X 的所有取值为 $0, 1, 2, ..., n$, 事件 A 在指定的 $k(0 \le k \le n)$ 次, 则在其他 $n-k$ 中不发生, 故总概率为 (记 $q = 1 - p$)

$$P\{X = k\} = C_n^k p^k q^{n-k}, k = 0, 1, 2, ..., n \tag{3}$$

显然,

$$P\{X = k\} \ge 0, k = 0, 1, 2, ..., n$$

$$\sum_{k=0}^{n} P\{X = k\} = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^n = 1$$

即 $P\{X = k\}$ 满足前述的 2 个性质. 其中,$C_n^k p^k q^{n-k}$ 正好是二项式 $(p+q)^n$ 的展开式中 p^k 的那一项, 因此我们称随机变量 X 服从参数为 n, p 的二项分布, 记为 $X \sim b(n, p)$.
特别地, 当 $n = 1$ 时二项分布 (3) 退化为0-1 分布.

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

例 (2)

按规定，某种型号电子元件的使用寿命超过 1500 小时的为一级品. 已知某一大批产品的一级品率为 0.2，现在从中随机地抽查 20 只. 问 20 只原件中恰有 k 只为一级品的概率是多少?

解 这是不放回抽样，但由于这批元件总数很大，且抽查的元件数量相对总数来说很小，因此可以当做放回抽样 (即可视为独立重复试验). 该例相当于 20 重伯努利试验，以 X 记 20 只元件中一级品的只数，那么 $X \sim b(20, 0.2)$，由公式 (3) 得

$$P\{X=k\} = C_{20}^{k} 0.2^{k} 0.8^{20-k}, k = 0, 1, 2, ..., 20$$

计算结果列表如下

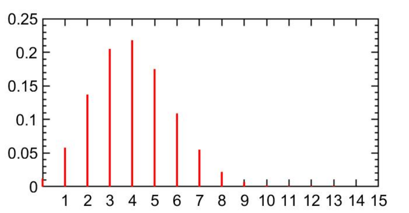
$P\{X=0\}=0.012$	$P\{X=1\}=0.058$	$P\{X=2\}=0.137$
$P\{X=3\}=0.205$	$P\{X=4\}=0.218$	$P\{X=5\}=0.175$
$P\{X=6\}=0.109$	$P\{X=7\}=0.055$	$P\{X=8\}=0.022$
$P\{X=9\}=0.007$	$P\{X=10\}=0.002$	$P\{X=k\}<0.001, for\ k\geq 11$

例 (2 续)

$$P\{X=k\}=C_{20}^{k}0.2^{k}0.8^{20-k}, k=0,1,2,...,20$$

$P\{X=0\}=0.012$	$P\{X=1\}=0.058$	$P\{X=2\}=0.137$
$P\{X=3\}=0.205$	$P\{X=4\}=0.218$	$P\{X=5\}=0.175$
$P\{X=6\}=0.109$	$P\{X=7\}=0.055$	$P\{X=8\}=0.022$
$P\{X=9\}=0.007$	$P\{X=10\}=0.002$	$P\{X=k\}<0.001, \textit{for } k\geq 11$

由图可知, 当 k 增加时, 概率 $P\{X=k\}$ 首先随之增加, 达到最大值, 然后单调减小. 对于二项分布 $b(n,p)$ 都具有这种性质.



例 (3)

某人进行射击, 设每次射击命中率为 0.02, 独立射击 400 次, 求至少击中两次的概率.

解 将每一次射击看出是一次试验. 击中次数为 X, 则 $X \sim b(400, 0.2)$,X 的分布律为

$$P\{X = k\} = C_{400}^{k} 0.02^{k} 0.98^{20-k}, k = 0, 1, 2, ..., 400$$

所求概率为

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

$$= 1 - 0.98^{400} - 400 \cdot 0.02 \cdot 0.98^{399} = 0.9972 \approx 1$$

例 (4)

设有 80 台同类型设备，各台工作是互相独立的，发生故障的概率都为 0.01，且一台设备的故障能由一个人处理．考虑两种配备维修工人的方法，其一是由 4 人维护，每人负责 20 台；其二是由 3 人共同维护 80 台．试比较这两种方法在设备发生故障时不能及时维修的概率．

解

方法一：以 X 记" 第 1 人维护的 20 台中同一时刻发生故障的台数"，以 $A_i(i=1,2,3,4)$ 表示事件" 第 i 人维护的 20 台中发生故障不能及时维修"，则知 80 台中发生故障而不能及时维修的概率为

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) \geq P(A_1) = P\{X \geq 2\}$$

而 $X \sim b(20, 0.01)$，则有

$$P\{X \geq 2\} = 1 - \sum_{k=0}^{1} P\{X = k\} = 1 - \sum_{k=0}^{1} C_{20}^{k} 0.01^k 0.99^{20-k} = 0.0169$$

即有 $P(A_1 \cup A_2 \cup A_3 \cup A_4) \geq 0.0169$

例 (4 续)

而 $X \sim b(20, 0.01)$, 则有

$$P\{X \geq 2\} = 1 - \sum_{k=0}^{1} P\{X = k\} = 1 - \sum_{k=0}^{1} C_{20}^{k} 0.01^{k} 0.99^{20-k} = 0.0169$$

即有 $P(A_1 \cup A_2 \cup A_3 \cup A_4) \geq 0.0169$

方法二: 以 Y 记 80 台中同一时刻发生故障的台数. 此时 $Y \sim b(80, 0.01)$, 故 80 台中发生故障而不能及时维修的概率为

$$P\{Y \geq 4\} = 1 - \sum_{k=0}^{3} C_{80}^{k} 0.01^{k} 0.99^{80-k} = 0.0087$$

总结: 方法二中虽然每人平均的任务重了, 但工作效率不降反升.

定义

设随机变量 X 所有可能取的值为 0,1,2,..., 而取各个值的概率为

$$P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!}, k=0,1,2,\ldots$$

其中 $\lambda>0$ 是常数, 称 X 服从参数为 λ 的泊松分布 (Poisson), 记作 $X\sim\pi(\lambda)$.

易知,$P\{X=k\}\geq 0$, 且有

$$\sum_{k=0}^{\infty}P\{X=k\}=\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1$$

即 $P\{X=k\}$ 满足前述的 2 个性质.

Proposition (泊松定理)

设常数 $\lambda > 0$, 任意正整数 n, 设 $np_n = \lambda$, 则对于任一固定的非负整数 k, 有

$$\lim_{n\to\infty} C_n^k p_n^k (1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

证明.

因为 $np_n = \lambda$, 有

$$C_n^k p_n^k (1-p_n)^{n-k} = \frac{n(n-1)...(n-k+1)}{k!}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k}$$

$$= \frac{\lambda^k}{k!}[1\cdot(1-\frac{1}{n})...(1-\frac{k-1}{n})](1-\frac{\lambda}{n})^n(1-\frac{\lambda}{n})^{-k}$$

对于任意固定的 k, 当 $n\to\infty$ 时

$$1\cdot(1-\frac{1}{n})...(1-\frac{k-1}{n}) \to 1, (1-\frac{\lambda}{n})^n \to e^{-\lambda}, (1-\frac{\lambda}{n})^{-k} \to 1$$

Proposition (泊松定理 (续))

定理的条件 $np_n = \lambda$(常数) 意味着当 n 很大时 p_n 必定很小, 因此上述定理表明有时有以下近似

$$C_n^k p_n^k (1 - p_n)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!} \tag{4}$$

其中 $\lambda = np$. 也就是说以 n, p 为参数的二项分布的概率值可以由参数 $\lambda = np$ 的泊松分布的概率值近似.

关系图

图: 著名学术大家关系图

例 (5)

计算机硬件公司制造某种特殊型号的微型芯片，次品率 0.1%，各芯片制造相互独立. 求 1000 只产品中至少有 2 只次品的概率. 以 X 记产品中的次品数,$X \sim b(1000, 0.001)$.

解 所求概率为

$$P\{X \geq 2\} = 1 - \sum_{k=0}^{1} P\{X = k\} = 1 - 0.999^{1000} - C_{1000}^{1} 0.999^{999} 0.001$$

$$= 1 - 0.3676954 - 0.3680635 \approx 0.2642411$$

利用公式 (4) 计算,$\lambda = np = 1000 \times 0.001 = 1$,

$$P\{X \geq 2\} = 1 - \sum_{k=0}^{1} P\{X = k\} = 1 - e^{-1} - e^{-1} = 0.2642411$$

显然用公式 (4) 计算更加方便. 一般地，当 $n \geq 20, p \leq 0.05$ 时，使用泊松定理进行近似效果颇佳.

2.3 随机变量的分布函数

动机:
①对于离散型随机变量 X 的取值可以一一列出, 但是对非离散型随机变量 X 无法这样用分布律表示.
②对于离散型随机变量 X 取单个值有概率, 但是对非离散型随机变量 X 取单个值的概率为 0.
③对于离散型随机变量 X 我们关心取单个值的概率, 但是对非离散型随机变量 X 我们关心其落在某个区间的概率.

定义

设 X 是一个随机变量,x 是任意实数, 称下述函数为 X 的分布函数(或累积分布函数)

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

对于任意实数 $x_1, x_2 (x_1 < x_2)$, 有

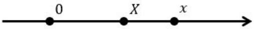
$$P\{x_1 < X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1\} = F(x_2) - F(x_1) \quad (5)$$

分布函数 $F(x)$ 具有以下基本性质:

1. $F(x)$ 是一个非减函数
   有公式 (5) 可知对任意实数
   $x_1, x_2(x_1 < x_2)$ 有

$$F(x_2) - F(x_1) = P\{x_1 < X \leq x_2\} \geq 0$$

2. $0 \leq F(x) \leq 1$, 且



$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0, F(\infty) = \lim_{x \to \infty} F(x) = 1$$

3. $F(x + 0) = F(x)$, 即 $F(x)$ 是右连续.

例 (1)

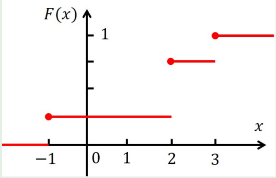
设随机变量 X 的分布律如下，求 X 得分布函数，并求 $P\{X \leq \frac{1}{2}\}$,$P\{\frac{3}{2} < X \leq \frac{5}{2}\}$,$P\{2 \leq X \leq 3\}$

X	-1	2	3
p_k	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

解 $F(x)$ 的值是 $X \leq x$ 的积累概率值，即为小于或等于 x 的那些 x_k 处的概率 p_k 之和

$$F(x) = \begin{cases} 0 & , \text{for } x < -1 \\ P\{X = -1\} = \frac{1}{4} & , \text{for } -1 \leq x < 2 \\ P\{X = -1\} + P\{X = 2\} = \frac{3}{4} & , \text{for } 2 \leq x < 3 \\ 1 & , \text{for } x \geq 3 \end{cases}$$

例 (1 续)



$F(x)$ 的图形如上图所示. 它是一条阶梯形的曲线, 在 $x = -1, 2, 3$ 处有跳跃点, 跳跃值分别为 $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$.

$$P\{X \leq \frac{1}{2}\} = F(\frac{1}{2}) = \frac{1}{4}$$

$$P\{\frac{3}{2} < X \leq \frac{5}{2}\} = F(\frac{5}{2}) - F(\frac{3}{2}) = \frac{1}{2}$$

$$P\{2 \leq X \leq 3\} = F(3) - F(2) + P\{X = 2\} = 1 - \frac{3}{4} + \frac{1}{2} = \frac{3}{4}$$

一般地，设离散型随机变量 X 的分布律为

$$P\{X = x_k\} = p_k, k = 1, 2, \dots$$

由概率的可列可加性得 X 的分布函数为

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

即

$$F(x) = \sum_{x_k \le x} p_k$$

例 (2)

有一个靶子的半径为 2m 的圆盘，设击中靶上任意同心圆盘上的点的概率与该圆盘的面积成正比，并假设每次射击都能中靶. 以 X 表示弹着点与圆心的距离. 试求随机变量 X 的分布函数.

解 ① 若 $x < 0$，则 $\{X \le x\}$ 是不可能事件，于是

$$F(x) = P\{X \le x\} = 0 \text{ , for x<0}$$

例 (2 续)

② 若 $0 \le x \le 2$, 则 $P\{0 \le X \le x\} = kx^2$. 取 $x = 2$, 有 $P\{0 \le X \le 2\} = 4k = 1$, 故 $k = \frac{1}{4}$, 即

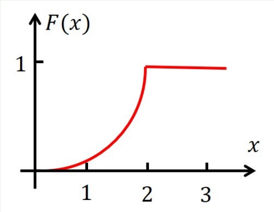
$$F(x) = P\{0 \le X \le x\} = \frac{x^2}{4} \text{ , for } 0 \le x \le 2$$

③ 若 $x \ge 2$, 则 $\{X \le x\}$ 是必然事件, 于是

$$F(x) = P\{X \le x\} = 1 \text{ , for } x \ge 2$$

综上,X 的分布函数为

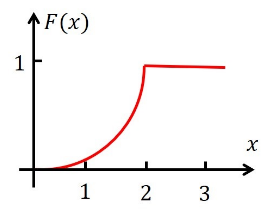
$$F(x) = \begin{cases} 0 & \text{, for } x < 0 \\ \frac{x^2}{4} & \text{, for } 0 \le x < 2 \\ 1 & \text{, for } x \ge 2 \end{cases}$$



例 (2 续)

X 的分布函数为

$$F(x) = \begin{cases} 0 & \text{, for } x < 0 \\ \frac{x^2}{4} & \text{, for } 0 \le x < 2 \\ 1 & \text{, for } x \ge 2 \end{cases}$$



可以观察到, 对于任意 x 可以写成如下形式

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

其中,

$$f(t) = \begin{cases} \frac{t}{2} & \text{, for } 0 < t < 2 \\ 0 & \text{, else} \end{cases}$$

$F(x)$ 是非负函数 $f(t)$ 在区间 $(-\infty, x]$ 上的积分, 这种情况下我们称 X 为连续型随机变量.

2.4 连续型随机变量及其概率密度

均匀分布
指数分布
正态 (高斯) 分布

定义

如上节例 2, 如果对于随机变量 X 的分布函数 $F(x)$, 存在非负可积函数 $f(x)$, 使得对于任意实数 x 有

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{6}$$

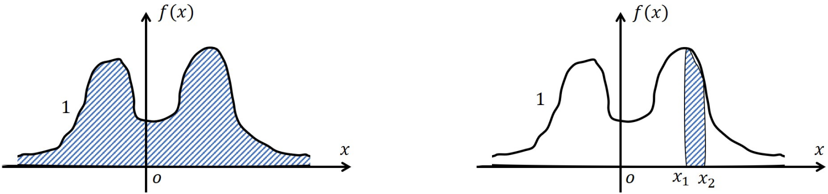
则称 X 为连续型随机变量,$f(x)$ 为 X 的概率密度函数, 简称概率密度.

概率密度 $f(x)$ 具有以下基本性质:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. 对于任意实数 $x_1, x_2 (x_1 \leq x_2)$ 有

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

4. 若 $f(x)$ 在点 x 处连续, 则有 $F'(x) = f(x)$



解读:

1. 由性质②可知介于曲线 $y = f(x)$ 与 Ox 轴间的面积为 1.
2. 由性质③可知 X 落在区间 $(x_1, x_2]$ 的概率 $P\{x_1 < X \le x_2\}$ 等于区间 $(x_1, x_2]$ 上曲线 $y = f(x)$ 之下的面积.
3. 由性质④可知在 $f(x)$ 的连续点 x 处有

$$f(x) = \lim_{\Delta x \to 0^+} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0^+} \frac{P\{x < X \le x + \Delta x\}}{\Delta x} \tag{7}$$

若不计高阶无穷小, 则有 $P\{x < X \le x + \Delta x\} \approx f(x)\Delta x$, 表示 X 落在小区间 $(x, x+\Delta x]$ 上的概率近似等于 $f(x)\Delta x$.

例 (1)

设随机变量 X 具有如下概率密度, 求 (1) 常数 k,(2)X 的分布函数 $F(x)$,(3)$P\{1 < X \leq \frac{7}{2}\}$.

$$f(x) = \begin{cases} kx & , \text{for } 0 \leq x < 3 \\ 2 - \frac{x}{2} & , \text{for } 3 \leq x \leq 4 \\ 0 & , \text{else} \end{cases}$$

解 (1) 由 $\int_{-\infty}^{\infty} f(x)dx = 1$ 得

$$\int_0^3 kxdx + \int_3^4 (2 - \frac{x}{2})dx = 1$$

解得 $k = \frac{1}{6}$, 于是 X 的概率密度为

$$f(x) = \begin{cases} \frac{x}{6} & , \text{for } 0 \leq x < 3 \\ 2 - \frac{x}{2} & , \text{for } 3 \leq x \leq 4 \\ 0 & , \text{else} \end{cases}$$

例 (1 续)

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

(2)X 的分布函数因此为

$$F(x) = \begin{cases} 0 & \text{, for } x < 0 \\ \int_0^x \frac{x}{6} dx = \frac{x^2}{12} & \text{, for } 0 \le x < 3 \\ \int_0^3 \frac{x}{6} dx + \int_3^x 2 - \frac{x}{2} dx = -3 + 2x - \frac{x^2}{4} & \text{, for } 3 \le x \le 4 \\ 1 & \text{, for } x \ge 4 \end{cases}$$

(3) 因此

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

注意: 对于连续型随机变量 X 取单个实数值的概率为 0 ($P\{X = a\} = 0$), 所以对于连续随机变量, 可以不区分开区间或者闭区间.

定义

若连续型随机变量 X 具有概率密度

$$f(x)=\begin{cases}\frac{1}{b-a} & \text{, for } a<x<b\\ 0 & \text{, else}\end{cases} \tag{8}$$

则称 X 在区间 (a,b) 上服从**均匀分布**，记 $X\sim U(a,b)$.

易知 (1)$f(x)\geq 0$,(2)$\int_{-\infty}^{\infty}f(x)dx=1$,(3)$X$ 落在 (a,b) 的子区间的概率只依赖于子区间长度而与位置无关

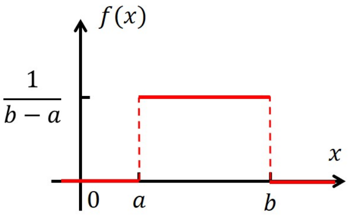
$$P\{c<X\leq c+L\}=\int_c^{c+L}f(x)dx=\int_c^{c+L}\frac{1}{b-a}dx=\frac{L}{b-a}$$

由公式 (8) 可得 X 的分布函数为

$$F(x)=\begin{cases}0 & \text{, for } x<a\\ \frac{x-a}{b-a} & \text{, for } a\leq x<b\\ 1 & \text{, for } x\geq b\end{cases} \tag{9}$$

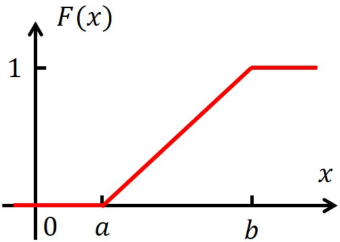
均匀分布的概率密度

$$f(x) = \begin{cases} \frac{1}{b-a} & , \text{for } a < x < b \\ 0 & , \text{else} \end{cases}$$



均匀分布的分布函数

$$F(x) = \begin{cases} 0 & , \text{for } x < a \\ \frac{x-a}{b-a} & , \text{for } a \leq x < b \\ 1 & , \text{for } x \geq b \end{cases}$$



例 (2)

设电阻值 R 是一个随机变量，均匀分布在 $900\Omega \sim 1100\Omega$. 求 R 得概率密度及 R 落在 $950\Omega \sim 1050\Omega$ 的概率.

解 R 的概率密度为

$$f(r) = \begin{cases} \frac{1}{1100-900} & , \text{for } 900 < r < 1100 \\ 0 & , \text{else} \end{cases}$$

故有

$$P\{950 < R \leq 1050\} = \int_{950}^{1050} \frac{1}{200} dr = 0.5$$

例 (ex)

在区间 $(-1,2)$ 上随机取一个数 X, 求 (1)X 的概率密度,(2)$P\{X>0\}$ 的概率,(3) 若取 10 个数, 则 10 个中恰有 2 个大于 0 的概率.

解 (1)X 是均匀分布, 其概率密度为

$$f(x)=\begin{cases} 1/(2-(-1))=1/3 & , \text{for } -1<x<2 \\ 0 & , \text{else} \end{cases}$$

(2)

$$P\{X>0\}=1-P\{X\le 0\}=1-\int_{-1}^{0} f(x)dx=1-1/3=2/3$$

(3) 设 10 个数中有 Y 个大于 0, 则 Y 服从 $Y\sim b(10,2/3)$

$$P\{Y=2\}=C_{10}^{2}(\frac{2}{3})^2(\frac{1}{3})^8$$

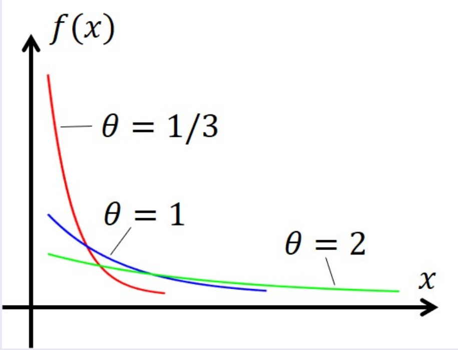
二、指数分布

定义

若连续型随机变量 X 的概率密度为

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & , \text{for } x > 0 \\ 0 & , \text{else} \end{cases} \tag{10}$$

其中 $\theta > 0$ 为常数, 则称 X 服从参数为 θ 的**指数分布**.



易知,(1)$f(x) \geq 0$,(2)$\int_{-\infty}^{\infty} f(x)dx = 1$. 由公式 (10) 可得 X 的分布函数为

$$F(x) = \begin{cases} 1 - e^{-x/\theta} & , \text{for } x > 0 \\ 0 & , \text{else} \end{cases} \tag{11}$$

指数分布具有无记忆性的性质 (如公式 (12)): 对于任意实数 $s, t > 0$ 有

$$P\{X > s + t | X > s\} = P\{X > t\} \tag{12}$$

证明.

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

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

$$= \frac{e^{-(s+t)/\theta}}{e^{-s/\theta}} = e^{-t/\theta} = P\{X > t\}$$

□

例 (ex)

某大型设备无故障运行的时间 T 服从参数为 $\theta = 1/\lambda$ 的指数分布,(1) 写下 T 的累积分布函数,(2) 已知设备无故障运行 10 小时, 求继续无故障运行 8 小时的概率.

解 (1) T 的概率密度和概率分布分别为

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & , \text{for } t > 0 \\ 0 & , \text{else} \end{cases} \qquad F(t) = \begin{cases} 1 - e^{-\lambda t} & , \text{for } t > 0 \\ 0 & , \text{else} \end{cases}$$

(2)

$$P\{T \geq (10+8) | T \geq 10\} = \frac{P\{T \geq 18\}}{P\{T \geq 10\}} = \frac{1 - F(18)}{1 - F(10)}$$

$$= \frac{e^{-18\lambda}}{e^{-10\lambda}} = e^{-8\lambda} = P\{T \geq 8\}$$

三、正态分布

定义

若连续型随机变量 X 的概率密度为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty \tag{13}$$

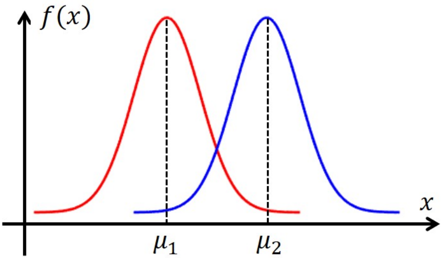
其中 $\mu,\sigma(\sigma>0)$ 为常数，则称 X 服从参数为 μ,σ 的正态分布或高斯分布 (Gaussian)，记为 $X\sim N(\mu,\sigma^2)$.

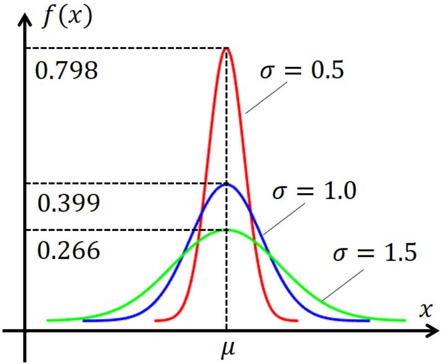
易知,(1)$f(x)\geq 0$,(2)$\int_{-\infty}^{\infty}f(x)dx=1$(证明见教材公式 4.11). 且具有如下性质:

1. 曲线关于 $x=\mu$ 对称，这表明对于任意 $h>0$ 有
$$P\{\mu-h<X\leq\mu\}=P\{\mu<X\leq\mu+h\}$$
2. 当 $x=\mu$ 时取得最大值
$$f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$$

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

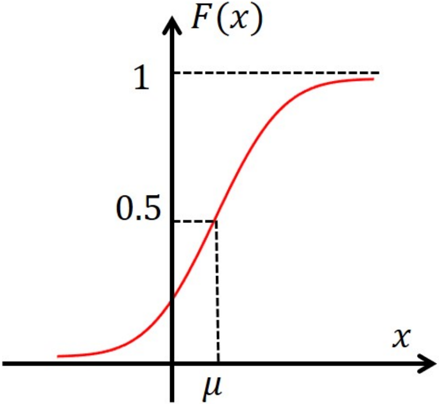
1. 正态分布的概率密度函数 $f(x)$ 在 $x=\mu\pm\sigma$ 处有拐点，曲线以 Ox 轴为渐近线.
2. 固定 σ，改变 μ 将沿着 Ox 轴平移曲线，而不改变形状.
3. 固定 μ，改变 σ 将改变 $f(x)$ 最大值.





由式 (13) 可得 X 的分布函数

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt \tag{14}$$

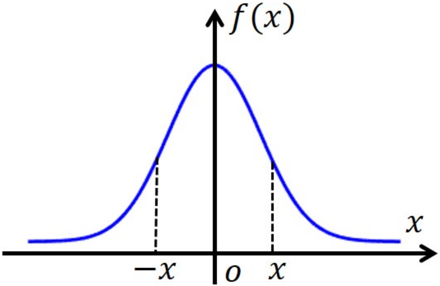


定义

特别地，当 $\mu=0,\sigma=1$ 时，称随机变量 X 服从**标准正态分布**，其概率密度和分布函数分别用 $\varphi(x)$ 和 $\Phi(x)$ 表示:

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \qquad (15) \qquad \Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt \qquad (16)$$

易知, $\Phi(-x)=1-\Phi(x)\rightarrow\Phi(x)+\Phi(-x)=1$.



引理

若 $X \sim N(\mu, \sigma^2)$，则随机变量 $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

证明.

随机变量 Z 的分布函数为

$$P\{Z \le x\} = P\{\frac{X-\mu}{\sigma} \le x\} = P\{X \le \mu+\sigma x\} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\mu+\sigma x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

令 $\frac{t-\mu}{\sigma} = y$，得

$$P\{Z \le x\} = \int_{-\infty}^{x} e^{-y^2/2} dy = \Phi(x)$$

因此，随机变量 $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$. □

利用以上引理可知，于是对随机变量 $X \sim N(\mu, \sigma^2)$，其分布函数可写为

$$F(x) = P\{X \leq x\} = P\{\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\} = \Phi(\frac{x-\mu}{\sigma})$$

例如，对于 $X \sim N(1, 2^2)$，可知

$$P\{0 < X \leq 1.6\} = \Phi(\frac{1.6-1}{2}) - \Phi(\frac{0-1}{2}) = \Phi(0.3) - \Phi(-0.5) = 0.3094$$

类似地，使用上述引理，

$$P\{\mu-\sigma < X < \mu+\sigma\} = \Phi(1)-\Phi(-1)$$

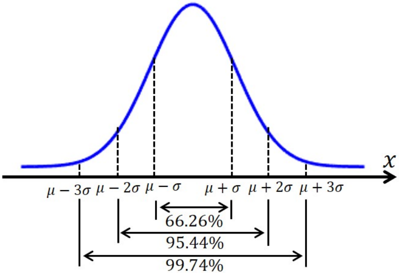
$$= 2\Phi(1) - 1 = 68.26\%$$

$$P\{\mu-2\sigma < X < \mu+2\sigma\} = \Phi(2)-\Phi(-2)$$

$$= 2\Phi(2) - 1 = 95.44\%$$

$$P\{\mu-3\sigma < X < \mu+3\sigma\} = \Phi(3)-\Phi(-3)$$

$$= 2\Phi(3) - 1 = 99.74\%$$



例 (3)

将一温度调节器放置在贮存着某种液体的容器内. 调节器整定在 d^oC, 液体的温度 X^oC 是一个随机变量,$X \sim N(d, 0.5^2)$.
求:(1) 若 $d = 90^oC$, 则 $X < 89^oC$ 的概率?(2) 若要求 $X \geq 80^oC$ 的概率不低于 0.99, 则 d 至少为多少?

解 (1)

$$P\{X < 89\} = P\{\frac{X-90}{0.5} < \frac{89-90}{0.5}\} = \Phi(-2) = 1 - \Phi(2) = 0.0228$$

(2)

$$0.99 \leq P\{X \geq 80\} = P\{\frac{X-d}{0.5} \geq \frac{80-d}{0.5}\} = 1 - P\{\frac{X-d}{0.5} < \frac{80-d}{0.5}\}$$

$$= 1 - \Phi(\frac{80-d}{0.5})$$

即

$$\Phi(\frac{d-80}{0.5}) \geq 0.99 = \Phi(2.327) \Rightarrow \frac{d-80}{0.5} \geq 2.327 \Rightarrow d \geq 81.1635$$

例 (ex1)

一批钢材长度 $X(cm)$ $N(\mu,\sigma^2)$,(1) 若 $\mu=100cm,\sigma=2cm$, 求这批钢材长度小于 $97.8cm$ 的概率,(2) 若 $\mu=100cm$, 要使这批钢材的长度有 90% 落在区间 $(97,103)$ 内,σ 至少取多少.

解 (1)

$$P\{X<97.8\}=\Phi(\frac{97.8-100}{2})=\Phi(-1.1)=1-\Phi(1.1)=0.1357$$

(2) 令 $P\{97<X<103\}\geq 90\%$, 即

$$\Phi(\frac{103-100}{\sigma})-\Phi(\frac{97-100}{\sigma})=2\Phi(\frac{3}{\sigma})-1\geq 90\%$$

$$\Rightarrow\Phi(\frac{3}{\sigma})\geq 0.95\Rightarrow\frac{3}{\sigma}\geq 1.645\Rightarrow\sigma\leq 1.8237$$

例 (ex2)

设某地区男子身高 $X(cm) \sim N(169.7, 4.1^2)$, 求 (1) 从该地区随机找一男子测身高，则他高于 $175cm$ 的概率,(2) 若从中随机找 5 男子，则至少有一人高于 $175cm$ 的概率是多少,(3) 恰有一人高于 $175cm$ 的概率是多少.

解 (1)

$$P\{X > 175\} = 1 - P\{X \le 175\} = 1 - \Phi(\frac{175 - 169.7}{4.1})$$

$$= 1 - \Phi(1.293) = 0.0985$$

(2) 设 5 人中有 Y 人身高大于 $175cm$, 则 $Y \sim b(5, 0.0985)$

$$P\{Y \ge 1\} = 1 - P\{Y = 0\} = 1 - (1 - 0.0985)^5 = 0.4045$$

(3)

$$P\{Y = 1\} = C_5^1(0.0985)^1(1 - 0.0985)^4 = 0.3253$$

2.5 随机变量的函数的分布

例 (1)

设随机变量 X 具有以下分布律, 求随机变量 $Y=(X-1)^2$ 的分布律.

X	-1	0	1	2
p_k	0.2	0.3	0.1	0.4

解 Y 的可能取值为 $0,1,4$

$$P\{Y=0\}=P\{(X-1)^2=0\}=P\{X=1\}=0.1$$

$$P\{Y=1\}=P\{X=0\}+P\{X=2\}=0.7$$

$$P\{Y=4\}=P\{X=-1\}=0.2$$

所以 $Y=(X-1)^2$ 的分布律为

Y	0	1	4
p_k	0.1	0.7	0.2

例 (2)

设随机变量 X 具有如下概率密度, 求随机变量 $Y=2X+8$ 的概率密度

$$f_X(x)=\begin{cases}\frac{x}{8} & \text{, for } 0<x<4\\ 0 & \text{, else}\end{cases}$$

解 记 X,Y 的分布函数分别为 $F_X(x),F_Y(y)$.

$$F_Y(y)=P\{Y\le y\}=P\{2X+8\le y\}=P\{X\le \frac{y-8}{2}\}=F_X(\frac{y-8}{2})$$

将 $F_Y(y)$ 关于 y 求导数, 得到 $Y=2X+8$ 的概率密度为

$$f_Y(y)=f_X(\frac{y-8}{2})(\frac{y-8}{2})'=\frac{(\frac{y-8}{2})}{8}\frac{1}{2}=\frac{y-8}{32}, 0<\frac{y-8}{2}<4$$

整理, 得

$$f_Y(y)=\begin{cases}\frac{y-8}{32} & \text{, for } 8<y<16\\ 0 & \text{, else}\end{cases}$$

例 (3)

设随机变量 X 具有概率密度 $f_X(x), -\infty < x < \infty$, 求 $Y = X^2$ 的概率密度.

解 记 X, Y 的分布函数分别为 $F_X(x), F_Y(y)$. 由于 $Y = X^2 \geq 0$, 故当 $y \leq 0$ 时,$F_Y(y) = 0$; 当 $y > 0$ 时,

$$F_Y(y) = P\{Y \leq y\} = P\{X^2 \leq y\}$$

$$= P\{-\sqrt{y} \leq y \leq \sqrt{y}\} = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

将 $F_Y(y)$ 关于 y 求导数, 得到 $Y = X^2$ 的概率密度为

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})] & , \text{for } y > 0 \\ 0 & , \text{else} \end{cases} \tag{17}$$

定义

设 $X \sim N(0,1)$, 则有式 (17) 可得 $Y = X^2$ 的概率密度

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2} & , \text{for } y > 0 \\ 0 & , \text{else} \end{cases} \tag{18}$$

称 Y 服从自由度为 1 的χ^2 分布.

Proposition

设随机变量 X 有概率密度 $f_X(x), -\infty < x < \infty$, 设函数 $g(x)$ 处处可导且恒有 $g'(x) > 0$(或 $g'(x) < 0$), 则 $Y = g(X)$ 是连续型随机变量, 其概率密度为

$$f_Y(y) = \begin{cases} f_X[h(y)] \cdot |h'(y)| & , for\ \alpha < y < \beta \\ 0 & , else \end{cases} \tag{19}$$

其中,α 和 β 是 $g(y)$ 在 $(-\infty, \infty)$ 的最小、大值,$h(y)$ 是 $g(x)$ 的反函数.

证明.

这里证明 $g'(x) > 0(g'(x) < 0$ 类似),$g(x)$ 在 $(-\infty, \infty)$ 单调增, 反函数 $h(y)$ 存在, 且在 (α, β) 单调增. 记 X, Y 的分布函数分别为 $F_X(x), F_Y(y)$.
(1) 因为 $Y = g(X) \in (\alpha, \beta)$, 当 $y \le \alpha$ 时,$F_Y(y) = P\{Y \le y\} = 0$;
(2) 当 $y \ge \beta$ 时,$F_Y(y) = P\{Y \le y\} = 1$;
(3) 当 $\alpha < y < \beta$ 时,

$$F_Y(y) = P\{Y \le y\} = P\{g(X) \le y\} = P\{X \le h(y)\} = F_X[h(y)]$$

将 $F_Y(y)$ 关于 y 求导数, 得到 $Y = g(X)$ 的概率密度为

$$f_Y(y) = \begin{cases} f_X[h(y)] \cdot h'(y) & , \text{for } \alpha < y < \beta \\ 0 & , \text{else} \end{cases} \tag{20}$$

对于 $g'(x) < 0$

$$f_Y(y) = \begin{cases} f_X[h(y)] \cdot -h'(y) & , \text{for } \alpha < y < \beta \\ 0 & , \text{else} \end{cases} \tag{21}$$

□

例 (4)

设随机变量 $X \sim N(\mu, \sigma^2)$, 证明 X 的线性函数 $Y = aX + b (a \neq 0)$ 也服从正态分布.

证

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$y = g(x) = ax + b, x = h(y) = (y - b)/a, h'(y) = 1/a$. 由式 (19) 可知

$$f_Y(y) = f_X[h(y)] \cdot |h'(y)| = \frac{1}{|a|} f_X(\frac{y-b}{a}) = \frac{1}{|a|\sigma\sqrt{2\pi}} e^{-\frac{[y-(b+a\mu)]^2}{2(a\sigma)^2}}$$

即 $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$. 若取 $a = 1/\sigma, b = -\mu/\sigma$, 则有

$$Y = \frac{X-\mu}{\sigma} \sim N(0,1)$$

即为上节的引理.

例 (5)

设电压 $V = A\sin\Theta$,A 为常数,Θ 是随机变量且 $\Theta \sim U(-\frac{\pi}{2}, \frac{\pi}{2})$, 求 V 得概率密度.

解 $v = g(\theta) = A\sin\theta$ 在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 上单调且恒有 $g'(\theta) = A\cos\theta > 0$, 反函数

$$\theta = h(v) = \arcsin\frac{v}{A}, h'(v) = \frac{1}{\sqrt{A^2 - v^2}}$$

Θ 的概率密度为

$$f_\Theta(\theta) = \begin{cases} \frac{1}{\pi} & , \text{for } -\frac{\pi}{2} < y < \frac{\pi}{2} \\ 0 & , \text{else} \end{cases} \tag{22}$$

由式 (19) 可知

$$f_V(v) = \begin{cases} \frac{1}{\pi} \cdot \frac{1}{\sqrt{A^2 - v^2}} & , \text{for } -A < v < A \\ 0 & , \text{else} \end{cases} \tag{23}$$

章节小结

知识点:

1. 随机变量
2. 离散型随机变量及其分布律
   ① 0-1 分布
   ② 二项分布
   ③ 泊松分布，泊松定理
3. 随机变量的分布函数
   分布函数或称为累计分布函数 $F(x) = P\{X \le x\}$
4. 连续型随机变量及其概率密度
   ① 均匀分布
   ② 指数分布
   ③ 正态分布，标准正态分布
5. 随机变量的函数的分布

$$f_Y(y) = \begin{cases} f_X[h(y)] \cdot |h'(y)| & , \text{for } \alpha < y < \beta \\ 0 & , \text{else} \end{cases}$$

The End