参考

第一章：随机事件与概率

1.1 随机事件及其运算

1.1.1 随机现象

例子

- 一天内进入某超市的顾客数
- 某种型号手机的寿命

定义

- 在一定的条件下，并不总出现相同结果的现象

特点

- 结果不止一个
- 事先不知道哪一个会出现

统计规律性

- 随机现象的各种结果会表现出一定的规律性

1.1.2 几种概念

随机试验

- 对**随机现象**进行的实验与观察
- 2个特定
  - 随机性
  - 重复性

样本空间Ω

- 所有**样本点**构成的集合
- 样本点
  - 随机试验的每一个可能结果

- 两类样本空间
  - 离散：样本点有限个或可列个
  - 连续：样本点无限不可列个
- 样本空间的分割
  - 若$A_i, A_2, ..., A_n$有这两点特点
    1. A_i互不相容
    2. $A_1 \cup A_2 \cup ... \cup A_n = \Omega$
  - 则称$A_i, A_2, ..., A_n$为Ω的一组分割

1.1.3 随机事件

定义

- 某些样本点组成的集合
- 样本空间Ω的子集，常用A、B、C、...表示

分类

- 基本事件
  - Ω的单点集
- 必然事件
  - Ω
- 不可能事件
  - 空集φ

事件的表示

- 语言：在试验中，A中某个样本点出现了，就说A出现了、发生了，记为A
- 集合：维恩图
- 随机变量

事件间的关系

- 包含关系
  - $A \subset B$
  - A发生导致B发生
- 相等关系
  - A=B
- 互不相容
  - A和B不可能同时发生

事件间的运算

- 并
  - $A \cup B$
  - A、B至少有一发生
- 交
  - $A \cap B$
  - A、B同时发生
- 差
  - A-B

    - A发生但B不发生
- 对立
    - $\bar{A}$
    - A不发生
- 对应维恩图

图片来源于网络

- 一些公式

图片来源于网络

图片来源于网络

图片来源于网络

事件域

- 设Ω为样本空间，Φ是由Ω的子集组成的集合类，若Φ满足以下三点，则称Φ为一个事件域，又称为σ域
    - $\Omega \in \Phi$
    - 若$A \in \Phi$，则对应事件$\bar{A} \in \Phi$
    - 若$A_n \in \Phi, n = 1, 2, 3, \ldots$，则可列并$\cup_{n=1}^{\infty} A_n \in \Phi$

练习

1

口袋中有a个白球、b个黑球，从中一个一个不返回地取球。

A = 取到最后一个是白球

B = 取到最后一段是白球

问A与B的关系?

2

若A是B的子事件，则$

- $$\begin{aligned} A \cup B &= ? \\ AB &= ? \end{aligned} \tag{1}$$

3

试用A\B\C表示下列事件

- 仅A出现
- 恰有一个出现
- 至多有一个出现
- 至多有两个出现
- 不都出现

1.1.4 随机变量

定义

- 表示随机现象结果的变量
- 常用X、Y、Z、...表示

1.1.5 本节小结

图片来源于网络

1.1.6 练习

- 课本练习题

1.2 概率的定义及其确定方法

1.2.1 概率的定义

直观定义

- 事件A出现的可能性大小

统计定义

- 事件A在大量重复试验下出现的频率的稳定值称为该事件的概率

公理化定义

- 非负性
  - $P(A) \geq 0$
- 正则性
  - $P(\Omega) = 1$
- 可列可加性
- 若$A_1, A_2, \ldots, A_n, \ldots$互不相容，则$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

1.2.2 概率的确定方法

排列组合

排列

- 从n个元素中任取r个，求取法有多少种
- 排列要按照顺序，是有序的
- 全排列
  - $P_n = n!$
- $0! = 1$
- 重复排列
  - n^r
- 选排列
  - $p_n^r = \frac{n!}{(n-r)!} = n(n-1)\ldots(n-r+1)$

组合

- $C_n^r = \frac{n!}{r!(n-r)!}$
- 重复组合
  - 重复组合(combination with repetition)是一种特殊的组合，从n个不同元素中可重复地选取m个元素，不管其顺序合成一组，称为从n个元素中取m个元素的可重复组合

  - C_{n+r-1}^{r}
- 组合是无序的

注意

- 求排列、组合时，要掌握和注意
  - 加法原则
  - 乘法原则

加法原则

- 完成某件事情有n类途径，在第一类途径中有m_1中方法，在第二类中有m_2种，依此类推，在第n类种有m_n种，则完成这件事共有$m_1+m_2+...+m_n$种不同的方法

乘法原理

- 完成某件事情需先后分成n个步骤，做第一步有m_1种，...，做第n步有m_n，则共有$m_1*...*m_n$种不同方法

练习

1

6根草，头两两相接，尾两两相接，求成环的概率？

hint：乘法原理

2

n个人围一圆桌坐，求甲、乙相邻而坐的概率？

3

n个人坐成一排，求甲、乙相邻而坐的概率？

频率方法

- 随机试验可大量重复进行
- 进行n次重复试验，记n(A)为事件A的频数
- 称$f_n(A)=\frac{n(A)}{n}$为时间A的频率
- 频率$f_n(A)$会稳定于某一常数（稳定值）
- 用该稳定值作为该事件的概率

古典方法

- 设Ω为样本空间，若
  - Ω只含有有限个样本点
  - 每个样本点出现的可能性相等
- 则事件A的概率为：
  - $P(A)=\frac{A\text{中样本点的个数}}{\text{样本点总数}}$

注意

- 抛一枚硬币三次等价于抛三枚硬币一次
- $\Omega_1=\{(0,0),(0,1),(1,0),(1,1)\}$
  - 此样本空间中的样本点等可能
- $\Omega_2=\{(\text{三正}),(\text{二正一反}),(\text{二反一正}),(\text{三反})\}$

- 此样本空间中的样本点不等可能

几何方法

- 若样本空间Ω充满某个区域，其度量（长度、面积等）为S_{Ω}
- 落在Ω中的任一子区域A的概率
- 只与子区域的度量S_A有关，而与子区域的位置无关（等可能的）
- 此时事件A的概率为：$P(A)=\frac{S_{\Omega}}{A_{\Omega}}$

例子

蒲丰投针问题

平面上画有间隔为d的等距平行线，向平面任意投掷一枚长为l（l<d）的针，求针与平行线相交的概率？

hint：参考

1.3 概率的性质

性质1-空集的概率

- $P(\Phi)=0$
- 逆不一定成立

性质2-有限可加性

- 若$AB=\Phi$
- 则$P(A\cup B)=P(A)+P(B)$
- 可推广到n个互不相容的事件

性质3-对立事件公式

- $P(\bar{A})=1-P(A)$

性质4-概率的单调性

- 若$A\supset B$，则$P(A-B)=P(A)-P(B)$且$P(A)\geq P(B)$
- $P(A-B)=P(A)-P(AB)$

性质5-加法公式

- $P(A\cup B)=P(A)+P(B)-P(AB)$
- $P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)$

例子

1

$AB=\Phi, P(A)=0.6, P(A\cup B)=0.8$，求B的对立事件的概率？

2

$P(A)=0.4, P(B)=0.3, P(A\cup B)=0.6$，求$P(A-B)$?

3

$P(A)=P(B)=P(C)=\frac{1}{4}, P(AB)=0, P(AC)=P(BC)=\frac{1}{6}$，求A、B、 C都不出现的概率？

4

口罩中有n-1个黑球，1个白球，每次从口袋中随机摸出一球，并换入一只黑球，求第k次取到黑球的概率？

hint：利用对立事件

5

口袋中有2个白球，每次从口袋中随机的摸出一球，并换入一只黑球，求第k次取到黑球的概率？

6

一颗骰子掷4次，求至少出现一次6点的概率？

hint：利用对立事件

7

从1-9中有返回取n次，求取出的n个数的乘积被10整除的概率？

hint：利用对立事件和加法公式，取到5且取到过偶数两个事件

8

甲掷硬币n+1次，乙掷n次，求甲掷出的正面数比乙掷出的正面数多的概率

hint：利用对称性，$P(\text{甲}_{\text{正}} > \text{乙}_{\text{正}}) = P(n+1-\text{甲}_{\text{正}} > n-\text{乙}_{\text{正}})$
$= P(\text{甲}_{\text{反}} - 1 < \text{乙}_{\text{反}}) = P(\text{甲}_{\text{反}} \le \text{乙}_{\text{反}}) = 1 - P(\text{甲}_{\text{正}} > \text{乙}_{\text{正}})$

常见模型

不返回抽样-超几何模型

N个产品，其中M个不合格产品，N-M个合格品，从中不返回任取n个，则此n个中有m个不合格品的概率为：

$$\begin{gathered}\frac{C_M^m * C_{N-M}^{n-m}}{C_N^n} \\ n \le N, m \le M \\ n-m \le N-M\end{gathered} \tag{2}$$

- 彩票问题

返回抽样

N个产品，其中M个不合格产品，N-M个合格品，从中有返回任取n个，则此n个中有m个不合格品的概率为：

$$\begin{gathered}\frac{C_n^m * M^m * (N-M)^{n-m}}{N^n} \\ m \le n\end{gathered} \tag{3}$$

盒子模型

n个不同球放入N个不同的盒子中，每个盒子中所放球不限，求恰有n个盒子中各有一球的概率（$n \le N$
）

$$\frac{P_N^n}{N^n} \tag{4}$$

- 生日问题-求n个人至少有两人生日相同的概率

配对模型

n个人、n顶帽子、任意取，至少一个人拿对自己帽子的概率

记A_i表示第i个人拿对自己帽子，则问题对应为求$P(\cup_{i=1}^n A_i)$

利用加法公式可求

性质6-连续性

事件序列的极限

- 若事件序列$\{F_n\}$满足$F_1 \subset F_2 \subset \ldots \subset F_n \subset \ldots$
  - 则称事件序列$\{F_n\}$为单独不减事件，
  - 其极限事件为：$\lim_{n\to\infty} F_n = \cup_{n=1}^{\infty} F_n$
- 若事件序列$\{F_n\}$满足$F_1 \supset F_2 \supset \ldots \supset F_n \supset \ldots$
  - 则称事件序列$\{F_n\}$为单独不增事件，
  - 其极限事件为：$\lim_{n\to\infty} F_n = \cap_{n=1}^{\infty} F_n$

集合函数的连续性

设P(.)为一个集合函数

- 若对任意单独不减事件序列$\{F_n\}$，有$P(\lim_{n\to\infty} F_n) = \lim_{n\to\infty} P(F_n)$
  - 则称P(.)为下连续的
- 若对任意单独不增事件序列$\{F_n\}$，有$P(\lim_{n\to\infty} F_n) = \lim_{n\to\infty} P(F_n)$
  - 则称P(.)为上连续的

概率的连续性

- 若P(.)是事件域σ上的一个概率函数，则P(.)为下连续的，也是上连续的

性质7-集合函数可列可加性充要条件

- 若P(.)是事件域σ上满足非负、正则的集合函数，则P(.)有可列可加性的充要条件是它具有有限可加性和下连续性

1.4 条件概率

10张卡片依次分给10个人，有3张写有中奖，已知第1个人没中后，第2个人中的概率是多少？

1.4.1 定义

- 对于事件A、B，若$P(B) > 0$，则称$P(A|B) = \frac{P(AB)}{P(B)}$为在B出现的条件下，A出现的条件概率

例子

10个产品有7个正品，3个次品，从中不返回的抽取2个，已知第一个取到次品，求第2个又取到次品的概率？

hint: 用定义求取$P(A|B) = \frac{P(AB)}{P(B)}$

条件概率是概率

- 条件概率满足概率的三条公理（哪三条？）
- 由此得：
  - $P(A\cup B|C)=P(A|C)+P(B|C)-P(AB|C)$
  - $P(\bar{A}|B)=1-P(A|B)$
  - $P(\Omega|B)=1$
  - $P(B|\Omega)\neq 1$
  - $P(A|\Omega)=P(A)$
  - $P(A|A)=1$

练习

$P(A)=0.6,P(A\cup B)=0.84,P(\Omega-B|A)=0.4,P(B)=?$

1.4.2 条件概率三大公式

乘法公式

性质

- 若$P(B)>0$，则$P(AB)=P(B)*P(A|B)$
- 若$P(A_1A_2...A_{n-1})>0$，则$P(A_1A_2...A_n)=P(A_1)P(A_2|A_1)...P(A_n|A_1A_2...A_{n-1})$

例子

一批零件共有100个，其中10个不合格品，从中一个个不返回取出，求第三次才取出不合格品得概率？

解答：

记A_i是第i次取出的是不合格品，B_i是第i次取出的是合格品，则求的是$P(B_1B_2A_3)$，用乘法公式即可求出

全概率公式

性质

- 若事件$B_1,B_2,...,B_n$是样本空间的一组分割，且$P(B_i)>0$，则

$$P(A)=\sum_{i=1}^{n}P(AB_i)=\sum_{i=1}^{n}P(B_i)P(A|B_i) \tag{5}$$

- 全概率用于求复杂事件的概率
- 使用全概率公式关键在于寻找另一组事件来“分割”样本空间
- 全概率公式最简单的形式

$$P(A)=P(B)P(A|B)+P(\bar{B})P(A|\bar{B}) \tag{6}$$

- 若事件$B_1,B_2,...,B_n$是互不相容的，且$P(B_i)>0$，则由$A\subset\cup_{i=1}^{n}B_i$可得公式（5）

例子

1

设10件产品中有3件不合格品，从中不返回的取2次，每次一件，求取出第二件为不合格品的概率？

hint: 设A=第一次取得不合格品，B=第二次取得不合格品，即求$P(B)$，用全概率公式

2

甲口袋有a只白球，b只黑球；乙口袋有n只白球，m只黑球。从甲口袋任取一球放入乙口袋，然后从乙口袋任取一球，求从乙口袋取出的是白球的概率？

hint：设A=从甲口袋取出的白球，B=从乙口袋取出的白球，即求$P(B)$，用全概率公式

贝叶斯公式

性质

- 若事件$B_1, B_2, ..., B_n$是样本空间Ω的一组分割，且$P(A) > 0, P(B_i) > 0$，则

$$P(B_i|A) = \frac{P(AB_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)} \tag{7}$$

- 事件$B_1, B_2, ..., B_n$可以看作是导致A发生的原因
- $P(B_j|A)$是在事件A发生的条件下，某个原因B_j发生的概率，称为后验概率
- 贝叶斯公式又称为后验概率公式
- $P(B_j)$为先验概率

例子

某商品由三个厂家供应，其供应量为：甲厂家是乙厂家的2倍；乙、丙两场相等。各厂产品的次品率为2%，2%，4%。若从市场上随机抽取一件此种商品，发现是次品，求它是甲厂生产的概率？

hint：贝叶斯公式，设A_i为取到第i厂的产品，B为取到次品，即求$P(A_{甲}|B)$

三大公式比较

- 乘法公式是求“几个事件同时发生”的概率
- 全概率公式是求“最后结果”的概率
- 贝叶斯公式是已知“最后结果”，求“原因”的概率

1.5 独立性

1.5.1 两个事件的独立性

定义

- 若事件A与B满足$P(AB) = P(A)P(B)$
- 则称A与B相互独立

结论

- A、B为两个事件，若$P(A) > 0$，则A与B独立等价于$P(B|A) = P(B)$

性质

- 若事件A与B独立，则A与$\bar{B}$独立、$\bar{A}$与B独立、$\bar{A}$与$\bar{B}$独立

怎么判断

- 实际应用中，往往根据经验来判断两个事件的独立性
- 例如：甲乙两人分别工作

1.5.2 多个事件的独立性

定义

- 若事件$A_1,A_2,...,A_n$满足：两两独立、三三独立、...、nn独立
- 则称事件$A_1,A_2,...,A_n$相互独立

结论

- 若A、B、C相互独立，则
  - $A\cup B$与C独立
  - $A\cap B$与C独立
  - $A-B$与C独立

例子

1

甲乙两人独立的对同一目标射击一次，其命中率分别为0.6和0.7，现已知目标被击中，求它是甲击中的概率？

hint：独立性，设A=甲中，B=乙中，C=被击中，即求$P(A|C)$，贝叶斯公式

$$\begin{aligned}P(A|C)=\frac{P(AC)}{P(C)}=\frac{P(A)}{P(A\cup B)}\\=\frac{P(A)}{P(A)+P(B)-P(AB)}\\=\frac{P(A)}{P(A)+P(B)-P(A)P(B)}\end{aligned}\tag{8}$$

2

两枪手轮流对同一目标进行设计，甲先射，谁先击中则得胜。每次射击中，甲、乙射中的概率分为为a和b，求甲得胜的概率？

hint：只考虑第一次和第二次，甲中+甲不中乙不中，这样又会开始新一轮循环，则

$P(\text{甲胜})=a+(1-a)(1-b)P(\text{甲胜})$

1.5.3 试验的独立性

定义

- 若试验E_1的任一结果与试验E_2的任一结果都是相互独立的事件，则称这两个试验相互独立

n重伯努利试验

伯努利试验

- 若某种试验只有两个结果，则称该试验为伯努利试验
- 在伯努利试验中，一般记“成功”的概率为p

n重伯努利试验

- n次独立重复的伯努利试验

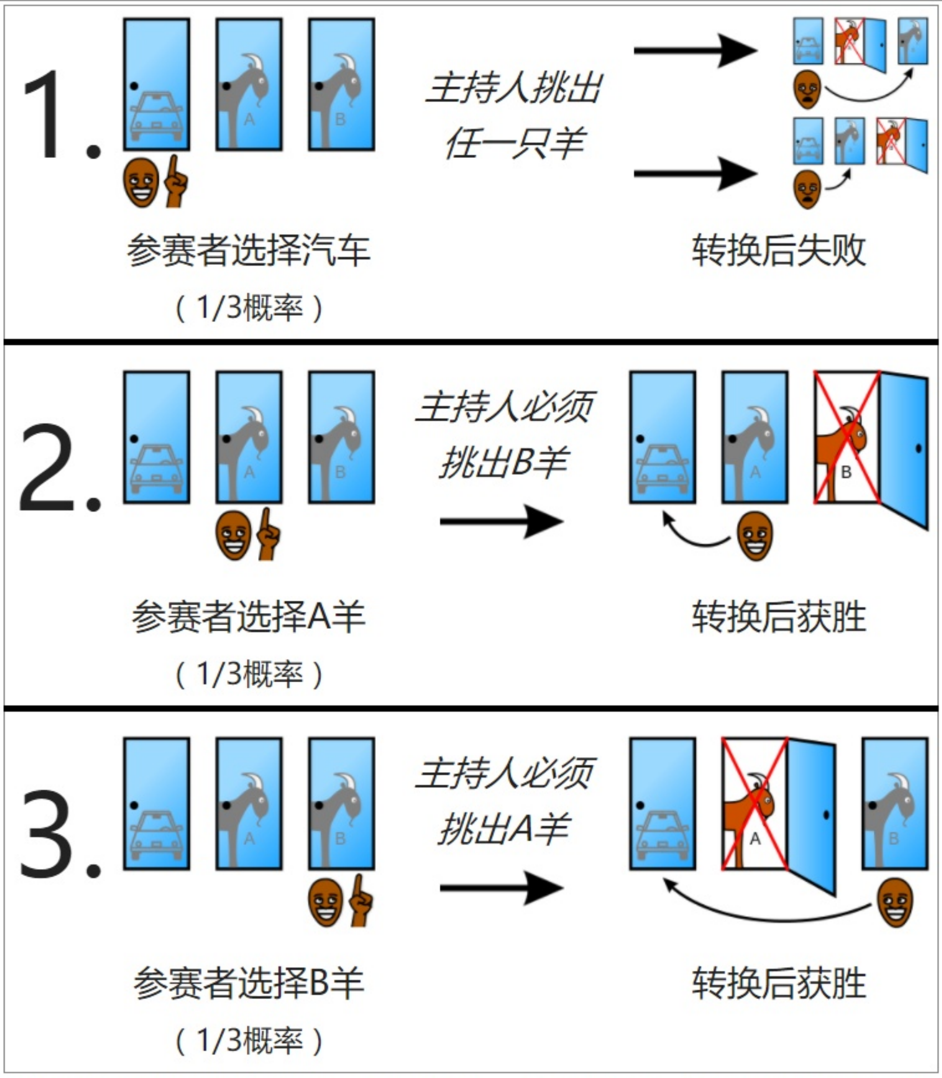
n重伯努利试验的成功次数

- 在n重伯努利试验中，记成功的次数为X
- X的可能取值为：0,1,...,n
- X取值为k的概率为：
  - $P(X=k)=C_n^kp^k(1-p)^{n-k}$

第二章：随机变量及其分布

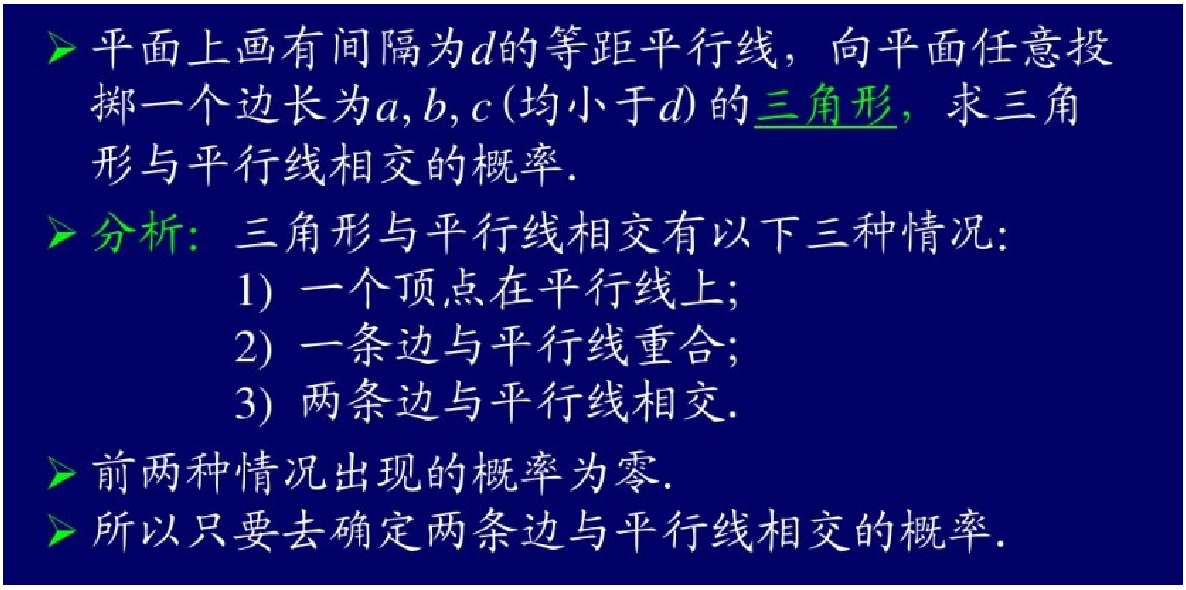
[参考](#)

- 三门问题



参赛者最初选择时有1/3的相同概率选择汽车、A羊和B羊，转换后的获胜概率为2/3。

- 三角形问题



解：记$P_{ab}, P_{ac}, P_{bc}, P_a, P_b, P_c$分别为边$ab, ac, bc, a, b, c$与平行线相交的概率，则所求概率为

$$p=P(\text{三角形与平行线相交})=P_{ab}+P_{ac}+P_{bc}.$$

由蒲丰投针问题知

$$P_a=2a/(d\pi),\quad P_b=2b/(d\pi),\quad P_c=2c/(d\pi).$$

因为 $P_a=P_{ab}+P_{ac},\quad P_b=P_{ab}+P_{bc},\quad P_c=P_{ac}+P_{bc}$

所以 $P_a+P_b+P_c=2(P_{ab}+P_{ac}+P_{bc})$,

由此得

$$p=P_{ab}+P_{ac}+P_{bc}=(P_a+P_b+P_c)/2=(a+b+c)/(d\pi).$$

2.1 随机变量及其分布

- 掷一颗骰子，出现的点数X
- n个产品中的不合格品个数Y
- 某商场一天内来的顾客数Z
- 某钟型号电视机的寿命T

2.1.1 随机变量的定义

定义2.1.1

- 设Ω为某随机现象的样本空间
- 称定义在Ω上的实值函数$X = X(\omega)$为随机变量

注意点（1）

- 随机变量$X(\omega)$是样本点ω的函数
  - 其定义域为ω，其值域为$R = (-\infty, +\infty)$
  - 若X表示掷一颗骰子出现的点数，则$\{X = 1.5\}$是不可能事件
- 若X为随机变量，则$\{X = k\}$、$\{a < X < b\}$、$\ldots$均为随机事件

注意点（2）

- 注意以下一些表达式
  - $\{X = k\} = \{X \le k\} - \{X < k\}$
  - $\{a < X < b\} = \{X \le b\} - \{X \le a\}$
  - $\{X > b\} = \Omega - \{X \le b\}$
- 同一样本空间可以定义不同的随机变量

两类随机变量

- 若随机变量X可能取值的个数为有限个或可列个，则称X为离散随机变量
- 若随机变量X的可能取值充满某个区间$[a, b]$，则称X为连续随机变量

2.1.2 随机变量的分布函数

定义2.1.2

- 设X为一个随机变量，对任意实数x，称$F(x)=P(X\leq x)$为X的分布函数

基本性质

1. $F(x)$单调不降
2. 有界
    1. $0\leq F(x)\leq 1$
    2. $F(-\infty)=0$
    3. $F(+\infty)=1$
3. 右连续

2.1.3 离散随机变量的分布列

定义2.1.3

- 设离散随机变量X的可能取值为：$x_1,x_2,\dots,x_n,\dots$
    - 称$p_i=P(X=x_i),i=1,2,\dots$为X的分布列

基本性质

- 非负性：$p_i\geq 0$
- 正则性：$\sum p_i=1$

注意点（1）

- 求离散随机变量的分布列应注意：
    - 确定随机变量的所有可能取值
    - 计算每个取值点的概率

注意点（2）

- 对离散随机变量的分布函数应注意
    - $F(x)$是递增的阶梯函数
    - 其间断点均为右连续的
    - 其间断点即为X的可能取值点
    - 其间断点的跳跃高度是对应的概率值

例2.1.1

已知X的分布列如下，求X的分布函数

X	0	1	2
p	1/3	1/6	1/2

hint：

$$F(x)=\begin{cases}0, & x<0\\ 1/3, & 0\leq x<1\\ 1/2, & 1\leq x<2\\ 1, & 2\leq x\end{cases}\tag{9}$$

例2.1.2

已知X的分布函数如下，求X的分布列

$$F(x)=\begin{cases}0, & x<0\\ 0.4, & 0\le x<1\\ 0.8, & 1\le x<2\\ 1, & 2\le x\end{cases} \tag{10}$$

hint：略

2.1.4 连续随机变量的密度函数

- 连续随机变量X的可能取值充满某个区间$[a,b]$
- 因为对连续随机变量X，有$P(X=x)=0$
    - 所以无法效仿离散随机变量用$P(X=x)$来描述连续随机变量X的分布

定义2.1.4

- 设随机变量X的分布函数为$F(x)$，若存在非负可积函数$p(x)$，满足：
    - $F(x)=\int_{-\infty}^{x}p(t)dt$
- 则称X为连续随机变量
- 称$p(x)$为概率密度函数，简称密度函数

基本性质

- 非负性：$p(x)\ge 0$
- 正则性：$\int_{-\infty}^{+\infty}p(x)dx=1$

例2.1.7

> **例 2.1.7** 向区间$(0,a)$上任意投点，用X表示这个点的坐标. 设这个点落在$(0,a)$中任一小区间的概率与这个小区间的长度成正比，而与小区间位置无关. 求X的分布函数和密度函数.
>
> **解** 记X的分布函数为$F(x)$，则
>
> 当$x<0$时，因为$\{X\leqslant x\}$是不可能事件，所以$F(x)=P(X\leqslant x)=0$；
>
> 当$x\geqslant a$时，因为$\{X\leqslant x\}$是必然事件，所以$F(x)=P(X\leqslant x)=1$；
>
> 当$0\leqslant x<a$时，有$F(x)=P(X\leqslant x)=P(0\leqslant X\leqslant x)=kx$，其中$k$为比例系数. 因为$1=F(a)=ka$，所以得$k=1/a$.
>
> 于是X的分布函数为
>
> $$F(x)=\begin{cases}0, & x<0,\\ \dfrac{x}{a}, & 0\leqslant x<a,\\ 1, & x\geqslant a.\end{cases}$$
>
> 下面求X的密度函数$p(x)$.
>
> 当$x<0$或$x>a$时，$p(x)=F'(x)=0$；
>
> 当$0<x<a$时，$p(x)=F'(x)=1/a$，
>
> 而在$x=0$和$x=a$处，$p(x)$可取任意值，一般就近取值为宜，这不会影响概率的计算，因为它们是几乎处处相等的密度函数. 于是X的密度函数为
>
> $$p(x)=\begin{cases}\dfrac{1}{a}, & 0<x<a,\\ 0, & 其他.\end{cases}$$

例2.1.8

例 2.1.8 某种型号电子元件的寿命 X（以小时计）具有以下的概率密度函数

$$p(x)=\begin{cases}\dfrac{1\,000}{x^2}, & x>1\,000,\\ 0, & \text{其他}.\end{cases}$$

现有一大批此种元件（设各元件工作相互独立），问：

（1）任取 1 只，其寿命大于 1 500 小时的概率是多少？

（2）任取 4 只，4 只寿命都大于 1 500 小时的概率是多少？

（3）任取 4 只，4 只中至少有 1 只寿命大于 1 500 小时的概率是多少？

（4）若已知一只元件的寿命大于 1 500 小时，则该元件的寿命大于 2 000 小时的概率是多少？

2.1.5 离散随机变量与连续随机变量的区别

1. 离散随机变量的分布函数总是右连续的阶梯函数，而连续随机变量的分布函数一定是整个数轴上的连续函数
2. 离散随机变量在其可能取值点上的概率不为0，而连续随机变量在其任意一点a上的概率恒为0
    1. 这表示不可能事件的概率为0，但概率为0的事件不一定是不可能事件
    2. 必然事件的概率为1，但概率为1的事件不一定是必然事件
3. 由于连续随机变量仅取一点的概率恒为0，从而在事件$\{a \le X \le b\}$中剔除$x=a$或$x=b$，不影响其概率
4. 由于在若干点上改变密度函数$p(x)$的值不影响其积分的值，从而不影响其分布函数的值，这意味着一个连续分布的密度函数不唯一

思考

存不存在既非离散又非连续的分布？

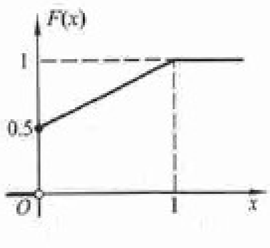


图 2.1.8 既非离散又非连续的分布函数示例

$$F(x)=\begin{cases}0, & x<0\\ \frac{1+x}{2}, & 0\le x<1\\ 1, & x\ge 1\end{cases} \tag{11}$$

习题

1. 口袋中有 5 个球，编号为 1，2，3，4，5. 从中任取 3 个，以 X 表示取出的 3 个球中的最大号码.

(1) 试求 X 的分布列；

(2) 写出 X 的分布函数，并作图.

2. 一颗骰子抛两次，求以下随机变量的分布列：

(1) X 表示两次所得的最小点数；

(2) Y 表示两次所得的点数之差的绝对值.

3. 口袋中有 7 个白球、3 个黑球.

(1) 每次从中任取一个不放回，求首次取出白球的取球次数 X 的概率分布列；

(2) 如果取出的是黑球则不放回，而另外放入一个白球，此时 X 的概率分布列如何.

4. 有 3 个盒子，第一个盒子装有 1 个白球、4 个黑球；第二个盒子装有 2 个白球、3 个黑球；第三个盒子装有 3 个白球、2 个黑球. 现任取一个盒子，从中任取 3 个球. 以 X 表示所取到的白球数.

(1) 试求 X 的概率分布列；

(2) 取到的白球数不少于 2 个的概率是多少？

5. 掷一颗骰子 4 次，求点数 6 出现的次数的概率分布.

6. 从一副 52 张的扑克牌中任取 5 张，求其中黑桃张数的概率分布.

7. 一批产品共有 100 件，其中 10 件是不合格品. 根据验收规则，从中任取 5 件产品进行质量检验，假如 5 件中无不合格品，则这批产品被接收，否则就要重新对这批产品逐个检验.

(1) 试求 5 件中不合格品数 X 的分布列；

(2) 需要对这批产品进行逐个检验的概率是多少？

8. 设随机变量 X 的分布函数为

$$F(x)=\begin{cases}0, & x<0,\\ 1/4, & 0\leqslant x<1,\\ 1/3, & 1\leqslant x<3,\\ 1/2, & 3\leqslant x<6,\\ 1, & x\geqslant 6.\end{cases}$$

试求 X 的概率分布列及 $P(X<3)$，$P(X\leqslant 3)$，$P(X>1)$，$P(X\geqslant 1)$.

9. 设随机变量 X 的分布函数为

$$F(x)=\begin{cases}0, & x<1,\\ \ln x, & 1\leqslant x<\mathrm{e},\\ 1, & x\geqslant \mathrm{e}.\end{cases}$$

试求 $P(X<2)$，$P(0<X\leqslant 3)$，$P(2<X<2.5)$.

10. 若 $P(X\geqslant x_1)=1-\alpha$，$P(X\leqslant x_2)=1-\beta$，其中 $x_1<x_2$，试求 $P(x_1\leqslant X\leqslant x_2)$.

11. 从1,2,3,4,5五个数中任取三个,按大小排列记为 $x_1<x_2<x_3$,令 $X=x_2$,试求:

(1) X 的分布函数;

(2) $P(X<2)$ 及 $P(X>4)$.

12. 设随机变量 X 的密度函数为

$$p(x)=\begin{cases}1-|x|, & -1\leqslant x\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

试求 X 的分布函数.

13. 如果随机变量 X 的密度函数为

$$p(x)=\begin{cases}x, & 0\leqslant x<1,\\ 2-x, & 1\leqslant x<2,\\ 0, & \text{其他}.\end{cases}$$

试求 $P(X\leqslant 1.5)$.

14. 设随机变量 X 的密度函数为

$$p(x)=\begin{cases}A\cos x, & |x|\leqslant \dfrac{\pi}{2},\\ 0, & |x|>\dfrac{\pi}{2}.\end{cases}$$

试求:

(1) 系数 A;

(2) X 落在区间 $(0,\pi/4)$ 内的概率.

15. 设连续随机变量 X 的分布函数为

$$F(x)=\begin{cases}0, & x<0,\\ Ax^2, & 0\leqslant x<1,\\ 1, & x\geqslant 1.\end{cases}$$

试求:

(1) 系数 A;

(2) X 落在区间 $(0.3,0.7)$ 内的概率;

(3) X 的密度函数.

16. 学生完成一道作业的时间 X 是一个随机变量,单位为小时.它的密度函数为

$$p(x)=\begin{cases}cx^2+x, & 0\leqslant x\leqslant 0.5,\\ 0, & \text{其他}.\end{cases}$$

(1) 确定常数 c;

(2) 写出 X 的分布函数;

(3) 试求在 20 min 内完成一道作业的概率;

(4) 试求 10 min 以上完成一道作业的概率.

17. 某加油站每周补给一次油.如果这个加油站每周的销售量(单位:千升)为一随机变量,其密度函数为

$$p(x)=\begin{cases}0.05\left(1-\dfrac{x}{100}\right)^4, & 0<x<100,\\ 0, & \text{其他}.\end{cases}$$

试问该油站的储油罐需要多大,才能把一周内断油的概率控制在5%以下?

18. 设随机变量 X 和 Y 同分布,X 的密度函数为

$$p(x)=\begin{cases}\dfrac{3}{8}x^2, & 0<x<2,\\ 0, & \text{其他}.\end{cases}$$

已知事件 $A=\{X>a\}$ 和 $B=\{Y>a\}$ 独立,且 $P(A\cup B)=3/4$,求常数 a.

19. 设连续随机变量 X 的密度函数 $p(x)$ 是一个偶函数,$F(x)$ 为 X 的分布函数,求证对任意实数 $a>0$,有

(1) $F(-a)=1-F(a)=0.5-\int_0^a p(x)\,\mathrm{d}x$;

(2) $P(|X|<a)=2F(a)-1$;

(3) $P(|X|>a)=2[1-F(a)]$.

2.2 随机变量的数学期望

2.2.1 数学期望的概念

定义2.2.1

- 设离散随机变量X的分布列为$p(x_i)=P(X=x_i),i=1,2,\ldots,n,\ldots$
- 如果$\sum_{i=1}^{\infty}|x_i|p(x_i)<\infty$，
- 则称$E(X)=\sum_{i=1}^{\infty}x_ip(x_i)$为随机变量X的数学期望
- 若不满足则称X的数学期望不存在

定义2.2.2

- 设连续随机变量X的密度函数为p(x)
- 如果$\int_{-\infty}^{+\infty}|x|p(x)dx<\infty$
- 则称$E(X)=\int_{-\infty}^{\infty}xp(x)dx$为X的数学期望
- 如果不满足条件，则称数学期望不存在

例2.2.4

设X服从区间(a,b)上的均匀分布，求E(X)

hint：X的密度函数为

$$p(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & other\end{cases} \tag{12}$$

所以

$$E(X)=\int_a^b x*\frac{1}{b-a}dx=\frac{a+b}{2} \tag{13}$$

例2.2.5

求柯西分布的数学期望?

hint：因为柯西分布的密度函数为$p(x)=\frac{1}{\pi}\frac{1}{1+x^2},-\infty<x<\infty$

所有$E(X)=\int_{-\infty}^{+\infty}|x|*\frac{1}{\pi}\frac{1}{1+x^2}dx=\infty$

故柯西分布的数学期望不存在

2.2.2 数学期望的性质

定理2.2.1

- 若随机变量X的分布用分布列$p(x_i)$或用密度函数$p(x)$表示，则X的某一函数$g(X)$的数学期望为

$$E[g(X)]=\begin{cases}\sum_i g(x_i)p(x_i), & 离散\\ \int_{-\infty}^{+\infty}g(x)p(x)dx, & 连续\end{cases} \tag{14}$$

性质2.2.1

- 若c是常数，则$E(c)=c$

性质2.2.2

- 对任意常数a，有$E(aX)=aE(X)$

性质2.2.3

- 对任意的两个函数$g_1(x)$和$g_2(x)$，有
- $E[g_1(X) \pm g_2(X)] = E[g_1(x)] \pm E[g_2(x)]$

习题2.2

> 17. 设随机变量 X 的概率密度函数为
>
> $$p(x)=\begin{cases}\frac{1}{2}\cos\frac{x}{2}, & 0 \le x \le \pi; \\ 0, & 其他.\end{cases}$$
>
> 对 X 独立重复观察 4 次，Y 表示观察值大于$\pi/3$ 的次数，求 Y^2 的数学期望.
>
> 解：Y 的全部可能取值为 0, 1, 2, 3, 4，因 $p=P\{X>\frac{\pi}{3}\}=\int_{\frac{\pi}{3}}^{\pi}\frac{1}{2}\cos\frac{x}{2}dx=\sin\frac{x}{2}\Big|_{\frac{\pi}{3}}^{\pi}=\sin\frac{\pi}{2}-\sin\frac{\pi}{6}=\frac{1}{2}$，
>
> 则 $P\{Y=0\}=(1-p)^4=\frac{1}{16}$，$P\{Y=1\}=\binom{4}{1}\cdot p(1-p)^3=\frac{4}{16}$，$P\{Y=2\}=\binom{4}{2}\cdot p^2(1-p)^2=\frac{6}{16}$，
>
> $P\{Y=1\}=\binom{4}{3}\cdot p^3(1-p)=\frac{4}{16}$，$P\{Y=4\}=p^4=\frac{1}{16}$，
>
> 故 $E(Y^2)=0^2\times\frac{1}{16}+1^2\times\frac{4}{16}+2^2\times\frac{6}{16}+3^2\times\frac{4}{16}+4^2\times\frac{1}{16}=\frac{80}{16}=5$.
>
> 18. 设随机变量 X 的密度函数为
>
> $$p(x)=\begin{cases}\frac{3}{8}x^2, & 0<x<2; \\ 0, & 其他.\end{cases}$$
>
> 试求 $\frac{1}{X^2}$ 的数学期望.
>
> 解：$E\left(\frac{1}{X^2}\right)=\int_0^2\frac{1}{x^2}\cdot\frac{3}{8}x^2dx=\int_0^2\frac{3}{8}dx=\frac{3}{4}$.

2.3 随机变量的方差和标准差

2.3.1 方差和标准差的定义

定义2.3.1

- 若随机变量X^2的数学期望$E(X^2)$存在，则称偏差平方$(X-E(X))^2$的数学期望为随机变量X的方差，记为

$$Var(X) = E(X-E(X))^2 = \begin{cases} \sum_i (x_i - E(x_i))^2 p(x_i), & 离散 \\ \int_{-\infty}^{+\infty} (x - E(x))^2 p(x) dx, & 连续 \end{cases} \tag{15}$$

- 称方差的正平方根$\sqrt{Var(X)}$为随机变量X的标准差，记为$\sigma(X)$

2.3.2 方差的性质

性质2.3.1

- $Var(X) = E(X^2) - [E(X)]^2$

性质2.3.2

- 常数的方差为0

性质2.3.3、

- 若a,b是常数，则$Var(aX+b) = a^2 Var(X)$

例子

例 2.3.3 设 X 为掷一颗骰子出现的点数，试求 $\mathrm{Var}(X)$.

解

$$E(X)=\frac{1}{6}(1+2+3+4+5+6)=\frac{7}{2},$$

$$E(X^2)=\frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2)=\frac{91}{6},$$

$$\mathrm{Var}(X)=\frac{91}{6}-\frac{49}{4}=\frac{35}{12}=2.917.$$

2.3.3 切比雪夫不等式

定理2.3.1

- 设随机变量X的数学期望和方差都存在，则对任意常数ϵ，有
- $P(|X-E(X)|\geq\epsilon)\leq\frac{Var(X)}{\epsilon^2}$或者
- $P(|X-E(X)|<\epsilon)\geq 1-\frac{Var(X)}{\epsilon^2}$

定理2.3.2

- 若随机变量X的方差存在，则$Var(X)=0$的充要条件是X几乎处处为某个常数a，即$P(X=a)=1$
- 证明如下

证明 充分性是显然的，下面证必要性. 设 $\mathrm{Var}(X)=0$，这时 $E(X)$ 存在. 因为

$$\{|X-E(X)|>0\}=\bigcup_{n=1}^{\infty}\left\{|X-E(X)|\geqslant\frac{1}{n}\right\},$$

所以有

$$\begin{aligned}P(|X-E(X)|>0)&=P\left(\bigcup_{n=1}^{\infty}\left\{|X-E(X)|\geqslant\frac{1}{n}\right\}\right)\\&\leqslant\sum_{n=1}^{\infty}P\left(|X-E(X)|\geqslant\frac{1}{n}\right)\\&\leqslant\sum_{n=1}^{\infty}\frac{\mathrm{Var}(X)}{(1/n)^2}=0,\end{aligned}$$

其中最后一个不等式用到了切比雪夫不等式. 由此可知

$$P(|X-E(X)|>0)=0,$$

因而有

$$P(|X-E(X)|=0)=1,$$

即

$$P(X=E(X))=1,$$

这就证明了结论，且其中的常数 a 就是 $E(X)$.

习题2.3

1. 设随机变量 X 满足 $E(X)=\mathrm{Var}(X)=\lambda$，已知 $E[(X-1)(X-2)]=1$，试求 λ.

2. 假设有 10 只同种电器元件，其中有两只不合格品. 装配仪器时，从这批元件中任取一只，如是不合格品，则扔掉重新任取一只；如仍是不合格品，则扔掉再取一只，试求在取到合格品之前，已取出的不合格品数的方差.

3. 已知 $E(X)=-2, E(X^2)=5$，求 $\mathrm{Var}(1-3X)$.

4. 设 $P(X=0)=1-P(X=1)$，如果 $E(X)=3\mathrm{Var}(X)$，求 $P(X=0)$.

5. 设随机变量 X 的分布函数为

$$F(x)=\begin{cases}\dfrac{e^x}{2}, & x<0,\\ \dfrac{1}{2}, & 0\leqslant x<1,\\ 1-\dfrac{1}{2}e^{-\frac{1}{2}(x-1)}, & x\geqslant 1,\end{cases}$$

试求 $\mathrm{Var}(X)$.

11. 设随机变量 X 取值 $x_1\leqslant\cdots\leqslant x_n$ 的概率分别是 $p_1,\cdots,p_n$，$\sum_{k=1}^{n}p_k=1$. 证明

$$\mathrm{Var}(X)\leqslant\left(\frac{x_n-x_1}{2}\right)^2.$$

12. 设 $g(x)$ 为随机变量 X 取值的集合上的非负不减函数，且 $E(g(X))$ 存在，证明：对任意的 $\varepsilon>0$，有

$$P(X>\varepsilon)\leqslant\frac{E(g(X))}{g(\varepsilon)}.$$

13. 设 X 为非负随机变量，$a>0$. 若 $E(e^{aX})$ 存在，证明：对任意的 $x>0$，有

$$P(X\geqslant x)\leqslant\frac{E(e^{aX})}{e^{ax}}.$$

14. 已知正常成人男性每升血液中的白细胞数平均是 7.3×10^9，标准差是 0.7×10^9. 试利用切比雪夫不等式估计每升血液中的白细胞数在 5.2×10^9 至 9.4×10^9 之间的概率的下界.

2.4 常用离散分布

2.4.1 二项分布

一、二项分布

定义

- 若X的分布列为$P(X=k)=C_n^kp^k(1-p)^{n-k},k=0,1,\ldots,n$
- 则称X服从二项分布记为$B(n,p)$

例2.4.1

例 2.4.1 某特效药的临床有效率为 0.95，今有 10 人服用，问至少有 8 人治愈的概率是多少？

解 设 X 为 10 人中被治愈的人数，则 $X \sim b(10, 0.95)$，而所求概率为

· 93 ·

$$P(X \geqslant 8)=P(X=8)+P(X=9)+P(X=10)$$
$$=\binom{10}{8}0.95^{8}0.05^{2}+\binom{10}{9}0.95^{9}0.05+\binom{10}{10}0.95^{10}$$
$$=0.0746+0.3151+0.5988=0.9885.$$

10 人中有 8 人以上被治愈的概率为 0.988 5.

例 2.4.2 设随机变量 $X \sim b(2, p)$，$Y \sim b(3, p)$. 若 $P(X \geqslant 1)=5/9$，试求 $P(Y \geqslant 1)$.

解 由 $P(X \geqslant 1)=5/9$，知 $P(X=0)=4/9$，所以 $(1-p)^{2}=4/9$，由此得 $p=1/3$. 再由 $Y \sim b(3, p)$ 可得

$$P(Y \geqslant 1)=1-P(Y=0)=1-\left(1-\frac{1}{3}\right)^{3}=\frac{19}{27}.$$

二、二点分布

- n=1时的二项分布称为二点分布，又称0-1分布，或称伯努利分布

三、二项分布的数学期望和方差

- $E(X) = np$
- $Var(X) = np(1-p)$

2.4.2 泊松分布

一、泊松分布

- 分布列为$P(X=k) = \frac{\lambda^k}{k!}\exp^{-\lambda}, k=0,1,2,\dots,\lambda>0$
- 记为$P(\lambda)$

二 、泊松分布的数学期望和方差

- $E(X) = \lambda$
- $Var(X) = \lambda$

三、二项分布的泊松近似

定理2.4.1

- 在n重伯努利试验种，记时间A在一次试验种发生的概率为p_n（与试验次数n无关），如果当$n \rightarrow \infty$时，有$np_n \rightarrow \lambda$，则
- $\lim_{n\rightarrow\infty} C_n^k p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!}\exp^{-\lambda}$

例子

已知某种疾病的发病率为0.001，某单位共有5000人，问该单位患有这种疾病的人数不超过5人的概率为多少？

hint：设患病人数为X，则X服从二项分布，即求

$$P(X \le 5) = \sum_{k=0}^{5} C_{5000}^{k} 0.001^k 0.999^{5000-k} \approx \sum_{k=0}^{5} \frac{5^k}{k!} \exp^{-5} \tag{16}$$

因为$\lambda = np = 5$

2.4.3 超几何分布

一、超几何分布

- $P(X=k) = \frac{C_M^k * C_{N-M}^{n-k}}{C_N^n}$

二、数学期望与方差

- $E(X) = nM/N$
- $Var(X) = \frac{nM(N-M)(N-n)}{N^2(N-1)}$

超几何分布的二项近似

- 当n远小于N时，超几何分布可用二项分布近似

$$\frac{C_M^k * C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k (1-p)^{n-k}, p = M/N \tag{17}$$

2.4.4 几何分布和负二项分布

一、几何分布

- $P(X=k) = (1-p)^{k-1}, k = 1, 2, \dots$
- $E(X) = 1/p$
- $Var(x) = \frac{1-p}{p^2}$
- [几何分布的无记忆性]，$P(X > m+n | X > m) = P(X > n)$

二、负二项分布

- $P(X=k) = C_{k-1}^{r-1} p^r (1-p)^{k-r}, k = r, r+1, r+2, \dots$

习题2.4

1. 一批产品中有 10% 的不合格品，现从中任取 3 件，求其中至多有一件不合格品的概率.

2. 一条自动化生产线上产品的一级品率为 0.8，现检查 5 件，求至少有 2 件一级品的概率.

3. 某优秀射手命中 10 环的概率为 0.7，命中 9 环的概率为 0.3. 试求该射手三次射击所得的环数不少于 29 环的概率.

4. 经验表明：预定餐厅座位而不来就餐的顾客比例为 20%. 如今餐厅有 50 个座位，但预定给了 52 位顾客，问到时顾客来到餐厅而没有座位的概率是多少？

5. 设随机变量 $X \sim b(n, p)$，已知 $E(X) = 2.4$，$\mathrm{Var}(X) = 1.44$，求两个参数 n 与 p 各为多少？

6. 设随机变量 X 服从二项分布 $b(2, p)$，随机变量 Y 服从二项分布 $b(4, p)$. 若 $P(X \geqslant 1) = 8/9$，试求 $P(Y \geqslant 1)$.

18. 令 $X(n,p)$ 表示服从二项分布 $b(n,p)$ 的随机变量，试证明：

$$P(X(n,p)\leqslant i)=1-P(X(n,1-p)\leqslant n-i-1).$$

19. 设随机变量 X 服从参数为 p 的几何分布，试证明：

$$E\left(\frac{1}{X}\right)=\frac{-p\ln p}{1-p}$$

20. 设随机变量 $X\sim b(n,p)$，试证明：

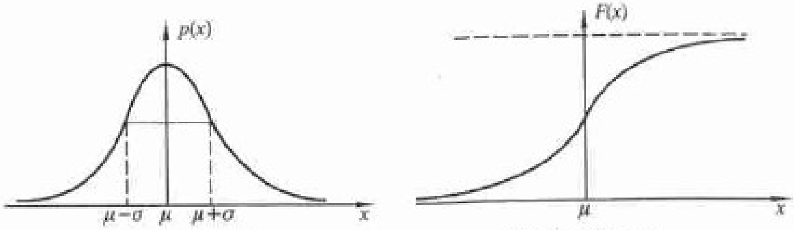
$$E\left(\frac{1}{X+1}\right)=\frac{1-(1-p)^{n+1}}{(n+1)p}.$$

2.5 常用连续分布

2.5.1 正态分布

一、正态分布的密度函数和分布函数

- 若随机变量X的密度函数为：$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty$
- 则称X服从正态分布，称X为正态变量，记作X\~$N(\mu,\sigma^2),\sigma>0$
- 它的分布函数为$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}\exp^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$



(a) 密度函数 $p(x)$　　(b) 分布函数 $F(x)$

图 2.5.1　正态分布

图 2.5.2 给出了在 μ 和 σ 变化时，相应正态密度曲线的变化情况.

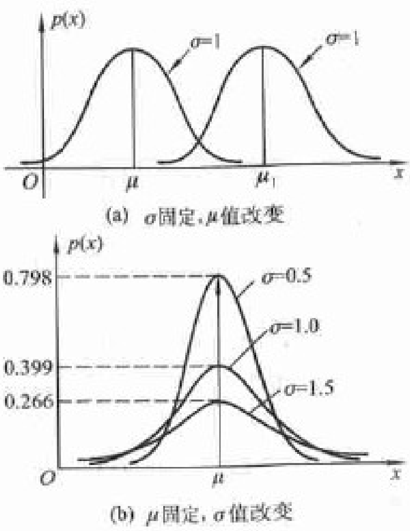


图 2.5.2　正态密度函数

二、标准正态分布

- 当$\mu=0,\sigma=1$时正态分布N(0,1)称为标准正态分布
- 此时记标准正态变量为U，密度函数为$\phi(u)$，分布函数为$\Phi(u)$

- 标准正态分布有以下特定
    - $\Phi(-u) = 1 - \Phi(u)$
    - $P(U > u) = 1 - \Phi(u)$
    - $P(a < U < b) = \Phi(b) - \Phi(a)$
    - $P(|U| < c) = 2\Phi(c) - 1 (c \geq 0)$

三、正态变量的标准化

- 正态分布家族

$$F = |N(\mu, \sigma^2) : -\infty < \mu < +\infty, sigma > 0| \tag{18}$$

定理2.5.1

- 若随机变量$X \sim N(\mu, \sigma^2)$，则$U = \frac{(X-\mu)}{\sigma} \sim N$（0，1）

例子2.5.2

> **例 2.5.2** 设随机变量 X 服从正态分布 $N(108, 3^2)$，试求：
> (1) $P(102<X<117)$；
> (2) 常数 a，使得 $P(X<a)=0.95$.
> **解** 利用公式(2.5.4)及查附表 2 得
> (1)
> $$P(102 < X < 117) = \Phi\left(\frac{117-108}{3}\right) - \Phi\left(\frac{102-108}{3}\right)$$
> $$= \Phi(3) - \Phi(-2) = \Phi(3) + \Phi(2) - 1$$
> $$= 0.9987 + 0.9772 - 1 = 0.9759.$$
> (2) 由
> $$P(X < a) = \Phi\left(\frac{a-108}{3}\right) = 0.95, \quad 或 \quad \Phi^{-1}(0.95) = \frac{a-108}{3},$$
> 其中 Φ^{-1} 为 Φ 的反函数. 从附表 2 由里向外反查得
> $$\Phi(1.64) = 0.9495, \qquad \Phi(1.65) = 0.9505,$$
> 再用线性内插法可得 $\Phi(1.645) = 0.95$，即 $\Phi^{-1}(0.95) = 1.645$，故
> $$\frac{a-108}{3} = 1.645,$$
> 从中解得 $a = 112.935$.

四、正态分布的数学期望和方差

- $E(X) = \mu$
- $Var(X) = \sigma^2$

五、正态分布的3σ原则

$$\begin{aligned} P(|X-\mu| < k\sigma) &= \Phi(k) - \Phi(-k) \\ &= 2\Phi(k) - 1 \\ &= \begin{cases} 0.6826, & k=1 \\ 0.9545, & k=2 \\ 0.9973, & k=3 \end{cases} \end{aligned} \tag{19}$$

2.5.2 均匀分布

一、均匀分布的密度函数和分布函数

$$p(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & \text{其他}\end{cases} \tag{20}$$

$$F(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\le x<b\\ 1, x\ge b\end{cases} \tag{21}$$

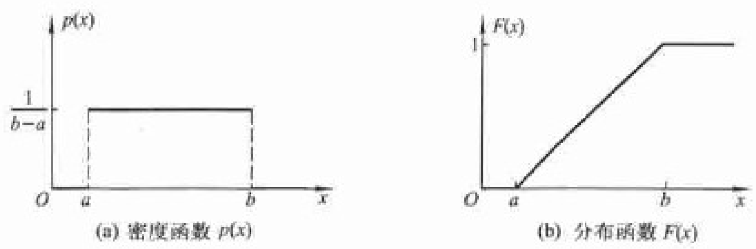


图 2.5.3 (a,b)上的均匀分布

二、数学期望和方差

- $E(X)=\frac{a+b}{2}$
- $Var(X)=\frac{(b-a)^2}{12}$

三、例子

例 2.5.4 设随机变量 X 服从$(0,10)$上的均匀分布，现对 X 进行4次独立观测，试求至少有3次观测值大于5的概率.

解 设随机变量 Y 是4次独立观测中观测值大于5的次数，则 $Y\sim b(4,p)$，其中 $p=P(X>5)$. 由 $X\sim U(0,10)$，知 X 的密度函数为

$$p(x)=\begin{cases}\frac{1}{10}, & 0<x<10,\\ 0, & \text{其他}.\end{cases}$$

所以

$$p=P(X>5)=\int_5^{10}\frac{1}{10}\mathrm{d}x=\frac{1}{2},$$

于是

$$P(Y\geqslant 3)=\binom{4}{3}p^3(1-p)+\binom{4}{4}p^4=4\left(\frac{1}{2}\right)^4+\left(\frac{1}{2}\right)^4=\frac{5}{16}.$$

2.5.3 指数分布

一、密度函数和分布函数

$$p(x)=\begin{cases}\lambda\exp^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases} \tag{22}$$

$$F(x)=\begin{cases}1-\exp^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases} \tag{23}$$

二、数学期望和方差

- $E(X)=\frac{1}{\lambda}$
- $Var(X)=\frac{1}{\lambda^2}$

三、指数分布的无记忆性

定理2.5.2

- 如果随机变量$X \sim Exp(\lambda)$，则对任意$s > 0, t > 0$，有
- $P(X > s + t | X > s) = P(X > t)$

例子

例 2.5.5 如果某设备在任何长为 t 的时间$[0,t]$内发生故障的次数 $N(t)$ 服从参数为 λt 的泊松分布，则相继两次故障之间的时间间隔 T 服从参数为 λ 的指数分布.

解 设 $N(t) \sim P(\lambda t)$，即

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \qquad k = 0, 1, \cdots.$$

注意到两次故障之间的时间间隔 T 是非负随机变量，且事件$\{T \geqslant t\}$说明此设备在$[0,t]$内没有发生故障，即$\{T \geqslant t\} = \{N(t) = 0\}$，由此我们得

当 $t<0$ 时，有 $F_T(t) = P(T \leqslant t) = 0$；

当 $t \geqslant 0$ 时，有

$$F_T(t) = P(T \leqslant t) = 1 - P(T > t) = 1 - P(N(t) = 0) = 1 - e^{-\lambda t},$$

所以 $T \sim Exp(\lambda)$，即相继两次故障之间的时间间隔 T 服从参数为 λ 的指数分布，图 2.5.5 示意其间关系.

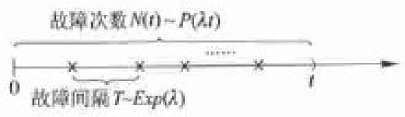


图 2.5.5 故障次数与故障间隔之间的关系

2.5.4 gamma分布

一、gamma函数

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \exp^{-x} dx \tag{24}$$

- $\Gamma(1) = 1$
- $\Gamma(1/2) = \sqrt{\pi}$
- $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
- 当α为自然数n时，有$\Gamma(n+1) = n\Gamma(n) = n!$

二、gamma分布

$$p(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{25}$$

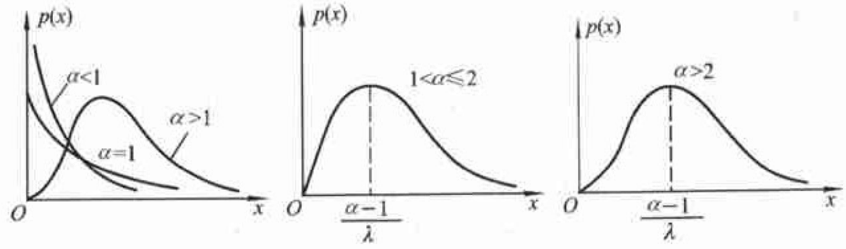


图 2.5.6　λ 固定、不同 α 的伽玛密度函数曲线

- 当 $0<\alpha<1$ 时，$p(x)$ 是严格下降函数，且在 $x=0$ 处有奇异点.
- 当 $\alpha=1$ 时，$p(x)$ 是严格下降函数，且在 $x=0$ 处 $p(0)=\lambda$.
- 当 $1<\alpha\leqslant 2$ 时，$p(x)$ 是单峰函数，先上凸、后下凸.

· 115 ·

- 当 $2<\alpha$ 时，$p(x)$ 是单峰函数，先下凸、中间上凸、后下凸. 且 α 越大，$p(x)$ 越近似于正态密度，但伽玛分布总是偏态分布，α 愈小其偏斜程度愈严重.

三、数学期望和方差

- $E(X)=\frac{\alpha}{\lambda}$
- $Var(X)=\frac{\alpha}{\lambda^2}$

四、两个特例

- $\alpha=1$时的gamma分布是指数分布，$Ga(1,\lambda)=Exp(\lambda)$
- $\lambda=1/2,\alpha=n/2$时的gamma分布式自由度为n的卡方分布，记为$\chi^2(n)$

例子

例 2.5.6 电子产品的失效常常是由于外界的"冲击引起". 若在$(0,t)$内发生冲击的次数$N(t)$服从参数为λt的泊松分布,试证第n次冲击来到的时间S_n服从伽玛分布$Ga(n,\lambda)$.

证 因为事件"第n次冲击来到的时间S_n小于等于t"等价于事件"$(0,t)$内发生冲击的次数$N(t)$大于等于n",即

$$\{S_n \leq t\} = \{N(t) \geq n\}.$$

于是,S_n的分布函数为

$$F(t) = P(S_n \leq t) = P(N(t) \geq n) = \sum_{k=n}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

用分部积分法可以验证下列等式

$$\sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} = \frac{\lambda^n}{\Gamma(n)} \int_t^{\infty} x^{n-1} e^{-\lambda x} dx. \qquad (2.5.16)$$

所以

$$F(t) = \frac{\lambda^n}{\Gamma(n)} \int_0^t x^{n-1} e^{-\lambda x} dx,$$

这就表明$S_n \sim Ga(n,\lambda)$. 证毕.

2.5.5 beta分布

一、beta函数

$$B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}dx, a>0, b>0 \qquad (26)$$

- $B(a,b)=B(b,a)$
- $B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

二、beta分布

$$p(x)=\begin{cases} \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} x^{a-1}(1-x)^{b-1}, & 0<x<1 \\ 0, & \text{其他} \end{cases} \qquad (27)$$

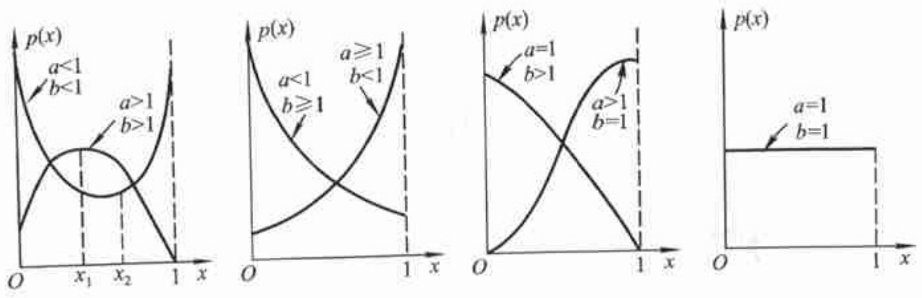


图 2.5.7　贝塔密度函数曲线

从上图可以看出：

- 当 $a<1,b<1$ 时，$p(x)$ 是下凸的 U 形函数.
- 当 $a>1,b>1$ 时，$p(x)$ 是上凸的单峰函数.
- 当 $a<1,b\geqslant1$ 时，$p(x)$ 是下凸的单调减函数.
- 当 $a\geqslant1,b<1$ 时，$p(x)$ 是下凸的单调增函数.
- 当 $a=1,b=1$ 时，$p(x)$ 是常数函数，且 $Be(1,1)=U(0,1)$.

因为服从贝塔分布 $Be(a,b)$ 的随机变量是仅在区间 $(0,1)$ 取值的，所以不合格品率、机器的维修率、市场的占有率、射击的命中率等各种比率选用贝塔分布作为它们的概率分布是恰当的，只要选择合适的参数 a 与 b 即可.

三、数学期望和方差

- $E(X)=\frac{a}{a+b}$
- $Var(X)=\frac{ab}{(a+b)^2(a+b+1)}$

习题2.5

习　题　2.5

1. 设随机变量 X 服从区间 $(2,5)$ 上的均匀分布，求对 X 进行 3 次独立观测中，至少有 2 次的观测值大于 3 的概率.

2. 在 $(0,1)$ 上任取一点记为 X，试求 $P\left(X^2-\frac{3}{4}X+\frac{1}{8}\geqslant0\right)$.

3. 设 K 服从 $(1,6)$ 上的均匀分布，求方程 $x^2+Kx+1=0$ 有实根的概率.

4. 若随机变量 $K\sim N(\mu,\sigma^2)$，而方程 $x^2+4x+K=0$ 无实根的概率为 0.5，试求 μ.

5. 设流经一个 2 Ω 电阻上的电流强度 I 是一个随机变量，它均匀分布在 9 A 至 11 A 之间. 试求此电阻上消耗的平均功率，其中功率 $W=2I^2$.

6. 某种圆盘的直径在区间 (a,b) 上服从均匀分布，试求此种圆盘的平均面积.

7. 设某种商品每周的需求量 X 服从区间 $(10,30)$ 上均匀分布，而商店进货数为区间 $(10,30)$ 中的某一整数，商店每销售 1 单位商品可获利 500 元；若供大于求则降价处理，每处理 1 单位商品亏损 100 元；若供不应求，则可从外部调剂供应，此时每 1 单位商品仅获利 300 元. 为使商店所获利润期望值不少于 9 280 元，试确定最少进货量.

8. 统计调查表明，英格兰 1875 年至 1951 年期间在矿山发生 10 人或 10 人以上死亡的两次事故之间的时间 T（以日计）服从均值为 241 的指数分布. 试求 $P(50<T<100)$.

9. 若一次电话通话时间 X（以 min 计）服从参数为 0.25 的指数分布，试求一次通话的平均时间.

10. 某种设备的使用寿命 X（以年计）服从指数分布，其平均寿命为 4 年. 制造此种设备的厂家规定，若设备在使用一年之内损坏，则可以予以调换. 如果设备制造厂每售出一台设备可赢利 100 元，而调换一台设备制造厂需花费 300 元. 试求每台设备的平均利润.

24. 某单位招聘员工，共有 10 000 人报考. 假设考试成绩服从正态分布，且已知 90 分以上有 359 人，60 分以下有 1 151 人. 现按考试成绩从高分到低分依次录用 2 500 人，试问被录用者中最低分为多少？

25. 设随机变量 X 服从正态分布 $N(60,3^2)$，试求实数 a,b,c,d 使得 X 落在如下五个区间中的概率之比为 7 : 24 : 38 : 24 : 7.

$$(-\infty,a],\quad (a,b],\quad (b,c],\quad (c,d],\quad (d,\infty).$$

26. 设随机变量 X 与 Y 均服从正态分布，X 服从正态分布 $N(\mu,4^2)$，Y 服从正态分布 $N(\mu,5^2)$，试比较以下 p_1 和 p_2 的大小.

$$p_1=P\{X\leqslant\mu-4\},\qquad p_2=P\{Y\geqslant\mu+5\}.$$

27. 设随机变量 X 服从正态分布 $N(0,\sigma^2)$，若 $P(|X|>k)=0.1$，试求 $P(X<k)$.

28. 设随机变量 X 服从正态分布 $N(\mu,\sigma^2)$，试问：随着 σ 的增大，概率 $P(|X-\mu|<\sigma)$ 是如何变化的？

29. 设随机变量 X 服从参数为 $\mu=160$ 和 σ 的正态分布，若要求 $P(120<X\leqslant200)\geqslant0.90$，允许 σ 最大为多少？

30. 设随机变量 $X\sim N(\mu,\sigma^2)$，求 $E|X-\mu|$.

31. 设随机变量 $X\sim N(0,\sigma^2)$，证明 $E|X|=\sigma\sqrt{\dfrac{2}{\pi}}$.

32. 设随机变量 X 服从伽玛分布 $Ga(2,0.5)$，试求 $P(X<4)$.

33. 某地区漏缴税款的比例 X 服从参数 $a=2,b=9$ 的贝塔分布，试求此比例小于 10% 的概率及平均漏缴税款的比例.

34. 某班级学生中数学成绩不及格的比率 X 服从 $a=1,b=4$ 的贝塔分布，试求 $P(X>E(X))$.

2.6 随机变量函数的分布

2.6.1 离散随机变量函数的分布

定义

- 设X是随机变量，X的分布列为

X	x_1	x_2	...	x_n	...
P	$p(x_1)$	$p(x_2)$	...	$p(x_n)$	...

- $Y=g(X)$的分布列为

X	$g(x_1)$	$g(x_2)$	...	$g(x_n)$	...
P	$p(x_1)$	$p(x_2)$	...	$p(x_n)$	...

例子

例 2.6.1 已知随机变量 X 的分布列如下，求 $Y=X^2+X$ 的分布列.

X	-2	-1	0	1	2
P	0.2	0.1	0.1	0.3	0.3

解 $Y=X^2+X$ 的分布列为

Y	2	0	0	2	6
P	0.2	0.1	0.1	0.3	0.3

再对相等的值合并，得

Y	0	2	6
P	0.2	0.5	0.3

2.6.2 连续随机变量函数的分布

分以下几种情况讨论$Y=g(X)$的分布

一、当$g(X)$为严格单调时

定理2.6.1

设X是连续随机变量，其密度函数为$p_X(x)$，$Y=g(X)$是另一个随机变量，若$y=g(x)$严格单调，其反函数$h(y)$有连续导函数，则$Y=g(X)$的密度函数为

$$p_Y(y)=\begin{cases}p_X[h(y)]|h^{'}(y)|, & a<y<b\\ 0, & \text{其他}\end{cases} \tag{28}$$
$$a=min\{g(-\infty),g(\infty)\}$$
$$b=max\{g(-\infty),g(\infty)\}$$

定理2.6.2

设随机变量X服从正态分布$N(\mu,\sigma^2)$，则当$a\neq 0$时，有$Y=aX+b\sim N(a\mu+b,a^2\sigma^2)$

定理2.6.3（对数正态分布）

设随机变量$X\sim N(\mu,\sigma^2)$，则$Y=\exp^X$的概率密度函数为

$$p_Y(y)=\begin{cases}\frac{1}{\sqrt{2\pi}y\sigma}\exp^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}, & y>0\\ 0, & y\leq 0\end{cases} \tag{29}$$

定理2.6.4

设随机变量X服从gamma分布$Ga(\alpha,\lambda)$，则当$k>0$时，有$Y=kX\sim Ga(\alpha,\frac{\lambda}{k})$

定理2.6.5

设随机变量X的分布函数$F_X(x)$为严格单调增的连续函数，其反函数$F_X^{-1}(y)$存在，则$Y=F_X(X)$服从(0,1)上的均匀分布$U(0,1)$

例子

例 2.6.2 （1）设随机变量 $X\sim N(10,2^2)$，试求 $Y=3X+5$ 的分布；

（2）设随机变量 $X\sim N(0,2^2)$，试求 $Y=-X$ 的分布.

解 （1）由定理 2.6.2 知 Y 仍是正态变量，其数学期望和方差分别为

$$E(Y)=E(3X+5)=3\times 10+5=35,$$

$$\mathrm{Var}(Y)=\mathrm{Var}(3X+5)=9\times 2^2=36.$$

所以 $Y=3X+5$ 的分布为 $N(35,6^2)$.

（2）Y 仍是正态变量，其数学期望和方差分别为

$$E(Y)=E(-X)=0,$$

$$\mathrm{Var}(Y)=\mathrm{Var}(-X)=2^2.$$

所以 $Y=-X$ 的分布仍为 $N(0,2^2)$. 这表明 X 与 $-X$ 有相同的分布，但这两个随机变量是不相等的. 所以我们要明确，分布相同与随机变量相等是两个完全不同的概念.

二、当$g(x)$为其他形式时

当使用定理2.6.1寻求$Y=g(X)$的分布有困难时，可直接由Y的分布函数$F_Y(y)=P(g(X)\le y)$出发，按函数$g(x)$的特点作个案处理

例子

例 2.6.3 设随机变量 X 服从标准正态分布 $N(0,1)$，试求 $Y=X^2$ 的分布.

解 先求 Y 的分布函数 $F_Y(y)$. 由于 $Y=X^2\geqslant 0$，故当 $y\leqslant 0$ 时，有 $F_Y(y)=0$，从而 $p_Y(y)=0$. 当 $y>0$ 时，有

$$F_Y(y)=P(Y\leqslant y)=P(X^2\leqslant y)=P(-\sqrt{y}\leqslant X\leqslant\sqrt{y})=2\Phi(\sqrt{y})-1.$$

因此 Y 的分布函数为

$$F_Y(y)=\begin{cases}2\Phi(\sqrt{y})-1, & y>0,\\ 0, & y\leqslant 0.\end{cases}$$

再用求导的方法求出 Y 的密度函数

$$p_Y(y)=\begin{cases}\varphi(\sqrt{y})\,y^{-\frac{1}{2}}, & y>0,\\ 0, & y\leqslant 0\end{cases}$$

$$=\begin{cases}\dfrac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}\mathrm{e}^{-\frac{y}{2}}, & y>0,\\ 0, & y\leqslant 0.\end{cases}$$

对照 χ^2 分布的密度函数，可以看出 $Y\sim\chi^2(1)$.

习题2.6

习 题 2.6

1. 已知离散随机变量 X 的分布列为

X	-2	-1	0	1	3
P	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{15}$	$\frac{11}{30}$

试求 $Y=X^2$ 与 $Z=|X|$ 的分布列.

2. 已知随机变量 X 的密度函数为

$$p(x)=\frac{2}{\pi}\cdot\frac{1}{e^x+e^{-x}},\qquad -\infty<x<\infty.$$

试求随机变量 $Y=g(X)$ 的概率分布,其中

$$g(x)=\begin{cases}-1, & 当 x<0,\\ 1, & 当 x\geqslant 0.\end{cases}$$

3. 设随机变量 X 服从 $(-1,2)$ 上的均匀分布,记

$$Y=\begin{cases}1, & X\geqslant 0,\\ -1, & X<0.\end{cases}$$

试求 Y 的分布列.

4. 设随机变量 $X\sim U(0,1)$,试求 $1-X$ 的分布.

5. 设随机变量 X 服从 $(-\pi/2,\pi/2)$ 上的均匀分布,求随机变量 $Y=\cos X$ 的密度函数 $p_Y(y)$.

6. 设圆的直径服从区间 $(0,1)$ 上的均匀分布,求圆的面积的密度函数.

7. 设随机变量 X 服从区间 $(1,2)$ 上的均匀分布,试求 $Y=e^{2X}$ 的密度函数.

8. 设随机变量 X 服从区间 $(0,2)$ 上的均匀分布.

(1) 求 $Y=X^2$ 的密度函数;(2) $P(Y<2)$.

10. 设随机变量 X 服从 $(0,1)$ 上的均匀分布,试求以下 Y 的密度函数:

(1) $Y=-2\ln X$;　　(2) $Y=3X+1$;

(3) $Y=e^X$;　　(4) $Y=|\ln X|$.

11. 设随机变量 X 的密度函数为

$$p_X(x)=\begin{cases}\frac{3}{2}x^2, & -1<x<1,\\ 0, & 其他.\end{cases}$$

试求下列随机变量的分布:

(1) $Y_1=3X$;(2) $Y_2=3-X$;(3) $Y_3=X^2$.

12. 设随机变量 $X\sim N(0,\sigma^2)$,求 $Y=X^2$ 的分布.

13. 设随机变量 $X\sim N(\mu,\sigma^2)$,求 $Y=e^X$ 的数学期望与方差.

14. 设随机变量 X 服从标准正态分布 $N(0,1)$,试求以下 Y 的密度函数:

(1) $Y=|X|$;　　(2) $Y=2X^2+1$.

15. 设随机变量 X 的密度函数为

$$p(x)=\begin{cases}e^{-x}, & 若 x>0,\\ 0, & 若 x\leqslant 0.\end{cases}$$

试求以下 Y 的密度函数:

(1) $Y=2X+1$;　　(2) $Y=e^X$;　　(3) $Y=X^2$.

16. 设随机变量 X 服从参数为 2 的指数分布,试证 $Y_1=e^{-2X}$ 和 $Y_2=1-e^{-2X}$ 都服从区间 $(0,1)$ 上的均匀分布.

17. 设随机变量 $X\sim LN(\mu,\sigma^2)$,试证:$Y=\ln X\sim N(\mu,\sigma^2)$.

18. 设随机变量 $Y\sim LN(5,0.12^2)$,试求 $P(Y<188.7)$.

2.7 分布的其他特征数

2.7.1 k阶矩

定义2.7.1

设X为随机变量，k为正整数。如果以下的数学期望都存在，则称$\mu_k = E(X^k)$为X的k阶原点矩，称$v_k = E(X - E(X))^k$为X的k阶中心矩

例子

例 2.7.1 设随机变量 $X \sim N(0, \sigma^2)$，则

$$\mu_k = E(X^k) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x^k \exp\left\{-\frac{x^2}{2\sigma^2}\right\} dx$$

$$= \frac{\sigma^k}{\sqrt{2\pi}}\int_{-\infty}^{\infty} u^k \exp\left\{-\frac{u^2}{2}\right\} du.$$

在 k 为奇数时，上述被积函数是奇函数，故

$$\mu_k = 0, \quad k = 1, 3, 5, \cdots.$$

在 k 为偶数时，上述被积函数是偶函数，再利用变换 $z = u^2/2$，可得

$$\mu_k = \sqrt{\frac{2}{\pi}}\sigma^k 2^{(k-1)/2}\int_0^{\infty} z^{(k-1)/2} e^{-z} dz = \sqrt{\frac{2}{\pi}}\sigma^k 2^{(k-1)/2}\Gamma\left(\frac{k+1}{2}\right)$$

$$= \sigma^k (k-1)(k-3)\cdots 1. \qquad k = 2, 4, 6, \cdots.$$

故 $N(0, \sigma^2)$ 分布的前四阶原点矩为

$$\mu_1 = 0, \quad \mu_2 = \sigma^2, \quad \mu_3 = 0, \quad \mu_4 = 3\sigma^4.$$

又因为 $E(X) = 0$，所以有原点矩等于中心矩，即 $\mu_k = \nu_k, k = 1, 2, \cdots$.

2.7.2 变异系数

- 方差、标准差反映了随机变量取值的波动程度，但只看方差、标准差就会产生不合理现象，原因如下：
  - 随机变量的取值有量纲，不同量纲的随机变量用其方差去比较波动大小不合理
  - 在取值量纲相同的情况下，取值大小有一个相对性问题
- 这就引出了变异系数

定义2.7.2

设随机变量X的二阶矩存在，则称比值$C_v(X) = \frac{\sigma(X)}{E(X)}$为X的变异系数

例子

例 2.7.2 用 X 表示某种同龄树的高度，其量纲是米(m)，用 Y 表示某年龄段儿童的身高，其量纲也是米(m). 设 $E(X) = 10, \mathrm{Var}(X) = 1, E(Y) = 1, \mathrm{Var}(Y) = 0.04$，你是否可以从 $\mathrm{Var}(X) = 1$ 和 $\mathrm{Var}(Y) = 0.04$ 就认为 Y 的波动小？这就有一个取值相对大小的问题. 在此用变异系数进行比较是恰当的. 因为 X 的变异系数为

$$C_v(X) = \frac{\sigma(X)}{E(X)} = \frac{1}{10} = 0.1,$$

而 Y 的变异系数为

$$C_v(Y) = \frac{\sigma(Y)}{E(Y)} = \frac{\sqrt{0.04}}{1} = 0.2,$$

这说明 Y(儿童身高)的波动比 X(同龄树高)的波动大.

2.7.3 分位数

定义2.7.3

设随机变量X的分布函数为$F(x)$，密度函数为$p(x)$。对任意$p \in (0,1)$，称满足条件$F(x_p) = \int_{-\infty}^{x_p} p(x)dx = p$的$x_p$为此分布的p分位数，又称下侧p分位数

同理称满足条件$1 - F(x_p') = \int_{x_p'}^{\infty} p(x)dx = p$的$x_p'$为此分布的上测分位数

如图

图 2.7.1(b)).

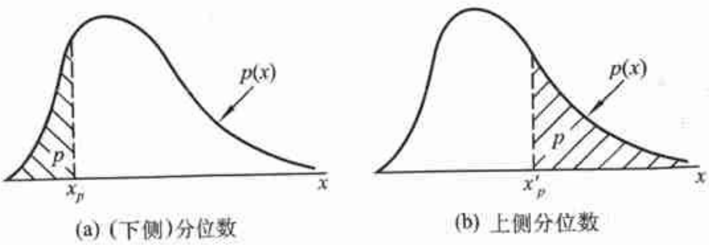


图 2.7.1 分位数与上侧分位数的区别

2.7.4 中位数

定义2.7.4

设连续随机变量X的分布函数为$F(x)$，密度函数为$p(x)$。称p=0.5时的p分位数$x_{0.5}$为此分布的中位数

如图

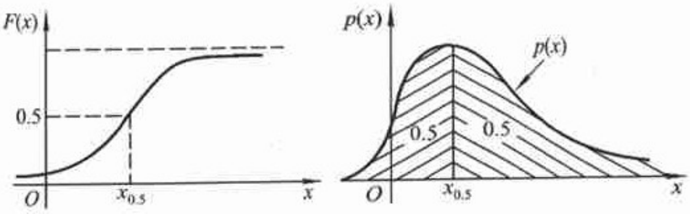


图 2.7.2 连续随机变量的中位数

例子

例 2.7.6 设连续随机变量 X 的密度函数为

$$p(x) = \begin{cases} 4x^3, & 0 < x < 1, \\ 0, & \text{其他}. \end{cases}$$

试求此分布的 0.95 分位数 $x_{0.95}$ 和中位数 $x_{0.5}$.

解 因为 X 的分布函数为

$$F(x) = \begin{cases} 0, & x < 0, \\ x^4, & 0 \leqslant x < 1, \\ 1 & 1 \leqslant x. \end{cases}$$

所以由 $F(x_{0.95}) = 0.95$ 可得：$x_{0.95}^4 = 0.95$，由此得 $x_{0.95} = \sqrt[4]{0.95} = 0.9873$. 同理由 $F(x_{0.5}) = 0.5$ 得 $x_{0.5}^4 = 0.5$，从中解得 $x_{0.5} = \sqrt[4]{0.5} = 0.8409$.

2.7.5 偏度系数

设随机变量X的前三阶矩存在，则如下比值$\beta_s=\frac{v_3}{v_2^{3/2}}=\frac{E(X-E(X))^3}{[Var(X)]^{3/2}}$为X的偏度系数

当$\beta_s<0$时，称为左偏，如图

偏度β_s是描述分布偏离对称性程度的一个特征数

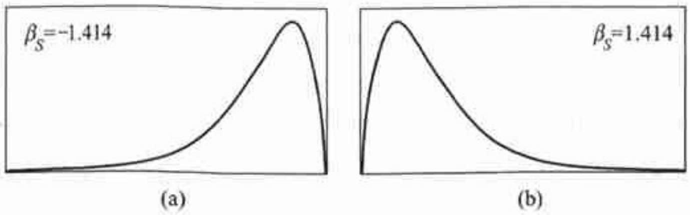


图 2.7.3 两个密度函数,一个左偏,另一个右偏

例子

例 2.7.7 讨论三个贝塔分布 $Be(2,8)$, $Be(8,2)$ 和 $Be(5,5)$ 的偏度.

解 设随机变量 X 服从贝塔分布 $Be(a,b)$,则可算得其前三阶原点矩:

$$E(X)=\frac{a}{a+b},$$

$$E(X^2)=\frac{a(a+1)}{(a+b)(a+b+1)},$$

$$E(X^3)=\frac{a(a+1)(a+2)}{(a+b)(a+b+1)(a+b+2)}.$$

以下为简化 β_s 的计算,特用 $a+b=10, b=10-a$ 代入可算得前三阶中心矩.

$$E(X)=\frac{a}{10},\quad \mathrm{Var}(X)=\frac{a(10-a)}{10^2\times 11},\quad E(X-EX)^3=\frac{a(10-a)(5-a)}{10^3\times 11\times 3}.$$

可得贝塔分布的偏度为

$$\beta_s=\frac{E(X-EX)^3}{[\mathrm{Var}(X)]^{3/2}}=\frac{\sqrt{11}(5-a)}{3\sqrt{a(10-a)}},$$

把 $a=2,5,8$ 分别代入可得

$Be(2,8)$ 的 $\beta_s=\sqrt{11}/4=0.829\ 2$,右偏(正偏)

$Be(5,5)$ 的 $\beta_s=0$,对称

$Be(8,2)$ 的 $\beta_s=-\sqrt{11}/4=-0.829\ 2$,左偏(负偏).

2.7.6 峰度系数

定义2.7.6

设随机变量X的前四阶矩存在，则称比值$\beta_k=\frac{v_4}{v_2^2}-3=\frac{E(X-E(X))^4}{[Var(x)]^2}-3$称为X的峰度系数

峰度是描述分布尖峭程度和尾部粗细的一个特征数，如图

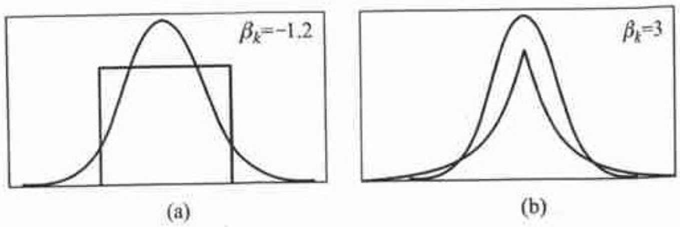


图 2.7.4　两个密度函数与标准正态分布密度函数的比较

它们的均值相等、方差相等、偏度皆为0(对称分布)，而峰度有很大差别

- 偏度和峰度都是描述分布形状的特征数

例子

例 2.7.8　计算伽玛分布 $Ga(\alpha,\lambda)$ 的偏度与峰度.

解　首先计算伽玛分布 $Ga(\alpha,\lambda)$ 的 k 阶原点矩：

$$\mu_k=E(X^k)=\alpha(\alpha+1)\cdots(\alpha+k-1)/\lambda^k.$$

当 $k=1,2,3,4$ 时可得前四阶原点矩

$$\mu_1=\alpha/\lambda,$$
$$\mu_2=\alpha(\alpha+1)/\lambda^2,$$
$$\mu_3=\alpha(\alpha+1)(\alpha+2)/\lambda^3,$$
$$\mu_4=\alpha(\alpha+1)(\alpha+2)(\alpha+3)/\lambda^4.$$

由此可得 2、3、4 阶中心矩

$$\nu_2=\mu_2-\mu_1^2=\alpha/\lambda^2,$$
$$\nu_3=\mu_3-3\mu_2\mu_1+2\mu_1^3=2\alpha/\lambda^3,$$
$$\nu_4=\mu_4-4\mu_3\mu_1+6\mu_2\mu_1^2-3\mu_1^4=3\alpha(\alpha+2)/\lambda^4.$$

最后可得伽玛分布 $Ga(\alpha,\lambda)$ 的偏度与峰度

$$\beta_S=\frac{\nu_3}{\nu_2^{3/2}}=\frac{2}{\sqrt{\alpha}},$$

$$\beta_k=\frac{\nu_4}{\nu_2^2}-3=\frac{6}{\alpha}.$$

可见，伽玛分布 $Ga(\alpha,\lambda)$ 的偏度与$\sqrt{\alpha}$成反比. 峰度与 α 成反比. 只要 α 较大，可使 β_S 与 β_k 接近于0，从而伽玛分布也愈来愈近似正态分布.

习题2.7

习 题 2.7

1. 设随机变量 $X\sim U(a,b)$，对 $k=1,2,3,4$，求 $\mu_k=E(X^k)$ 与 $\nu_k=E(X-E(X))^k$. 进一步求

此分布的偏度系数和峰度系数.

2. 设随机变量 $X\sim U(0,a)$，求此分布的变异系数.

3. 求以下分布的中位数：

(1) 区间 (a,b) 上的均匀分布；

(2) 正态分布 $N(\mu,\sigma^2)$；

(3) 对数正态分布 $LN(\mu,\sigma^2)$.

4. 设随机变量 $X\sim Ga(\alpha,\lambda)$，对 $k=1,2,3$，求 $\mu_k=E(X^k)$ 与 $\nu_k=E(X-E(X))^k$.

5. 设随机变量 $X\sim Exp(\lambda)$，对 $k=1,2,3,4$，求 $\mu_k=E(X^k)$ 与 $\nu_k=E(X-E(X))^k$. 进一步求此分布的变异系数、偏度系数和峰度系数.

6. 设随机变量 X 服从正态分布 $N(10,9)$，试求 $x_{0.1}$ 和 $x_{0.9}$.

7. 设随机变量 X 服从双参数韦布尔分布，其分布函数为

$$F(x)=1-\exp\left\{-\left(\frac{x}{\eta}\right)^m\right\},\qquad x>0,$$

其中 $\eta>0,m>0$. 试写出该分布的 p 分位数 x_p 的表达式，且求出当 $m=1.5,\eta=1\ 000$ 时的 $x_{0.1}$，$x_{0.5}$，$x_{0.8}$ 的值.

8. 自由度为 2 的 χ^2 分布的密度函数为

$$p(x)=\frac{1}{2}e^{-\frac{x}{2}},\qquad x>0.$$

试求出其分布函数及分位数 $x_{0.1}$，$x_{0.5}$，$x_{0.8}$.

9. 设随机变量 X 的分布密度函数 $p(x)$ 关于 c 点是对称的，且 $E(X)$ 存在，试证：

(1) 这个对称点 c 既是均值又是中位数，即 $E(X)=x_{0.5}=c$；

(2) 如果 $c=0$，则 $x_p=-x_{1-p}$.

10. 试证随机变量 X 的偏度系数与峰度系数对位移和改变比例尺是不变的，即对任意的实数 $a,b(b\neq 0)$，$Y=a+bX$ 与 X 有相同的偏度系数与峰度系数.

11. 设某项维修时间 T（单位：分）服从对数正态分布 $LN(\mu,\sigma^2)$.

(1) 求 p 分位数 t_p；

(2) 若 $\mu=4.127\ 1$，求该分布的中位数；

(3) 若 $\mu=4.127\ 1,\sigma=1.036\ 4$，求完成 95% 维修任务的时间.

12. 某种绝缘材料的使用寿命 T（单位：小时）服从对数正态分布 $LN(\mu,\sigma^2)$. 若已知分位数 $t_{0.2}=5\ 000$ 小时，$t_{0.8}=65\ 000$ 小时，求 μ 和 σ.

13. 某厂决定按过去生产状况对月生产额最高的 5% 的工人发放高产奖. 已知过去每人每月生产额 X（单位：千克）服从正态分布 $N(4\ 000,60^2)$，试问高产奖发放标准应把生产额定为多少？