主讲老师：张宇

概率论与数理统计

第一讲　随机事件与概率

一、基本概念

1. 随机试验

我们称一个试验为**随机试验**，如果它满足以下三个条件：

(1) 试验可以在相同的条件下重复进行；

(2) 试验所有可能结果是明确可知道的，并且不止一个；

(3) 每一次试验会出现哪一个结果，事先不能确定.

我们是通过研究随机试验来研究随机现象的，为方便起见，将随机试验简称为**试验**，并用字母 E 或 $E_1,E_2,\cdots$ 表示.

【注】 在不少情况下，我们不能确切知道某一随机试验的全部可能结果，但可以知道它不超出某个范围. 这时，也可以用这个范围来作为该试验的全部可能结果. 例如，我们需要记录某个城市一天的交通事故数量，则试验结果将是非负数 x. 我们无法确定 x 的可能取值的确切范围，但可以把这范围取为 $[0,+\infty)$，它总能包含一切可能的试验结果，尽管我们明知某些结果，如 $x>10\,000$ 是不会出现的. 我们甚至可以把这范围取为 $(-\infty,+\infty)$ 也无妨. 这里就有了一定的数学抽象，它可以带来很大的方便.

2. 随机事件

在一次试验中可能出现，也可能不出现的结果称为**随机事件**，简称为**事件**，并用大写字母 A,B,C 等表示. 为讨论需要，将每次试验一定发生的事件称为**必然事件**，记为 Ω. 每次试验一定不发生的事件称为**不可能事件**，记为 $\varnothing$.

【注】 随机事件在一次试验中是否发生虽然不能事先确定，但是在大量重复试验的情况下，它的发生呈现出一定的规律性，这门学问正是要研究这种规律性，读者应在研究这门课程后，对此有较为深刻的认识.

3. 样本空间

随机试验每一个最简单、最基本的结果称为**基本事件**（或**样本点**），记为 ω. 基本事件

(或样本点)的全体称为**基本事件空间**(或**样本空间**),记为Ω,即$\Omega=\{\omega\}$. 随机事件A总是由若干个基本事件组成,即A是Ω的子集,$A\subset\Omega$. 事件A发生等价于构成A的基本事件有一个发生.

4. 事件的关系与运算

(1) 如果事件A发生必导致事件B发生,则称事件B**包含**事件A(或A被B包含),记为$A\subset B$.

(2) 如果$A\subset B$且$B\subset A$,则称事件A与B**相等**,记为$A=B$. A与B相等,事实上也就是说,A与B由完全同样的一些试验结果构成,它不过是同一事件表面上看来不同的两个说法而已.

(3) 称"事件A与B同时发生"的事件为事件A与B的**交(或积)**,记为$A\cap B$或AB.

【注】 称"有限个(或可列个)事件$A_1,A_2,\cdots,A_n(\cdots)$同时发生"的事件为事件$A_1$,$A_2,\cdots,A_n(\cdots)$的交(或积),记为$\bigcap\limits_{i=1}^{n}A_i$(或$\bigcap\limits_{i=1}^{\infty}A_i$).

(4) 若$AB\neq\varnothing$,则称事件"A和B**相容**";若$AB=\varnothing$,则称"事件A与B**互不相容**",也叫**互斥**. 如果一些事件中任意两个事件都互斥,则称这些事件是**两两互斥**的,或简称**互斥**的.

(5) 称"事件A与B至少有一个发生"的事件为事件A与B的**并(或和)**,记为$A\cup B$.

【注】 称"有限个(或可列个)事件$A_1,A_2,\cdots,A_n(\cdots)$至少有一个发生"的事件为事件$A_1,A_2,\cdots,A_n(\cdots)$的并(或和),记为$\bigcup\limits_{i=1}^{n}A_i$(或$\bigcup\limits_{i=1}^{\infty}A_i$).

(6) 称"事件A发生而事件B不发生"的事件为事件A与B的**差**,记为$A-B$;称"事件A不发生"的事件为事件A的**逆事件**或**对立事件**,记为$\overline{A}$.

由定义易知　$A-B=A-AB=A\overline{B}$,　$B=\overline{A}\Leftrightarrow AB=\varnothing$且$A\cup B=\Omega$.

(7) 称有限个(或可列个)事件$A_1,A_2,\cdots,A_n(\cdots)$构成**一个完备事件组**,如果$\bigcup\limits_{i=1}^{n}A_i$(或$\bigcup\limits_{i=1}^{\infty}A_i$)$=\Omega$,$A_iA_j=\varnothing$(对一切$i\neq j$).

事件的关系和运算法则:

(1) 吸收律　若$A\subset B$,则$A\cup B=B$,$A\cap B=A$;

(2) 交换律　$A\cup B=B\cup A$,$A\cap B=B\cap A$;

(3) 结合律　$(A\cup B)\cup C=A\cup(B\cup C)$,$(A\cap B)\cap C=A\cap(B\cap C)$;

(4) 分配律　$A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$,$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$,$A\cap(B-C)=(A\cap B)-(A\cap C)$;

(5) 对偶律(德摩根律)　$\overline{A\cup B}=\overline{A}\cap\overline{B}$,$\overline{A\cap B}=\overline{A}\cup\overline{B}$.

【注】 ① 事件运算顺序约定为先进行逆运算，而后交运算，最后并或差运算.

② 事件的关系、运算与集合的关系、运算相当，且具有相同的运算法则，所以我们可以对比着理解记忆，并要学会用集合关系去思考事件关系.

二、用古典概型求概率

1. 定义

称随机试验(随机现象)的概率模型为**古典概型**，如果其基本事件空间(样本空间)满足：

(1) 只有有限个基本事件(样本点)；

(2) 每个基本事件(样本点)发生的可能性都一样.

如果古典概型的基本事件总数为n，事件A包含k个基本事件，也叫作有利于A的基本事件为k个，则A的概率定义为

$$P(A)=\frac{k}{n}=\frac{\text{事件}A\text{所含基本事件的个数}}{\text{基本事件总数}}.$$

由上式计算的概率称为A的**古典概率**.

2. 方法与例题

计数方法：

(1) 穷举法：个数不多时，直接数数即可.

(2) 集合对应法：

① 加法原理——完成一件事有n类方法，第一类方法中有m_1种办法，第二类方法中有m_2种方法，…，第n类方法中有m_n种办法，则完成此事共有$\sum\limits_{i=1}^{n}m_i$种办法.

② 乘法原理——完成一件事有n个步骤.

第一步有m_1种方法，第二步有m_2种方法，…，第n步有m_n种办法，则完成此事共有$\prod\limits_{i=1}^{n}m_i$种办法.

③ 排列——从n个不同的元素中取出$m(\leqslant n)$个元素，并按照一定顺序排成一列，叫排列. 所有排列的个数叫排列数. 记作$\mathrm{P}_n^m=n(n-1)(n-2)\cdots(n-m+1)=\frac{n!}{(n-m)!}$

当$m=n$时，$\mathrm{P}_n^n=\frac{n!}{0!}=n!$叫全排列.

④ 组合——从n个不同的元素中取出$m(\leqslant n)$个元素，并成一组，叫组合. 所有组合

的个数叫组合数. 记作

$$C_n^m = \frac{P_n^m}{m!}$$

(3) 对立事件思想 —— 若研究 A 复杂,则转而研究 $\overline{A}$,用 $n - n_{\overline{A}} = n_A$(总数一易算出).

【例 1】袋中 5 球,3 白 2 黑

(1) 从袋中先后有放回取 2 球.

(2) 先后无放回取 2 球.

(3) 任取 2 球.

求 $P(A) = P\{至少一白\}$.

【分析】

【例 2】袋中 100 个球,40 个白 60 黑.

① 先后无放回取 20 个,求 $P\{15\text{ 白 }5\text{ 黑}\}$;

② 先后无放回取 20 个,求 $P\{第\ 20\ 次取到白\}$;

③ 先后有放回取 20 个,求 $P\{15\text{ 白 }5\text{ 黑}\}$;

④ 先后有放回取 20 个,求 $P\{第\ 20\ 次取到白\}$.

【分析】

三、用几何概型求概率

1. 引例与定义

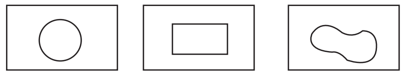
假设明天早上 9:00,天上会掉一个馅饼(当作质点)到你所在学校的操场上,请你用食堂的饭盆去接(如下图).

两个问题:一是你站在哪里接呢?二是你要带多大的饭盆和什么形状的饭盆呢?认真思考后,不难得出如下结论.

第一个问题的答案是:站在何处去接馅饼都可以.这里就要求读者懂得一个重要思想——**等可能性思想——我们没有任何理由认为这个馅饼更有可能落在操场区域中的某个位置,只好认为它落在此区域中的任何位置都具有相等的可能性**.这个思想在很多问题中都有重要应用,比如说:一个袋子中有 10 个同质球,其中有 1 个白球,9 个黑球,现在请你从该袋子中随机取出一球,问取出的球是白球的概率.我们会毫不犹豫地回答:十分之一.这里用到的就是这个思想:由于球是同质的,我们没有任何理由认为取到某一个球更具有可能性,只好认为取到 10 个球中任何一球都具有等可能性,所以取到白球的概率自然是十分之一.

第二个问题的答案是:你所带的饭盆大小至关重要,但是饭盆是什么形状却无关紧要.为什么?因为馅饼落在操场区域的任何位置都具有等可能性,于是,读者容易想到,如果饭盆的面积是操场面积的千分之一,那么你接到馅饼的概率就是千分之一;如果饭盆的面积是操场面积的二分之一,那么你接到馅饼的概率也就是二分之一;如果,我是说如果,你的饭盆面积和操场面积一样大,那么毫无疑问,馅饼就是你的了.至于你是用圆形、还是矩形、甚至是奇形怪状的(如下图),真是随你了(这里要注意一点:如果你的饭盆面积和操场面积一样大,你的饭盆和操场形状自然是一样的,这是特殊情况).这里又可以

把第一个问题所述的等可能性等价地，或者说更加专业地描述为：**馅饼落到操场区域的任意子区域上的概率与该子区域的面积大小成正比.**



以上所述涉及**求概率的几何概型**这一重要知识点.

于是，在本书中你会看到如下定义与公式.

称随机试验(随机现象)的概率模型为**几何概型**，如果：

(1) 样本空间(基本事件空间)Ω 是一个可度量的几何区域；

(2) 每个样本点(基本事件)发生的可能性都一样，即样本点落入 Ω 的某一可度量的子区域 S 的可能性大小与 S 的几何度量成正比，而与 S 的位置及形状无关.

在几何概型随机试验中，如果 S_A 是样本空间 Ω 的一个可度量的子区域，则事件 $A=\{$样本点落入区域 $S_A\}$ 的概率定义为

$$P(A)=\frac{S_A\text{ 的几何度量}}{\Omega\text{ 的几何度量}}.$$

由上式计算的概率称为 A 的**几何概率**.

需要指出的是，上述问题是考研重点，是命题人手里的"香饽饽".

2. 方法

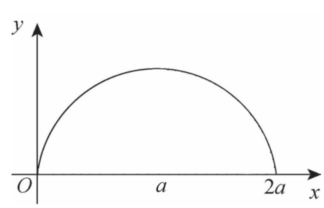
几何概型的特点是总样本数和事件所含样本的点数不能如古典概型那样可以数数，而具有几何特性，如长度、面积或体积等，解题时应将数量关系用几何直观图形表现出来，其概率通常是事件所含样本数和总样本数对应区域大小之比. 前言中对此有详细叙述，可参考之.

3. 例题

【例】[取自《张宇概率论与数理统计 9 讲》P15，例 1.18]

随机地向半圆 $0<y<\sqrt{2ax-x^2}\,(a>0)$ 内投掷一点，点均匀落在半圆内任何一个区域，求该点和原点连线同 x 轴夹角 $\theta\leqslant\frac{\pi}{4}$ 的概率.

【分析】



四、用重要公式求概率

1. 重要公式

(1) 逆事件概率公式

对于任一事件 A，有 $P(\overline{A}) = 1 - P(A)$.

(2) 加法公式

对于任意两个事件 A, B，有 $P(A \cup B) = P(A) + P(B) - P(AB)$.

【注】 上式还能推广到多个事件的情况.

① 设 A_1, A_2, A_3 为任意三个事件，则有

$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3)$.

② 对于任意 n 个事件 $A_1, A_2, \cdots, A_n$，可以用数学归纳法证得

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leqslant i < j \leqslant n} P(A_iA_j) + \sum_{1 \leqslant i < j < k \leqslant n} P(A_iA_jA_k) + \cdots + (-1)^{n-1} P(A_1A_2\cdots A_n).$$

(3) 减法公式

$P(A - B) = P(A) - P(AB) = P(A\overline{B})$.

(4) 条件概率公式

设 A, B 为任意两个事件，若 $P(A) > 0$，我们称在已知事件 A 发生的条件下，事件 B 发生的概率为**条件概率**，记为 $P(B \mid A)$，并定义

$$P(B \mid A) = \frac{P(AB)}{P(A)}.$$

【注】 ① 条件概率 $P(\cdot \mid A)$ 是概率，概率的一切性质和重要结果对条件概率都适用.

例如：
$$P(\overline{B} \mid A) = 1 - P(B \mid A),$$
$$P(B - C \mid A) = P(B \mid A) - P(BC \mid A)$$

等等.

② 条件概率就是在附加了一定的条件之下所计算的概率. 当说到“条件概率”时，总是指另外附加的条件，其形式可归结为“已知某事件发生了”.

(5) 乘法公式

如果 $P(A) > 0$，则 $P(AB) = P(A)P(B \mid A)$.

一般地，如果 $P(A_1\cdots A_{n-1}) > 0$，则 $P(A_1A_2\cdots A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid$

$A_1A_2)\cdots P(A_n \mid A_1\cdots A_{n-1})$.

(6) 全概率公式

如果 $\bigcup\limits_{i=1}^{n} A_i = \Omega, A_iA_j = \varnothing (i \neq j; i,j = 1,2,\cdots,n), P(A_i) > 0$,则对任一事件 B,有

$$B = \bigcup_{i=1}^{n} A_iB,\quad P(B) = \sum_{i=1}^{n} P(A_i)P(B \mid A_i).$$

(7) 贝叶斯公式(又称逆概公式)

如果 $\bigcup\limits_{i=1}^{n} A_i = \Omega, A_iA_j = \varnothing (i \neq j), P(A_i) > 0$,则对任一事件 B,只要 $P(B) > 0$,就有

$$P(A_j \mid B) = \frac{P(A_j)P(B \mid A_j)}{\sum\limits_{i=1}^{n} P(A_i)P(B \mid A_i)} (i,j = 1,2,\cdots,n).$$

【注】 ① 要注意 $P(B)$ 与 $P(B \mid A)$ 的区别与联系,虽然二者都是计算事件 B 的概率,但前者 $P(B)$ 实际上是指在样本空间 Ω 下计算的,后者 $P(B \mid A)$ 则是在 A 已经发生的条件下(即样本空间现在缩减至 A) 计算的.

② 全概率公式是用于计算某个“结果”B 发生的可能性大小. 如果一个结果 B 的发生总是与某些前提条件(或原因、因素或前一阶段结果)A_i 相联系,那么在计算 $P(B)$ 时,我们总是用 A_i 对 B 作分解:

$$B = \bigcup_{i=1}^{\infty} A_iB,$$

应用全概率公式计算 $P(B)$,我们常称这种方法为**全集分解法**. 如果在 B 发生的条件下探求导致这一结果的各种“原因”A_i 发生的可能性大小 $P(A_i \mid B)$,则要应用贝叶斯公式.

(8) 关于事件的独立性

① **描述性定义(直观性定义)** 设 A,B 为两个事件,如果其中任何一个事件发生的概率不受另外一个事件发生与否的影响,则称事件 A 与 B **相互独立**. 设 $A_1,A_2,\cdots,A_n$ 是 n 个事件,如果其中任何一个或几个事件发生的概率都不受其余的某一个或几个事件发生与否的影响,则称事件 $A_1,A_2,\cdots,A_n$ 相互独立.

② **数学定义** 设 A,B 为事件,如果 $P(AB) = P(A)P(B)$,则称事件 A 与 B **相互独立**,简称为 A 与 B **独立**.

【注】 设 $A_1,A_2,\cdots,A_n$ 为 $n(n \geqslant 2)$ 个事件,如果对其中任意有限个事件 $A_{i_1},A_{i_2},\cdots,A_{i_k}(k \geqslant 2)$,有

$$P(A_{i_1}A_{i_2}\cdots A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k}),$$

则称 n 个事件 $A_1,A_2,\cdots,A_n$ 相互独立.

考研中常考的是 $n=3$ 的情形.细致说来,设 A_1,A_2,A_3 为 3 个事件,若

$$P(A_1A_2)=P(A_1)P(A_2), \tag{1.1}$$

$$P(A_1A_3)=P(A_1)P(A_3), \tag{1.2}$$

$$P(A_2A_3)=P(A_2)P(A_3), \tag{1.3}$$

$$P(A_1A_2A_3)=P(A_1)P(A_2)P(A_3), \tag{1.4}$$

则称 A_1,A_2,A_3 **相互独立**.当去掉上述(1.4)式后,称只满足(1.1)(1.2)(1.3)的 A_1,A_2,A_3 **两两独立**.

2. 例题

【例 1】[取自《张宇概率论与数理统计 9 讲》P10,例 1.9]

(1)设事件 A,B 仅发生一个的概率为 0.3,且 $P(A)+P(B)=0.5$.求 A,B 至少有一个不发生的概率;

(2)设 X,Y 为随机变量,且 $P\{X\geqslant 0,Y\geqslant 0\}=\dfrac{3}{7}$,$P\{X\geqslant 0\}=P\{Y\geqslant 0\}=\dfrac{4}{7}$.试求下列事件的概率:

$A=\{\max\{X,Y\}\geqslant 0\}$; $B=\{\max\{X,Y\}<0,\min\{X,Y\}<0\}$; $C=\{\max\{X,Y\}\geqslant 0,\min\{X,Y\}<0\}$.

【分析】

【例 2】[取自《张宇概率论与数理统计 9 讲》$P17$,例 1.26]

设有两箱同种零件,第一箱内装 50 件,其中 10 件一等品;第二箱内装 30 件,其中 18 件一等品.先从两箱中随机挑出一箱,然后从该箱中先后随机取出两个零件(取出的零件均不放回).试求:

(1) 先取出的零件是一等品的概率 p;

(2) 在先取出的是一等品的条件下,后取出的零件仍然是一等品的条件概率 q.

【分析】

【例 3】[取自《张宇概率论与数理统计 9 讲》$P21$,例 1.27]

设有两批数量相同的零件,已知有一批产品全部合格,另一批产品有 25% 不合格.从两批产品中任取 1 只,经检验是正品,放回原处,并在原所在批次再取 1 只,试求这只产品是次品的概率.

【分析】

第二讲　一维随机变量及其分布

一、基本概念

1. 随机变量的概念

随机变量就是“其值会随机而定”的变量. 设随机试验E的样本空间为$\Omega=\{w\}$，如果对每一个$w\in\Omega$，都有唯一的实数$X(w)$与之对应，并且对任意实数x，$\{w\,|\,X(w)\leqslant x, w\in\Omega\}$是随机事件，则称定义在$\Omega$上的实单值函数$X(w)$为**随机变量**. 简记为随机变量$X$. 一般用大写字母$X,Y,Z,\cdots$或希腊字母$\xi,\eta,\zeta,\cdots$来表示随机变量.

【注】 (1) 随机事件是从静态的观点来研究随机现象，而随机变量则是一种动态的观点，一如高等数学中常量与变量的区别与联系.

(2) 随机变量的实质是“实单值函数”，这个定义不同于高等数学中函数的定义(其“定义域”一定是实数集)，它的“定义域”不一定是实数集，这点需要读者注意.

2. 分布函数的概念及性质

(1) **概念**

设X是随机变量，x是任意实数，称函数$F(x)=P\{X\leqslant x\}(x\in\mathbf{R})$为随机变量$X$的分布函数，或称$X$服从分布$F(x)$，记为$X\sim F(x)$.

【注】 分布函数完整地描述了随机变量的概率规律性.

(2) **性质(也是充要条件)**

①$F(x)$是x的单调不减函数，即对任意$x_1<x_2$，有$F(x_1)\leqslant F(x_2)$；

②$F(x)$是x的右连续函数，即对任意$x_0\in\mathbf{R}$，有$\lim\limits_{x\to x_0^+}F(x)=F(x_0+0)=F(x_0)$；

③$F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0$，$F(+\infty)=\lim\limits_{x\to+\infty}F(x)=1$.

【注】 (1) 务必记住分布函数是事件的概率，由此知$0\leqslant F(x)\leqslant 1$，即$F(x)$是有界函数.

(2) 满足以上三条性质的函数$F(x)$必是某个随机变量的分布函数，所以，这三条性质也是判断某一函数$F(x)$是否为某一随机变量X的分布函数的充要条件.

二、常见的两类随机变量 —— 离散型随机变量和连续型随机变量

1. 离散型随机变量及其概率分布

如果随机变量 X 只可能取有限个或可列个值 $x_1, x_2, \cdots$，则称 X 为**离散型随机变量**，称

$$p_i = P\{X = x_i\}, i = 1, 2, \cdots$$

为 X 的**分布列**、**分布律**或**概率分布**，记为 $X \sim p_i$，概率分布常常用表格形式或矩阵形式表示，即

X	x_1	x_2	$\cdots$
P	p_1	p_2	$\cdots$

或 $X \sim \begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}$.

数列$\{p_i\}$是离散型随机变量的概率分布的充要条件是：$p_i \geqslant 0 (i = 1, 2, \cdots)$，且$\sum\limits_i p_i = 1$.

设离散型随机变量 X 的概率分布为 $p_i = P\{X = x_i\}$，则 X 的分布函数

$$F(x) = P\{X \leqslant x\} = \sum_{x_i \leqslant x} P\{X = x_i\},$$

$$p_i = P\{X = x_i\} = P\{X \leqslant x_i\} - P\{X < x_i\} = F(x_i) - F(x_i - 0),$$

并且对实数轴上的任一集合 B 有

$$P\{X \in B\} = \sum_{x_i \in B} P\{X = x_i\}.$$

特别是 $\qquad P\{a < X \leqslant b\} = P\{X \leqslant b\} - P\{X \leqslant a\} = F(b) - F(a).$

2. 连续型随机变量及其概率密度

对这种变量的概率分布，不能用像离散型变量那种方法去描述. 原因在于，这种变量的取值充满一个区间，无法一一排出，若指定一个值 a，则变量 X 恰好是 a 一丝不差，事实上不可能. 如在称量误差的例中，如果你认定天平上的读数（刻度）是“无限精细”，则“误差正好为 $\pi - 3$”虽原则上不能排除，但可能性也极微，以至于只能取为 0；如在靶面上指定一个几何意义下的点（即只有位置而无任何向度），则“射击时正好命中该点”的概率也只能取为 0.

那么，该如何刻画连续型随机变量呢？

如果随机变量 X 的分布函数可以表示为

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t (x \in \mathbf{R}),$$

其中 $f(x)$ 是非负可积函数，则称 X 为**连续型随机变量**，称 $f(x)$ 为 X 的**概率密度函数**，简

称为**概率密度**或**密度函数**，记为 $X \sim f(x)$.

$f(x)$ 为某一随机变量 X 的概率密度的充分必要条件是：$f(x) \geqslant 0$，且 $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$（由此可知，改变 $f(x)$ 有限个点的值后，$f(x)$ 仍然是概率密度）.

设 X 为连续型随机变量，$X \sim f(x)$，则对任意实数 c 有 $P\{X = c\} = 0$；对实数轴上任一集合 B 有

$$P\{X \in B\} = \int_B f(x)\mathrm{d}x.$$

特别是

$$P\{a < X < b\} = P\{a \leqslant X < b\} = P\{a < X \leqslant b\} = P\{a \leqslant X \leqslant b\} = \int_a^b f(x)\mathrm{d}x = F(b) - F(a).$$

【注】 (1)“概率密度”这名词的由来可解释如下：取定一个点 x，则按分布函数的定义，事件 $\{x < X \leqslant x+h\}$ 的概率（$h > 0$，为常数），应为 $F(x+h) - F(x)$，所以，比值 $[F(x+h) - F(x)]/h$ 可以解释为在 x 点附近 h 这么长的区间 $(x, x+h]$ 内，单位长所占有的概率. 令 $h \to 0$，则这个比的极限，即 $F'(x) = f(x)$，也就是在 x 点处（无穷小区间段内）单位长的概率，或者说，它反映了概率在 x 点处的“密集程度”. 你可以设想一条极细的无穷长的金属杆，总质量为 1，概率密度相当于杆上各点的质量密度.

(2) $P\{a < X < b\} = \int_a^b f(x)\mathrm{d}x$ 意味着 X 落入某一区间的概率等于该区间之上、概率密度之下曲边梯形的面积，应用概率的这种几何意义，常常有助于问题的分析与求解.

(3) 设 $X \sim f(x)$，则 X 的分布函数 $F(x)$ 是 x 的连续函数；在 $f(x)$ 的连续点 x_0 处有 $F'(x_0) = f(x_0)$；如果 $F(x)$ 是连续函数，除有限个点外，$F'(x)$ 存在且连续，则 X 为连续型随机变量，且 $f(x) = F'(x)$（在 $F'(x)$ 不存在的地方可以令 $f(x) = 0$ 或取其他值）.

三、常见的随机变量分布类型

1. 离散型

(1) 0－1 分布 $B(1, p)$

如果 X 的概率分布为 $X \sim \begin{pmatrix} 1 & 0 \\ p & 1-p \end{pmatrix}$，即 $P\{X = 1\} = p$，$P\{X = 0\} = 1 - p$，则称 X 服从参数为 p 的 **0－1 分布**，记为 $X \sim B(1, p)$ $(0 < p < 1)$.

(2) 二项分布 $B(n, p)$

如果 X 的概率分布为 $P\{X = k\} = C_n^k p^k (1-p)^{n-k}$ $(k = 0, 1, \cdots, n; 0 < p < 1)$，则称 X

服从参数为(n,p)的**二项分布**，记为$X\sim B(n,p)$.

【注】 如果X是n重伯努利试验中事件A发生的次数，则$X\sim B(n,p)$，其中$p=P(A)$. 这个结论在解题中我们会经常用到.

(3) 泊松分布$P(\lambda)$

如果X的概率分布为$P\{X=k\}=\dfrac{\lambda^k}{k!}e^{-\lambda}(k=0,1,\cdots;\lambda>0)$，则称$X$服从参数为$\lambda$的**泊松分布**，记为$X\sim P(\lambda)$.

(4) 几何分布$G(p)$

如果X的概率分布为$P\{X=k\}=q^{k-1}p\ (k=1,2,\cdots;0<p<1;q=1-p)$，则称$X$服从参数为$p$的**几何分布**，记为$X\sim G(p)$.

【注】 设X表示伯努利试验中事件A首次发生所需做的试验次数，则$X\sim G(p)$，其中$p=P(A)$，2015年数学一、数学三的第(22)题中求数学期望，就是在对这个背景做深入的考查.

(5) 超几何分布$H(n,N,M)$

如果X的概率分布为$P\{X=k\}=\dfrac{C_M^kC_{N-M}^{n-k}}{C_N^n}\ (k=0,1,\cdots,\min\{M,n\};M,N,n$为正整数且$M\leqslant N,n\leqslant N)$，则称$X$服从参数为$(n,N,M)$的**超几何分布**，记为$X\sim H(n,N,M)$.

2. 连续型

(1) 均匀分布$U(a,b)$

如果X的概率密度或分布函数分别为

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b,\\ 0, & \text{其他},\end{cases}\text{或 }F(x)=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\leqslant x<b,\\ 1, & x\geqslant b,\end{cases}$$

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

【注】 区间(a,b)可以是闭区间$[a,b]$；几何概型是均匀分布的实际背景，用几何概率计算事件概率时已假设点在区域内服从均匀分布；几何概率可以用均匀分布计算.

(2) 指数分布$E(\lambda)$

如果X的概率密度或分布函数分别为

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0,\\ 0, & \text{其他},\end{cases}\text{或 }F(x)=\begin{cases}1-e^{-\lambda x}, & x\geqslant 0,\\ 0, & x<0\end{cases}(\lambda>0),$$

则称X服从参数为λ的**指数分布**，记为$X\sim E(\lambda)$.

(3) 正态分布 $N(\mu,\sigma^2)$

如果 X 的概率密度

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

其中 $-\infty<\mu<+\infty,\sigma>0$,则称 X 服从参数为 (μ,σ^2) 的**正态分布**或称 X 为**正态变量**,记为 $X\sim N(\mu,\sigma^2)$. 此时 $f(x)$ 的图形关于直线 $x=\mu$ 对称,即 $f(\mu-x)=f(\mu+x)$,并在 $x=\mu$ 处有唯一最大值 $f(\mu)=\dfrac{1}{\sqrt{2\pi}\sigma}$.

称 $\mu=0,\sigma=1$ 时的正态分布 $N(0,1)$ 为**标准正态分布**,通常记标准正态分布的概率密度为 $\varphi(x)=\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}$,分布函数为 $\Phi(x)$. 显然 $\varphi(x)$ 为偶函数,$\Phi(0)=\dfrac{1}{2}$,$\Phi(-x)=1-\Phi(x)$.

若 $X\sim N(0,1)$,$P\{X\geqslant\mu_\alpha\}=\alpha$,则称 μ_α 为标准正态分布的**上侧 α 分位数.**

如果 $X\sim N(\mu,\sigma^2)$,则

$$F(x)=P\{X\leqslant x\}=\Phi\left(\frac{x-\mu}{\sigma}\right),$$

$$F(\mu-x)+F(\mu+x)=1,$$

$$P\{a<X<b\}=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right),$$

$$aX+b\sim N(a\mu+b,a^2\sigma^2)(a\neq 0).$$

四、例题

【例 1】[取自《张宇概率论与数理统计 9 讲》P36,例 2.6]

设 $f_1(x)$ 为标准正态分布的概率密度,$f_2(x)$ 为 $[-1,3]$ 上均匀分布的概率密度,若

$$f(x)=\begin{cases}af_1(x), & x\leqslant 0,\\ bf_2(x), & x>0\end{cases}\ (a>0,b>0)$$

为概率密度,则 a,b 应满足(　　).

(A) $2a+3b=4$　　(B) $3a+2b=4$　　(C) $a+b=1$　　(D) $a+b=2$

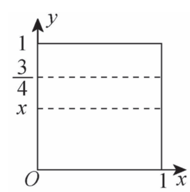
【分析】

【例 2】[取自《张宇概率论与数理统计 9 讲》$P48$，例 2.29]

一电路装有三个同种电气元件，其工作状态相互独立，且无故障工作时间服从参数为 $\lambda > 0$ 的指数分布. 当三个元件都无故障工作时，线路工作状态正常，试求电路正常工作时间 T 的概率分布.

【分析】

【例 3】[取自《张宇概率论与数理统计 9 讲》$P50$，例 2.32]



一水渠出口闸门挡板是边长为 1(单位) 的正方形. 已知初始水面高为 $\frac{3}{4}$(单位)，现发现挡板某一部位出现一个小孔(小孔等可能出现在挡板的任一位置)，水经过小孔流出，求剩余液面高度 X 的分布函数 $F(x)$.

【分析】

五、随机变量函数的分布

1. 概念及分布

(1) 概念

设 X 为随机变量，函数 $y=g(x)$，则以随机变量 X 作为自变量的函数 $Y=g(X)$ 也是随机变量，称之为**随机变量 X 的函数**. 例如：$Y=aX^2+bX+c$，$Y=|X-a|$，$Y=\begin{cases}X, & X\leqslant 1,\\ 1, & X>1\end{cases}$ 等等.

(2) 随机变量函数的分布

1) 离散型

设 X 为离散型随机变量，其概率分布为 $p_i=P\{X=x_i\}(i=1,2,\cdots)$，则 X 的函数 $Y=g(X)$ 也是离散型随机变量，其概率分布为 $P\{Y=g(x_i)\}=p_i$，即

$$Y\sim\begin{pmatrix}g(x_1) & g(x_2) & \cdots\\ p_1 & p_2 & \cdots\end{pmatrix}.$$

如果有若干个 $g(x_k)$ 相同，则合并诸项为一项 $g(x_k)$，并将相应概率相加作为 Y 取 $g(x_k)$ 值的概率.

2) 连续型

设 X 为连续型随机变量，其分布函数、概率密度分别为 $F_X(x)$ 与 $f_X(x)$，随机变量 $Y=g(X)$ 是 X 的函数，则 Y 的分布函数或概率密度可用下面两种方法求得：

定义法(分布函数法)

直接由定义求 Y 的分布函数

$$F_Y(y)=P\{Y\leqslant y\}=P\{g(X)\leqslant y\}=\int_{g(x)\leqslant y}f(x)\mathrm{d}x.$$

如果 $F_Y(y)$ 连续，且除有限个点外，$F'_Y(y)$ 存在且连续，则 Y 的概率密度 $f_Y(y)=F'_Y(y)$.

2. 例题

【例 1】[取自《张宇概率论与数理统计 9 讲》P57，例 3.1]

设 X 是仅可能取 6 个值的离散型随机变量，其分布如下：

X	-2	-1	0	1	2	3
P	0.05	0.15	0.20	0.25	0.20	0.15

求 $Y=2X+1, Z=X^2$ 的概率分布.

【分析】

【例 2】[取自《张宇概率论与数理统计 9 讲》$P58$，例 3.5]

设随机变量 X 的概率密度为 $f_X(x)=\dfrac{k}{1+x^2}\ (x\in\mathbf{R})$. 求：

(1) 常数 k；

(2) 随机变量 $Y=1-\sqrt[3]{X}$ 的概率密度 $f_Y(y)$.

【分析】

第三讲　多维随机变量及其分布

一、基本概念

① 二维 $r.v.(X,Y)$；

② 联合分布函数 $F(x,y) \overset{\Delta}{=} P\{X \leqslant x, Y \leqslant y\}$；

③ 边缘分布函数，若 $(X,Y) \sim F(x,y)$，则

$F_X(x) \overset{\Delta}{=} P\{X \leqslant x\} = P\{X \leqslant x, Y < +\infty\} = F(x,+\infty) = \lim\limits_{y \to +\infty} F(x,y)$；

$F_Y(y) \overset{\Delta}{=} P\{Y \leqslant y\} = P\{X < +\infty, Y \leqslant y\} = F(+\infty, y) = \lim\limits_{x \to +\infty} F(x,y)$；

④ 独立性 $(X,Y) \sim F(x,y)$，

若 $F(x,y) = F_X(x)F_Y(y) \Leftrightarrow$ 则 X,Y 互相独立；

⑤ 离散型 $(X,Y) \sim P_{ij}$ 联合分布律

X \ Y	y_1	y_2	$\cdots$	y_j	$\cdots$	
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}		$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	

其中：$P(X = x_i) = P_{i\cdot}$，$P(Y = y_j) = P_{\cdot j}$，且 $P_{i\cdot} \cdot P_{\cdot j} = P_{ij}$，$\forall i,j \Leftrightarrow X,Y$ 独立.

【注】借鉴 $P(A \mid B) = \dfrac{P(AB)}{P(B)}$，引入“条件分布律”：

$$P(X = x_i \mid Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{P_{ij}}{P_{\cdot j}};$$

$$P(Y = y_j \mid X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{P_{ij}}{P_{i\cdot}}.$$

⑥ 连续型 $r.v.(X,Y)\sim f(x,y)$ 联合概率密度

$$\text{边缘}\begin{cases} f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x \end{cases}$$

独立性:$f_X(x)f_Y(y)=f(x,y)\Leftrightarrow X,Y$ 独立.

条件概率密度:$f_{X|Y}(x\mid y)=\dfrac{f(x,y)}{f_Y(y)}(f_Y(y)>0)f_{Y|X}(y\mid x)=\dfrac{f(x,y)}{f_X(x)}(f_X(x)>0)$.

二、例题

【例 1】[取自《张宇概率论与数理统计 9 讲》P73,例 4.1]

设 X 与 Y 相互独立,下表是随机变量 X 与 Y 的联合分布律以及两个边缘分布律中的部分数值,试将其余数值填入表中空白处.

$X \backslash Y$	y_1	y_2	y_3	$P\{X=x_i\}$
x_1		$\frac{1}{8}$		
x_2	$\frac{1}{8}$			
$P\{Y=y_j\}$	$\frac{1}{6}$			1

【分析】

【例 2】[取自《张宇概率论与数理统计 9 讲》$P78$,例 4.9]

已知二维随机变量(X,Y)在G上服从均匀分布,G由直线$x-y=0$,$x+y=2$与$y=0$围成,求:

(1) 边缘概率密度$f_X(x)$和$f_Y(y)$;

(2) 条件概率密度$f_{X|Y}(x \mid y)$.

【分析】

第四讲　数字特征

一、基本概念

1. 数学期望

①$X \sim p_i \Rightarrow EX = \sum\limits_i x_i p_i$

如 $X \sim \begin{pmatrix} 1 & 2 & 5 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$，则 $EX = 1 \times \frac{1}{2} + 2 \times \frac{1}{3} + 5 \times \frac{1}{6}$，

$X \sim p_i$，且 $Y = g(X) \rightarrow EY = \sum\limits_i g(x_i) p_i$；

②$X \sim f(x) \Rightarrow EX = \int_{-\infty}^{+\infty} x f(x) dx$，

$X \sim f(x)$，且 $Y = g(X) \Rightarrow EY = \int_{-\infty}^{+\infty} g(x) f(x) dx$.

2. 方差

$DX \overset{\Delta}{=} E(X - EX)^2$.

(1) 定义法：$Y = (X - 数)^2 = g(X)$，

$$\begin{cases} 1) X \sim p_i \Rightarrow DX = E(X-EX)^2 = EY = \sum\limits_i (x_i - EX)^2 p_i; \\ 2) X \sim f(x) \Rightarrow DX = E(X-EX)^2 = EY = \int_{-\infty}^{+\infty} (x - EX)^2 f(x) dx. \end{cases}$$

(2) 公式法

$$\begin{aligned} DX &= E(X-EX)^2 = E[X^2 - 2X \cdot EX + (EX)^2] \\ &= EX^2 + E[-2X \cdot EX] + E[(EX)^2] \\ &= EX^2 + (-2EX) \cdot EX + (EX)^2 = EX^2 - (EX)^2. \end{aligned}$$

3. 协方差 $\mathrm{Cov}(X,Y)$

设 $DX, DY > 0$，称 $\mathrm{Cov}(X,Y) = E(X-EX)(Y-EY)$ 为协方差.

（若 $X = Y$，$\mathrm{Cov}(X,X) = E(X-EX)(X-EX) = DX$）

$\mathrm{Cov}(X,Y) = E(XY - X\cdot EY - Y\cdot EX + EXEY) = E(XY) - EX\cdot EY$.

4. 相关系数

称 $\rho_{XY} = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{DX}\ \sqrt{DY}}$ 为相关系数. 描述 X,Y 之间线性相关程度.

二、例题

【例 1】[取自《张宇概率论与数理统计 9 讲》P132，例 6.20]

设 A,B 为两个随机事件，且 $P(A) = \dfrac{1}{4}$，$P(B \mid A) = \dfrac{1}{3}$，$P(A \mid B) = \dfrac{1}{2}$.

令
$$X = \begin{cases} 1, & A\text{ 发生}, \\ 0, & A\text{ 不发生}, \end{cases} \quad Y = \begin{cases} 1, & B\text{ 发生}, \\ 0, & B\text{ 不发生}, \end{cases}$$

求：(1) 二维随机变量 (X,Y) 的概率分布；

(2) X,Y 的相关系数 ρ_{XY}；

(3) $Z = X^2 + Y^2$ 的概率分布.

【分析】

第五讲　数理统计初步

一、基本概念

1. 总体与样本

(1) 总体:研究对象的某个指标的全体 $X\sim F$;

(2) 样本:简单随机样本 $X_i\overset{iid}{\sim}F$.

2. 估计方法

(1) 矩估计　$EX\xlongequal[\text{令}]{\text{人为}}\overline{X}$.

【例 1】[取自《张宇概率论与数理统计 9 讲》$P187$,例 9.1]

设来自总体 X 的简单随机样本 $X_1,X_2,\cdots,X_n$,总体 X 的概率分布为

$$X\sim\begin{pmatrix}-1 & 0 & 2\\ 2\theta & \theta & 1-3\theta\end{pmatrix},$$

其中 $0<\theta<\dfrac{1}{3}$. 试求未知参数 θ 的矩估计量.

【分析】

【例 2】[取自《张宇概率论与数理统计 9 讲》$P187$,例 9.2]

已知总体 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{6x}{\theta^3}(\theta-x), & 0<x<\theta,\\ 0, & \text{其他},\end{cases}$$

$X_1,X_2,\cdots,X_n$ 为来自总体 X 的一个简单随机样本,求参数 θ 的矩估计量 $\hat{\theta}$;

【分析】

（2）最大似然估计

方法

【例 3】［取自《张宇概率论与数理统计 9 讲》$P188$，例 9.5］

设来自总体 X 的简单随机样本 $X_1,X_2,\cdots,X_n$，总体 X 的概率分布为

$$X \sim \begin{pmatrix} 1 & 2 & 3 \\ \theta^2 & 2\theta(1-\theta) & (1-\theta)^2 \end{pmatrix},$$

其中 $0<\theta<1$. 分别以 ν_1,ν_2 表示 $X_1,X_2,\cdots,X_n$ 中 1,2 出现的次数，试求

（1）未知参数 θ 的最大似然估计量；

（2）未知参数 θ 的矩估计量；

（3）当样本值为 1,1,2,1,3,2 时的最大似然估计值和矩估计值.