简明概率论讲义（安徽大学 2022）

第一章 随机事件及其概率

1.1 随机事件及其运算

$$\text{自然界中的现象}\begin{cases}\text{必然（确定性）现象}\\ \text{随机（偶然性）现象}\begin{cases}\text{个别随机现象}\\ \text{大量随机现象}\end{cases}\end{cases}$$

必然现象--------在一定的条件下，一定会出现（发生）的现象 .
随机现象--------在一定的条件下，可能出现也可能不出现的现象 .

注：① 在概率论中，“出现”与“发生”同义。概率论主要研究大量随机现象（统计规律性），但是也不排斥个别随机现象 .

$$\text{随机性的作用}\begin{cases}\text{很小以致可以忽略}\\ \text{基本的，如布朗运动}\end{cases}$$

② 随机现象中的“不确定性”有两层含义，一则指客观结果的不确定性；一则指主观猜测的不确定性，后者融入了观察者个人的信念.

统计规律性--------大量随机现象所具有的规律性，概率论主要研究此 .

对于随机现象，即使条件完全相同，它们的出现所产生的结果也不尽相同，此之谓“现象的随机性”；那么随机性产生的原因是什么呢？

任何随机现象都是相互联系和相互影响的，它的行为受许多因素地支配或制约；而控制所有这些因素原则上做不到，往往只限于决定该现象状态的最基本因素 .并且除此之外还有许多时隐时现，转瞬即逝，无法控制的偶然因素；当随机现象重复出现时，这些因素产生的效应是不同的，不确定的，从而使得现象带有随机性 .

试验--------凡对现象的观察或为此而进行的实验 .
决定性试验--------凡对决定性（必然）现象的观察或为此而进行的实验 .
随机试验--------凡对随机现象的观察或为此而进行的实验，常记为 E 或 E_i .

事件（试验的结局）---------试验观察的结果 .

无论何种试验，都包含两个方面，即试验的条件和试验的结果 .随机试验的条件有的是“人为的”，如在一定的条件下观察“射击是否命中目标”；有的是“不依人的意志为转移的”，如花粉微粒的无规则运动（布朗运动）.

随机试验的共同特点为：
① 在相同的条件下可重复进行；
② 每次试验的结果可能不止一个，但事先明确所有可能的结果；
③ 试验之前不能确定那个结果会出现 .

称随机试验的最基本结果为样本点；称随机试验的所有样本点的集合为样本空间；样本空间从集合的角度可划分为有限（穷）样本空间、可列（数）（无穷）样本空间和不可列（数）样本空间（举例说明）.

事件的运算

一　事件的并运算：　设 A, B 为两事件，则“ A, B 至少一个发生”这种情况可能出现也可能不出现，其作为一个随机事件，称之为 A, B 的并（事件），记作：$A \cup B$；

若 $A_1, A_2, \cdots, A_n$ 均为事件，则“ $A_1, A_2, \cdots, A_n$ 至少一个发生”作为事件，称之为 $A_1, A_2, \cdots, A_n$ 的并事件，记作：$A_1 \cup A_2 \cup \cdots \cup A_n$，简记为：$\bigcup\limits_{i=1}^{n} A_i$；

若 $A_1, A_2, \cdots, A_n, \cdots$ 为一列事件，则“ $A_1, A_2, \cdots, A_n, \cdots$ 至少一个发生”作为事件，称之为 $A_1, A_2, \cdots, A_n, \cdots$ 的并事件，记作：$A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots$，简记为：$\bigcup\limits_{n=1}^{\infty} A_n$（可列并）.

二　事件的交运算：设 A, B 为两事件，则“ A, B 同时发生”这种情况可能出现也可能不出现，其作为一个随机事件，称之为 A, B 的交（积）（事件），记作：$A \cap B$；

若 $A_1, A_2, \cdots, A_n$ 均为事件，则“ $A_1, A_2, \cdots, A_n$ 同时发生”作为事件，称之为 $A_1, A_2, \cdots, A_n$ 的交事件，记作：$A_1 \cap A_2 \cap \cdots \cap A_n$，简记为：$\bigcap\limits_{i=1}^{n} A_i$；

若 $A_1, A_2, \cdots, A_n, \cdots$ 为一列事件，则“ $A_1, A_2, \cdots, A_n, \cdots$ 同时发生”作为事件，称之为 $A_1, A_2, \cdots, A_n, \cdots$ 的交事件，记作：$A_1 \cap A_2 \cap \cdots \cap A_n \cap \cdots$，简记为：$\bigcap\limits_{n=1}^{\infty} A_n$（可列交）.

注：若 A, B 两事件不可能同时发生，则称 A, B 互不相容（互斥），记作：$A \cap B = \Phi$；若 $A_1, A_2, \cdots, A_n$ 互不相容（两两互不相容），则又称 $A_1, A_2, \cdots, A_n$ 的并为 $A_1, A_2, \cdots, A_n$ 的和，记作：$\sum\limits_{i=1}^{n} A_i$；即有：$\bigcup\limits_{i=1}^{n} A_i = \sum\limits_{i=1}^{n} A_i$；若 $A_1, A_2, \cdots, A_n, \cdots$ 互不相容（两两互不相容），则又称 $A_1, A_2, \cdots, A_n, \cdots$ 的并（可列并）为 $A_1, A_2, \cdots, A_n, \cdots$ 的和，记作：$\sum\limits_{n=1}^{\infty} A_n$；即有：

$$\bigcup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} A_n \quad .$$

三　事件的差运算：设 A, B 为两事件，则“ A 发生而 B 不发生”这种情况可能出现也可能不出现，其作为一个随机事件，称之为 A, B 的差事件，记作：$A - B(A / B)$.

利用差运算，可以将“有交并”转化为“不交并”，如：

$$A \cup B = (A - B) \cup B = (A - B) + B \left[(B - A) \cup A = (B - A) + A \right] \ .$$

四　事件的逆（余）运算：设 A 为事件，则“ A 不发生”这种情况可能出现也可能不出现，其作为一个随机事件，称之为 A 的逆（余）（事件），记作：$\overline{A}\left(A^c\right)$.

差积转化律：　　　　$A - B = A\overline{B}$　　　　　　　　.

五　事件列的极限（运算）：规定：$\bigcup_{n=1}^{\infty} A_n = \lim_{n\to\infty} \bigcup_{i=1}^{n} A_i$ ，$\bigcap_{n=1}^{\infty} A_n = \lim_{n\to\infty} \bigcap_{i=1}^{n} A_i$.

依此若事件列$\{A_n, n \geq 1\}$，

(a) 单调递增，即：$A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots$；则定义事件列的极限为

$$\lim_{n\to\infty} A_n = \lim_{n\to\infty} \bigcup_{i=1}^{n} A_i = \bigcup_{n=1}^{\infty} A_n .$$

(b) 单调递减，即：$A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$；则定义事件列的极限为

$$\lim_{n\to\infty} A_n = \lim_{n\to\infty} \bigcap_{i=1}^{n} A_i = \bigcap_{n=1}^{\infty} A_n .$$

对于任意的随机事件列$\{A_n, n \geq 1\}$，令，

① $B_n = \bigcup_{k=n}^{\infty} A_k, n \geq 1$,则$\{B_n, n \geq 1\}$单调递减，从而有：$\lim_{n\to\infty} B_n = \lim_{n\to\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ ；

称此极限为$\{A_n, n \geq 1\}$的上极限，记作：$\limsup_{n\to\infty} A_n$，即有：$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$.

② $C_n = \bigcap_{k=n}^{\infty} A_k, n \geq 1$，则$\{C_n, n \geq 1\}$单调递增，从而有：$\lim_{n\to\infty} C_n = \lim_{n\to\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$ ；

称此极限为$\{A_n, n \geq 1\}$的下极限，记作：$\liminf_{n\to\infty} A_n$，即有：$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$.

③ 若$\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n$，则称此极限为$\{A_n, n \geq 1\}$的极限，也即：

$$\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = \lim_{n\to\infty} A_n .$$

例 1 在某系任选一名学生，A＝“被选学生是男生”，B＝{被选学生是三年级学生},C＝{被选学生是运动员}，

1）试述$AB\overline{C}$的意义； 2）在何情况下，$ABC = C$？ 3）在何情况下，$\overline{A} = B$？

例 2 若A表示“产品甲畅销且产品乙滞销”，试述 $\overline{A}$.

1.2　随机事件的概率

几何学中，线段的长度、平面（空间）图形的面积（体积）、物理学中物质的某些量等都可用数值来度量；我们观察一个随机试验的诸事件，总发现“有些事件出现的可能性大些，有些事件出现的可能性小些”，而这些“可能性大小”自然也可用数值来度量，这是由于事件出现的“可能性大小”是客观存在的。这个刻划事件发生可能性大小的数值至少应该满足两个要求：

① 它具有一定的客观性，不能随意改变，而且理论上应可通过在“相同条件下”大量的重复试验加以识别和检验；
② 它必须符合一般常理，如：事件发生可能性大（小）的，这个值就应该大（小）些；必然事件的值最大为1，不可能事件的值最小为0 .

概率的一般（直观）定义：刻划（随机）事件发生（出现）可能性大小的数值（数量指标），又称或然率或几（机）率，它介于0与1之间 .

（事件）σ 域（代数）：设 Ω 是某试验 E 的样本空间，F 是由 Ω 的一些子集所成的集族（类），若 F 满足：

① $\Omega \in F$ ； ② 若 $A \in F$，则 $\overline{A} \in F$ ；

③ 若 $\forall n \geq 1,\ A_n \in F$，也即 $\{A_n, n \geq 1\} \subset F$，则 $\bigcup_{n=1}^{\infty} A_n \in F$ ；

则称 F 为（事件）σ 域（代数），F 中的集称为事件（F 可测集）.

易证：F 关于有限并（交）、可列交、差、上（下）极限等诸多（**Borel**）运算均封闭。

可测空间：将任一样本空间 Ω 以及建立在其上由其若干子集组成的 σ 域（代数）F 组成的二元体 (Ω, F) 称为具有 σ －域（代数）结构的样本空间或简称可测空间 .

概率（函数，测度）与概率空间（场）：设 (Ω, F) 为一可测空间，对每一集合（事件）$A \in F$，定义实值集（合）函数 $P(A)$，使之满足：

①（非负性）对每一 $A \in F$，有 $0 \leq P(A) \leq 1$ ；

②（规范性）对必然事件 Ω，有 $P(\Omega) = 1$ ；

③（可列可加性）若 $\forall n \geq 1, A_n \in F$，且 $A_i A_j = \Phi$， 则

$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ ；称 $P(\cdot)$ 为定义于 (Ω, F) 上的概率测度（函数），简称概率；并称三元体 (Ω, F, P) 为概率空间 .

由概率的公理化定义可导出概率的如下性质：

1. $P(\Phi) = 0$.

2．可列可加性 $\Rightarrow$ 有限可加性；由性质 **2** 可得如下常用公式：

① $A \in F, P(\overline{A}) = 1 - P(A)$.

② 若 $A, B \in F$，则 $P(A - B) = P(A) - P(AB)$.

特别地，若 $A \supset B$，则有 $P(A - B) = P(A) - P(B)$（减法公式或可减性）；由此也可得，若 $A \supset B$，则 $P(A) \geq P(B)$（概率的单调性）.

3．概率的连续性：

①【下连续性】若 $A_n \in F, n \geq 1,$ 且 $A_n \uparrow$ ，则有

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\lim_{n\to\infty} \bigcup_{i=1}^{n} A_i\right) = P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n);$$

注：所谓“连续性”，即是指可以交换极限运算与概率运算的次序 .

②【上连续性】若 $A_n \in F, n \geq 1,$ 且 $A_n \downarrow$，则有

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = P\left(\lim_{n\to\infty} \bigcap_{i=1}^{n} A_i\right) = P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n) .$$

③【连续性】若 $A_n \in F, n \geq 1$, 且 $\lim\limits_{n\to\infty} A_n$ 存在，则有

$$P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P\left(A_n\right) .$$

注：由上述，不难验证：可列可加性 $\Leftrightarrow$ 连续性＋有限可加性 .

4．加法公式（多除少补原理）－**Jordan**（若尔当）公式：

设 $A_i \in F, i \geq 1$，则 $P\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = S_1 - S_2 + S_3 - \cdots + \left(-1\right)^{n-1} S_n$ ，其中：

$$S_1 = \sum_{i=1}^{n} P(A_i)，\ S_2 = \sum_{1\leq i\leq j\leq n} P(A_i A_j)，S_3 = \sum_{1\leq i\leq j\leq k\leq n} P(A_i A_j A_k)，\cdots, S_n = P\left(A_1 A_2 \cdots A_n\right) .$$

特别地，若 $A, B, C \in F$，则有：

① $P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) - P\left(AB\right)$ ；

② $P\left(A \cup B \cup C\right) = P\left(A\right) + P\left(B\right) + P\left(C\right) - P\left(AB\right) - P\left(AC\right) - P\left(BC\right) + P\left(ABC\right)$.

5. 次可加性：设 $A_i \in F, i \geq 1$ ，

①（次有限可加性）$P\left(\bigcup\limits_{i=1}^{n} A_i\right) \leq \sum\limits_{i=1}^{n} P(A_i)$.②（次可列可加性）$P\left(\bigcup\limits_{i=1}^{\infty} A_i\right) \leq \sum\limits_{i=1}^{\infty} P(A_i)$.

例 3 设 $P\left(A\right) = 0.4, P\left(B\right) = 0.5, P\left(A \cup B\right) = 0.7$ ，试求 $P\left(A - B\right)$ 和 $P\left(B - A\right)$.

例 4 设 $A, B \in F$ ，$P\left(A\right) = 0.5, P\left(B\right) = 0.7$ ，试求 $P\left(AB\right)$ 和 $P\left(A \cup B\right)$ 的最值 .

例 5 设 $A, B, C \in F$ ，且 $P\left(A\right) = P\left(B\right) = P\left(C\right) = \dfrac{1}{4}$ ，$P\left(AB\right) = 0$，$P\left(AC\right) = P\left(BC\right)$ $= \dfrac{1}{6}$，试求“ A, B, C 全不发生”的概率 .

例 6 给定一概率空间 $\left(\Omega, F, P\right)$，$A, B \in F$，且 $P\left(A\right) = P\left(B\right) = 1$（几乎必然事件），试求：$P\left(AB\right), P\left(A\overline{B}\right), P\left(A \cup B\right), P\left(A \cup \overline{B}\right)$.

例 7 试证明概率不等式：$\left|P(AB) - P(A)P(B)\right| \leq \dfrac{1}{4}$.

例 8 ① 若 $AB = \overline{A}\overline{B}$，且 $P\left(A\right) = a$，则 $P\left(B\right)$ 为多少 .

② 已知 $P\left(AB\right) = P\left(\overline{A}\overline{B}\right)$，且 $P\left(A\right) = a$，求 $P\left(B\right)$.

在概率论的发展史上，人们曾针对不同的问题，从不同的角度给出了定义概率和计算概率的各种方法；然而这些“定义的概率”都存在一定的缺陷——与其称它们为概率的定义，弗如称它们为计算概率的方法 .下面我们将陆续介绍之 .

以下介绍两种等可能概型（概率模型），先考虑一种特殊的随机试验，譬如掷一枚（质地均匀）硬币，人们自然会想到硬币正反两面对称，故出现正面与反面应是一样，从而有理由认为出现正面和反面的可能性都是0.5；在其他一些例子中，人们得出类似的结论也多是由于人们利用了研究对象的物理均衡性或者几何对称性 .

古典概型（等可能概型 I）：古典概率—计算概率的古典方法-----Laplace

由于它的简单，曾是历史上最早被研究的一类概型，归纳它的特点如下：如果一个随机试验的所有可能结果只有有限个，而且每个结果的出现都是等可能的，则称这个随机试验是古典概型 .

设有古典概型 E ,以 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$ 表示它的样本空间；则对于任意的事件 A，若它恰好包含其中的 m 个样本点，则称事件 A（发生）的（古典）概率（即在古典概型背景下计算概率的古典方法）为 $\frac{m}{n}$,记作：$P(A)=\frac{m}{n}$；也即在古典概型下，由古典方法计算出的事件 A 的（古典）概率为 $P(A)=\frac{|A|}{|\Omega|}$，这里 $|A|,|\Omega|$ 分别表示 A,Ω 所包含的样本点数 .

.

例 9（错例分析）掷两枚硬币，求“出现一正一反”的概率 .

例 10 ① 从 $1,2,\cdots,n$ 中任取两数，求 A =｛ 两数之和为偶数 ｝的概率 .

② 从 **N** 中任取两数，求 A =“两数之和为偶数”的概率 .

例 11 袋中有 a 只黑球和 b 只白球，将球随机一一取出，取后不放回，试球求 A =｛ 第 k（$1\le k\le a+b$）次取出黑球 ｝ 的概率（四种方法）.

“古典概率”有如下性质：

①【非负性】设是 A 古典概型中任一事件，则 $0\le P(A)\le 1$.

②【规范性】（又称规一性或正则性）对必然事件 Ω，$P(\Omega)=1$.

③【有限可加性】设同一古典概型中的事件 $A_1,A_2,\cdots,A_n$ 互不相容，则

$$P\left(\bigcup_{i=1}^{n}A_i\right)=P\left(\sum_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i) .$$

对于古典概型，“样本点出现的等可能性”是个基本假设； 在实际问题中，往往并不知道这个假设是否.满足，而只凭主观对物理均衡性质，几何对称性质来判断是不完全确切的通常，人们认为“判断事件发生的可能性大小最可靠的办法”即是通过大量重复试验；特别是当样本点不可能判定为等可能出现时尤其要采取这个方法，这便有了“统计概率”的称谓 .

统计概率－计算概率的统计（频率）方法

频数：在相同的条件下进行 n 次试验，在这 n 次试验中 A 发生的次数 n_A，即为事件 A 的频数 .

频率：比值称为 n_A/n 事件发生的频率，记之为 $f_n(A)=n_A/n$.

频率具有如下特点：

① 频率的大小能体现事件发生的可能性大小，频率大（小）的发生的可能性应该大（小）些；
② 频率具有一定的随机波动性；
③ 随着重复试验次数的增加，频率呈现出相对的稳定性 .

事实上，频率具有稳定性这一事实，说明了刻划事件发生可能性大小的概率具有一定

的客观存在性，而且频率的稳定性也正反映出大量随机现象的统计规律性 .

基于人们不自觉的共识，人们常用频率 $f(A)$ 作为事件 A 的概率的一种量度，这样计算的概率称为统计概率，即为计算概率的统计方法；不妨将 $f(A)$ 写作 $P(A)$，从而“统计概率”满足如下性质：

①【非负性】设 A 是任一事件，则 $0 \le P(A) \le 1$.

②【规范性】$P(\Omega)=1$.

③【有限可加性】设事件 $A_1, A_2, \cdots, A_n$ 互不相容，则 $P\left(\bigcup_{i=1}^{n} A_i\right)=P\left(\sum_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i)$.

“统计概率”同样也有理论和应用上的缺陷性，我们没有理由可以这样认为：

① 取试验次数为 $n+1$ 来计算频率总会比取试验次数为 n 来计算频率更准确；

② 实际情况是，我们不知道试验次数究竟取多大；甚至很难保证：在多次（大量）重复试验时，每次试验的条件完全一样 .

几何概型（等可能概型 II）：几何概率－计算概率的几何方法

在概率论的发展早期，人们就已注意到只考虑随机现象的可能结果只有有穷个是不够的，还需考虑无穷个的情形；事实上，当试验的可能结果无穷多时，当然不能简单地通过样本点的计数来计算概率；如

【引例】在区间 $(0,1)$ 内任取两个数，求事件 A ={ 两数之和小于 $6/5$ }的概率 .

归纳这类例子的共同特点，即可以通过空间集合的几何度量（如：长度、面积、体积等）来计算概率 .

几何概型（概率）： 设试验 E 的样本空间为某可度量的几何区域 Ω，且 Ω 中任一子区域（事件）出现的可能性大小与该区域的几何度量成正比，而与该区域的位置和形状无关，则称试验 E 为几何概型 .若 A 是 Ω 中一区域，且 A 可度量，则定义事件 A 的概率

$$P(A)=\frac{\mu(A)}{\mu(\Omega)} ,$$

这里，$\mu(A), \mu(\Omega)$ 分别表示 A, Ω 的几何度量；其中若 Ω 是一维的，二维的或三维的，那么 Ω 的几何度量分别是长度、面积或体积，称这样定义的概率为几何概率（即计算概率的几何方法）.

由定义，计算概率的几何方法和古典方法类似，也是由一个比值来刻划，只是前者是后者的推广 .

注：求解几何概型归纳起来一般有如下关联的四个步骤：

① 明确问题的实质，即是否为几何概型；

② 明确等可能性的几何元素－任何一个几何概型其样本点都可归纳为具有某种等可能性的几何元素；

③ 用几何区域（如；区间，平面区域，空间区域等）来表示样本点数的总和；

④ 利用初等几何或微积分知识求出样本空间 Ω 的几何区域的几何度量 $\mu(\Omega)$ 和随机事件 A 的几何区域的几何度量 $\mu(A)$，最终由几何方法得到 $P(A)=\dfrac{\mu(A)}{\mu(\Omega)}$,即为 A 的（几

何）概率 .

通常几何概型大体可分为两类;

(a) 样本空间具有明显的几何意义，这类问题结构简单，易于求得;

(b) 样本空间所对应的几何区域没有直接给出,这类问题比较复杂;一般可从代数方法入手，引入适当变量，利用代数和几何的联系找出几何区域，再依几何方法计算概率 .

例 **12**（会面问题）甲乙两人相约 7 点到 8 点的某个时刻在某地会面，先到者等 20 分钟，过时即离去，试求**{** 两人能会面 **}**的概率 .

例 **13**（贝特朗**[Betrand]**奇论）在一半径为 r 的圆内“任意”地作一弦，试求“此弦长度 l 大于圆内接正三角形边长 $\sqrt{3}\ r$ ”的概率 .

例 **14**（几乎不可能事件）向 $[0,1]$ 区间随机地投掷一点，令 A ＝**{** 点落在“$1/2$ ”处 **}**，A_n ＝**{** 点落在 $\left(\frac{1}{2}-\frac{1}{n},\frac{1}{2}+\frac{1}{n}\right)$ **}** ，$n \geq 2$ ，试求 $P(A)$.

“几何概率”具有如下性质：

①【非负性】设 A 是几何概型中任一事件，则 $0 \leq P(A) \leq 1$.

②【规范性】对必然事件 Ω，有 $P(\Omega)=1$.

③【可列可加性】(又称完全可加性或 σ －可加性）设同一几何概型中一列事件 $\{A_n, n \geq 1\}$ 互不相容，则 $P\left(\bigcup_{i=1}^{\infty} A_i\right)=P\left(\sum_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i)$.

几何概型提出了许多新课题；比如：第一、在几何概型中只有样本空间（可测集）的可测（可度量）子集才能计算概率，因此只有可测子集才能称为事件，不可测子集不能称为事件，它们不是概率论的研究对象；第二、在几何概型中，概率为 **0** 的事件（即零概率事件）未必是不可能事件，概率为 **1** 的事件未必是必然事件等等 .

主观概率－计算概率的主观方法

在现实生活中，有些随机现象是不能重复或不能大量重复的，此时该如何确定概率呢？统计界的贝叶斯学派认为：一个事件的概率是人们根据经验对该事件发生的可能性所给出的个人信念，这样给出的概率称为“主观概率” .

如：用主观方法确定“大学生考试作弊”的概率 .

注：① 主观概率不同于主观臆造，前者要求当事人对所考察的事件有丰富的经验和透彻的了解，并能对现在和历史信息进行仔细分析 .

② 用主观方法得出的可能性大小本质上是对随机事件概率的一种推断和估计，虽然结论有待检验和修正，但可信性在统计上是有价值的；况且在遇到随机现象无法大量重复时，用主观方法作出决策和推断是适合的，其至少是统计（频率）方法的一种补充 .当然，主观方法也应符合以上的公理化性质 .

1.3 条件概率与全概率公式

至此，我们一直是在没有试验其他信息的情况下计算事件的概率的；但一旦知道试验中某个情况已经发生，总是希望在试验中能利用新得到的信息，从而重新去认识随机事件的不确定性 .那么，如何去认识利用这些信息呢?

【引例】E：掷一颗骰子，令 $A=\{$掷出“2”点$\}$，$B=\{$掷出偶数点$\}$；易知，$P(A)=\frac{1}{6}$，

$P(B)=\frac{1}{2}$；假若已经知道某次试验“已掷出偶数点”，那么此时“掷出‘2’点”的“概率”应为 $\frac{1}{3}$；记之为：$P(A|B)$ 或 $P_B(A)=\frac{1}{3}$；它并不同于原先的概率 $P(A)=\frac{1}{6}$；有这样一种观点：事件 B 的发生改变了试验 E 原先的样本空间 $\Omega=\{\omega_1,\omega_2,\cdots\omega_6\}$，（其中 ω_i 一掷出“i”点）或者确切地说缩减了样本空间；由于事件 B 的发生，原样本空间 Ω 中的可能结果 $\omega_1,\omega_3,\omega_5$ 已被排除在外，此时缩减后的样本空间应该是：$\Omega_B=B=\{\omega_2,\omega_4,\ \omega_6\}$，基于 Ω_B 的等可能性分析，事件 A 发生的概率应是 $\frac{1}{3}$.

以上观点即是关于条件概率的直观解释，以下我们遵循这种直观解释（尽管有些以偏概全），基于古典概型和几何概型来讨论“条件概率的定义” .

① 由古典概型入手：设事件 A,B 和 AB 分别包含样本空间总数为 n 的样本点中 n_A,n_B,n_{AB} 个，由条件概率的直观解释，$P(A|B)=\dfrac{n_{AB}}{n_B}=\dfrac{n_{AB}/n}{n_B/n}=\dfrac{P(AB)}{P(B)}$.

② 由几何概型入手：考虑试验 E：向平面有界区域 Ω 内随机地投掷一点, 令 $A=\{$ 随机点落入子区域 $A\}$ ，$B=\{$随机点落入子区域 $B\}$ ，$AB=\{$随机点落入子区域 $AB\}$，分别以 $\mu(\Omega),\mu(A),\mu(B),\mu(AB)$ 表示区域 Ω,A,B,AB 的面积；则有：

$$P(B)=\frac{\mu(B)}{\mu(\Omega)},P(AB)=\frac{\mu(AB)}{\mu(\Omega)}\ ,\ P(A|B)=\frac{\mu(AB)}{\mu(B)}=\frac{\mu(AB)/\mu(\Omega)}{\mu(B)/\mu(\Omega)}=\frac{P(AB)}{P(B)}\ .$$

条件概率的一般定义：设 (Ω,F,P) 为一概率空间，$A,B\in F$，且 $P(B)>0$ ，在已知事件 B 发生的条件下，事件 A 发生的条件概率定义为：

$$P(A|B)=P(AB)/P(B)\ .$$

注：① 在初等概率论中，没有测度论的支持，我们无法给出条件概率的严格定义，也无法给出 $P(B)=0$ 时的条件概率 $P(A|B)$.

② 为了区别于 $P(A|B)$，常称 $P(A)$ 为事件 A 的无条件概率 .

③ 取 $B=\Omega$,则 $P(A|\Omega)=\dfrac{P(A\Omega)}{P(\Omega)}=P(A)$，故无条件概率只是条件概率的一种特殊情形 .

由上述定义并不能确切地给出 $P(A)$ 与 $P(A|B)$ 的必然联系；我们可以借助维恩（$Venn$）图 来讨论二者的关系：

(a) $A\cap B=\Phi,P(A|B)=0\le P(A)$ ；(b) $A\subset B,P(A|B)=P(A)/P(B)\ge P(A)$ ；

(c) $A\supset B,P(A|B)=1\ge P(A)$ ；(d) $A\cap B\ne\Phi,P(A/B),P(A)$ 无法比较 .

注：关于条件概率 $P\left(A|B\right)$ 可视具体情况运用如下两种方法来计算：

① 在缩减后的样本空间 Ω_B 中依条件概率的直观解释计算 .

② 在原来的样本空间 Ω 中直接依定义计算 .

例 **15** ① 若 $P\left(\overline{A}\right)=0.3$，$P\left(B\right)=0.4$，$P\left(A\overline{B}\right)=0.5$，试求 $P\left(B\middle|A\cup\overline{B}\right)$.

② 现有 5 件产品，3 件正品，2 件次品，现进行不放回抽样，每次取一件，令 $A_i=\{$ 第 i 次取到次品 $\}$，$i=1,2,\cdots,5$，试求 $P\left(A_2\middle|\overline{A_1}\right)$.

例 **16** ① 设一批产品中一、二、三等品各占 60%、35%、5%，从中任取一件，结果不是三等品，求"取到的是一等品"的概率 .

② 设 10 件产品中有 4 件是不合格品，从中任取两件，已知其中一件是不合格品，求"另一件也是不合格品"的概率 .

例 **17** 已知 $0<P\left(A\right),P\left(B\right)<1$，$P\left(B|A\right)=1$，试求 $P\left(\overline{A}\middle|\overline{B}\right)$.

例 **18**（抽签问题）竹筒中有 10 支签，其中有 2 支好签，10 人依次抽取 1 支，取后不放回，令 $A_i=\{$ 第 i 人抽得好签 $\}$，$i=1,2,\cdots,10$，试求 $P\left(A_i\right)$.

例 **19** 以 A_t 表示"一分子在 $(0,t]$ 内不与其他分子碰撞"，假设"分子在 $(0,t]$ 内不发生碰撞的条件下，在 $(t,t+\Delta t]$ 内发生碰撞"的概率为 $\lambda\Delta t+o\left(\Delta t\right)$，试求 $P\left(A_t\right)$.

条件概率 $P\left(\cdot|B\right)$ 满足如下性质：

①【非负性】对每一 $A\in F$，有 $0\le P\left(A|B\right)\le 1$.

②【规范性】$P\left(\Omega|B\right)=1$.

③【可列可加性】若 $\forall n\ge 1, A_n\in F$，且 $A_iA_j=\Phi$，则

$$P\left(\bigcup_{i=1}^{\infty}A_i|B\right)=P\left(\sum_{i=1}^{\infty}A_i|B\right)=\sum_{i=1}^{\infty}P\left(A_i|B\right) .$$

从而，$P\left(\cdot|B\right)$ 为定义于 $\left(\Omega,F\right)$ 上的概率测度（函数）；它理应满足概率的所有性质，如：有限可加性、加法公式、单调性、连续性等等。记 $P\left(\cdot|B\right)=P_B\left(\cdot\right)$，即有 $\left(\Omega,F,P_B\right)$ 也是一个概率空间，称之为条件概率空间；此即表明：将概率空间 $\left(\Omega,F,P\right)$ 限制在事件 B 上则得到条件概率空间 $\left(\Omega,F,P_B\right)$.

设 A 为概率空间 $\left(\Omega,F,P\right)$ 上的正概率事件，即：$P\left(A\right)>0$，若 $B\in F$，且 $P_A(B)>0$，则对条件概率空间 (Ω,F,P_A) 而言，$\forall C\in F$，有

$$P_A(C|B)=\frac{P_A(BC)}{P_A(B)}=\frac{P(BC|A)}{P(B|A)}=P(C|A\cap B) .$$

条件概率有三大用途：乘法公式、全概率公式和贝叶斯公式 .

乘法公式：设$\left(\Omega, F, P\right)$为一概率空间，若$A_1, A_2, \cdots, A_n \in F$，且$P(\bigcap_{j=1}^{n-1} A_j) > 0$，则：

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P\left(A_n\middle|\bigcap_{i=1}^{n-1} A_i\right);$$

特别地，若

$$A, B \in F, P(AB) = P(A)P(B|A) \ .$$

例 20 袋中有6只白球和4只红球，现任意一一取出，取后不放回，令A_k ={ 第k次取出红球 }，$1 \le k \le 10$，试求$P(A_1A_2)$，$P(A_2)$（本题也可用古典方法计算）.

例 21 一批产品共100件，其中有次品10件，合格品90件；现从中任取一件，取后不放回，接连取三次，试求“第三次才取到合格品”的概率（乘法公式，古典方法）.

全概率公式是一个重要且实用的公式，它的思想贯穿了概率论的始终，成为概率论的一个重要特色；其思想很简单，即：分解；意指引入一个适当的划分将复杂事件分解为若干简单事件，通过计算出简单事件的概率而得到复杂事件的概率；这种思想也是我们通常解决复杂问题的一般思路 .

划分（分割，剖分）：设$\left(\Omega, F, P\right)$为概率空间，若$A_i \in F$，$i \ge 1$，且$A_i \bigcap A_j = \Phi, i \ne j$；并有$\bigcup_{i=1}^{n} A_i = \Omega$ ；则称 $A_1, A_2, \cdots, A_n$ 为样本空间Ω的一个（有限，穷）划分（分割，剖分），或称{$A_1, A_2, \cdots, A_n$}为样本空间Ω的一个完备事件组 . 类似，可给出可列情形的定义 .

全概率公式：① 设$\left(\Omega, F, P\right)$为概率空间，$A_1, A_2, \cdots, A_n$ 为样本空间Ω的一个有限划分，且 $P(A_i) > 0, i = 1, 2, \cdots, n$ ；则 $\forall B \in F$ ，有

$$P(B) = P(B\Omega) = P(B \bigcap \{\bigcup_{i=1}^{n} A_i\}) = P(\bigcup_{i=1}^{n} BA_i) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(A_i)P(B|A_i) \ .$$

② 设$\left(\Omega, F, P\right)$为概率空间，$A_1, A_2, \cdots, A_n, \cdots$为样本空间的一个可列划分，且 $P(A_i) > 0, i = 1, 2, \cdots$ ；则 $\forall B \in F$ ，有

$$P(B) = P(B\Omega) = P(B \bigcap \{\bigcup_{i=1}^{\infty} A_i\}) = P(\bigcup_{i=1}^{\infty} BA_i) = \sum_{i=1}^{\infty} P(BA_i) = \sum_{i=1}^{\infty} P(A_i)P(B|A_i) \ .$$

③ 设$\left(\Omega, F, P\right)$为概率空间，$B, A_i \in F$ 且 $P(A_i) > 0, i = 1, 2, \cdots, n$ ；当$i \ne j$时，$A_i \bigcap A_j = \Phi$ ，并有 $B \subset \bigcup_{i=1}^{n} A_i$ ，则有

$$P(B) = P(B \bigcap \{\bigcup_{i=1}^{n} A_i\}) = P(\bigcup_{i=1}^{n} BA_i) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(A_i)P(B|A_i) \ .$$

④ 设$\left(\Omega, F, P\right)$为概率空间，$B, A_i \in F$ 且 $P(A_i) > 0, i = 1, 2, \cdots, n$ ；当$i \ne j$时，$A_i \bigcap A_j = \Phi$ ，并有 $P(\sum_{i=1}^{n} A_i) = 1$ ；则有

$$P(B) = P(B\Omega) = P(\bigcup_{i=1}^{n} BA_i) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(A_i)P(B|A_i) \ .$$

例 22（敏感问题调查）在调查家庭暴力（或婚外恋，吸毒等敏感问题）所占家庭的比例 p 时,被调查者往往不愿回答真相，这使得调查数据失真.为得到实际的 p 同时又不侵犯个人隐私，调查人员将袋中放入比例是 p_0 的红球和比例是 $q_0=1-p_0$ 的白球 .被调查者在袋中任取一球窥视后放回，并承诺取得红（白）球就讲真（假）话 .被调查者只需在匿名调查表中填“是”（即有家庭暴力）或“否”，然后将调查表放入投票箱，没人知道被调查者是否讲真话和回答的是什么 .

① 如果声称有家庭暴力的家庭比例是 p_1，求 p .

② 如果袋中装有30个红球，50个白球，调查了320个家庭，其中有195个家庭回答“是”，试估算 p .

例 23 甲袋中有2只白球1只黑球，乙袋中有1只白球2只黑球，今从甲袋中任取一球放入乙袋，再从乙袋中任取一球，求“此球是白球”的概率 .

例 24【末步分析法】

① 连续 n 次掷一枚硬币，第一次掷出正面的概率为 a，“第二次以后每次出现与前一次相同面”的概率为 b ，求“第 n 次掷出正面”的概率 .

② 设有 n 只袋子，每只袋中有 a 只黑球和 b 只白球，现从第一只袋中任取一球放入第二只袋中，然后从第二只袋中任取一球放入第三只袋子，依此下去，问：“从第 n 只袋中任取一球是黑球”的概率 .

例 25【首步分析法】

①（玻利亚『polya』概型）罐中有 a 只黑球和 b 只白球，每次从中任取一球并连同 c 个同色球一起放回，如此反复进行，试求“第 n 次取球时取出黑球”的概率【 $c=0(-1)$ 时即有（无）放回抽样】.

② 袋中有 n 只黑球2只白球，每次任取一球并放入一只黑球，试求“第 k 次取出黑球”的概率 .

贝叶斯公式（又称逆概率公式）：设 (Ω,F,P) 为概率空间，$A_1,A_2,\cdots,A_n$ 为样本空间 Ω 的一个有限划分，且 $P(A_i)>0,i=1,2,\cdots,n$；则 $\forall B\in F$，且满足 $P(B)>0$，有

$$P(A_j|B)=\frac{P(A_jB)}{P(B)}=\frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n}P(A_i)P(B|A_i)}.$$

关于其他情形的贝叶斯公式，这里不再累述！

例 26（伊索寓言－孩子与狼）试由 **Bayes** 公式分析大人（村民）对放羊小孩的信任程度是如何下降的 .

例 27 在回答有四个选项 A,B,C,D 的选择题时，由于题目较难，全班只有百分之五的学生能解出正确答案；假设能解出答案的学生回答正确的概率是 **0.99**，不能解出答案的学生随机猜测答案；试求“答题正确的学生是猜对答案”的概率并评价这样出题是否合适 ？

例 28 设有来自三个地区的分别有10名，15名和25名考生的报名表，其中女生报名表分别有3份，7份和5份；现随机地抽取一个地区的报名表，从中先后抽出两份，

① 求“先抽到的是一份女生表”的概率 .

② 已知后抽到的是一份男生表，求“先抽到的是一份女生表”的概率 .

注：有时，我们干脆把贝叶斯公式中的 A_i 看作是导致事件 B 发生的各种可能原因或者影响事件 B 发生的各种可能因素，那么全概率公式提示了我们：事件 B 发生的概率恰好是

事件 B 在这些“原因”事件发生下的条件概率的加权平均，权重为这些“原因”事件的概率 $P\left(A_i\right)$；常称这些 $P\left(A_i\right)$ 为“先验（验前）概率”；贝叶斯公式告诉我们：倘若在一次试验中，某正概率事件 B 已经发生，而事件 $A_1, A_2, \cdots$ 都可能导致或影响事件 B 的发生.“事件 B 已发生”这个信息自然应该被用来重新估计先验概率，贝叶斯公式正是从数量上刻划了这个变化，有时称 $P(A_i|B)$ 为事件 A_i 的“后验（验后）概率”。一言以蔽之，如果说全概率公式是由“原因”求“结果”(由因及果)的话，那么贝叶斯公式则是由“结果”求“原因”（由果溯因）.

例 29 有两箱零件，第一箱装 50 件，其中 20 件是一等品；第二箱装 30 件，其中 18 件是一等品；现从两箱中随意挑出一箱，然后从该箱中先后任取两个零件，求：

（**1**） “第一次取出一等品”的概率 .

（**2**） “第二次取出一等品”的概率 .

（**3**） 在第一次取出一等品的条件下，“第二次取出的仍然是一等品”的概率 .

（**4**） 在第二次取出一等品的条件下，“第一次取出的仍然是一等品”的概率 .

1.4 随机事件的独立性

在概率空间 (Ω, F, P) 中，$A, B \in F, P(A)$ 与 $P(A|B)(P_B(A))$ 常常是有所区别的；从而意识到：事件 B 的发生（已知信息）有可能会改变事件 A 发生的可能性大小 .

考虑一种情形：$P\left(A\right)=P\left(A\middle|B\right)$ **(a)**，此时即有：$P_B\left(A\right)=\dfrac{P\left(AB\right)}{P\left(B\right)}=P\left(A\right)$，

也即： $P\left(AB\right)=P\left(A\right)P\left(B\right)$ **(b)**；进一步，若有 $0<P\left(B\right)<1$，则有：

$$P\left(A\middle|\overline{B}\right)=P_{\overline{B}}\left(A\right)=\frac{P\left(A\overline{B}\right)}{P\left(\overline{B}\right)}=\frac{P\left(A\right)-P\left(AB\right)}{1-P\left(B\right)}=P\left(A\right) \quad \textbf{(c)}；$$

因此，若事件 B 的发生没有影响 A 发生的可能性（**(a)**式成立），则事件 B 不发生也不会影响 A 发生的可能性（**(c)**式成立）；从而可以说，事件 B 的发生与否不影响事件 A 的发生 .反之，若又有 $P\left(A\right)>0$，则由**(b)**式易得：

$$P\left(B\middle|A\right)\left[P_A\left(B\right)\right]=\frac{P\left(AB\right)}{P\left(A\right)}=P\left(B\right) \quad \textbf{(d)}$$

；若再有条件：$P\left(A\right)<1$，则又有：

$$P\left(B\middle|\overline{A}\right)\left[P_{\overline{A}}\left(B\right)\right]=\frac{P\left(\overline{A}B\right)}{P\left(\overline{A}\right)}=\frac{P\left(B\right)-P\left(AB\right)}{1-P\left(A\right)}=P\left(B\right) \quad \textbf{(e)}$$

；因此，若以上所提及的条件若均满足，则有：若事件 B 的发生与否不影响事件 A 的发生，则事件 A 的发生与否不影响事件 B 的发生;也就是所谓 A, B 的发生与否相互不影响，以后常称之为（相互）独立!

事件的独立性 I：设 $\left(\Omega, F, P\right)$ 为概率空间，若 $A, B \in F$，且满足：$P\left(AB\right)=P\left(A\right)P\left(B\right)$，则称 A, B（统计）相互独立（又称两两独立），简称独立；否则称为（统计）相依 .

注：① 在独立性的定义式中采用了**(b)**式，这是由于在该式中，A, B 的地位是对等的，更能体现出独立的相互性 .

② **(b)**式的内涵也比其他四式更加丰富；可以验证：Φ, Ω，几乎不可能事件，几乎必然事件与任意事件都是相互（统计）独立的 .

③ 可以验证：若$\{A,B\},\{\overline{A},B\},\{A,\overline{B}\},\{\overline{A},\overline{B}\}$这四组事件中有一组相互独立，则其他三组必相互独立 .

例 30 设A事件与其自身独立，则 $P(A)=0$ 或 1 .

例 31 A,B独立，且两事件中“仅A发生”和“仅B发生”的概率均为$\frac{1}{4}$，求$P(A)$和$P(B)$.

例 32 甲乙两人“独立”地对同一目标射击一次，其命中率分别为0.8和0.7，现已知目标被击中，求“甲命中”的概率 .

事件的独立性 **II**：设(Ω,F,P)为概率空间，$A,B,C\in F$ ，若以下条件：

① $P(AB)=P(A)P(B)$ ；② $P(AC)=P(A)P(C)$ ；③ $P(BC)=P(B)P(C)$ ；

④ $P(ABC)=P(A)P(B)P(C)$ ；则称A,B,C相互独立 .

注：① 对于三个或三个以上的事件来说，相互独立必两两独立，但两两独立未必相互独立.

② 可以验证：任选三事件中的A作为主体，其与$B\bigcup C,B\bigcap C,B-C$等等均（相互，两两）独立 .

③ 可以验证：若$\{A,B,C\},\{\overline{A},B,C\},\{A,\overline{B},C\},\{A,B,\overline{C}\},\{\overline{A},B,\overline{C}\},\{A,\overline{B},\overline{C}\},\{\overline{A},\overline{B},C\}$，$\{\overline{A},\overline{B},\overline{C}\}$这八组事件中有一组相互独立，则其他七组必相互独立 .

例 33 三人独立地破译一个密码，他们单独译出的概率分别为$\frac{1}{3}$、$\frac{1}{4}$、$\frac{1}{5}$，求“此密码被译出”的概率 .

例 34 三门高射炮独立地向一飞机射击，已知“飞机中一弹被击落”的概率为0.4，“飞机中两弹被击落”的概率为0.8，中三弹则必然被击落；假设每门高射炮的命中率为0.6，现三门高射炮各对飞机射击一次，求“飞机被击落”的概率 .

事件的独立性 **III**：设(Ω,F,P)为概率空间，$A_1,A_2,\cdots,A_n\in F$ ，如果对于任意的$s,2\le s\le n$ ，以及任意的$1\le i_1<i_2<\cdots<i_s\le n$ ，都有$P(A_{i_1}A_{i_2}\cdots A_{i_s})=P(A_{i_1})P(A_{i_2})\cdots P(A_{i_s})$，则称事件$A_1,A_2,\cdots,A_n$相互独立 .

注：该定义中给出的“概率等式”实际上包含2^n-1-n个等式；同样，对于多个事件的情形，相互（整体）独立必两两独立，反之则不成立！

试验的独立性：若试验E_1和试验E_2的所有可能结果（事件）之间相互独立，则称试验E_1与试验E_2独立；

独立重复试验－**Bernoulli** 试验：若试验$E_1,E_2,\cdots,E_n$相互独立，且$E_i=E,\Omega_i=\Omega$，则称$E_1,E_2,\cdots,E_n$为n次独立重复试验；若每次试验只关心某个事件A发生与否，则称之为“n重 **Bernoulli** 试验”.

设在n重 **Bernoulli** 概型中，事件A在此n次独立重复试验中指定的某k次试验（不妨设前k次）中发生，而在其余的$n-k$次试验中不发生，则其概率为：

$$P\left(A_1A_2\cdots A_k\overline{A_{k+1}}\cdots\overline{A_n}\right)=P\left(A_1\right)P\left(A_2\right)\cdots P\left(A_k\right)P\left(\overline{A_{k+1}}\right)\cdots P\left(\overline{A_n}\right)=p^kq^{n-k},q=1-p,$$

这里，$A_i=\left\{第i次试验中A事件发生\right\},i=1,2,\cdots,n$；从而，在$n$重 **Bernoulli** 试验中，“事件$A$恰好发生$k$次”的概率$P$为：

$$P=P\left(\bigcup_{1\le i_1<i_2<\cdots<i_k\le n}A_{i_1}A_{i_2}\cdots A_{i_k}\overline{A_{i_{k+1}}}\cdots\overline{A_{i_n}}\right)=\sum_{1\le i_1<i_2<\cdots<i_k\le n}P\left(A_{i_1}A_{i_2}\cdots A_{i_k}\overline{A_{i_{k+1}}}\cdots\overline{A_{i_n}}\right)=\sum_{1\le i_1<i_2<\cdots<i_k\le n}p^kq^{n-k}$$

$$=\frac{n!}{k!(n-k)!}p^kq^{n-k},k=0,1,\cdots,n$$；称之为“二项概率公式”.

例 35 一射手对同一目标独立地射击四次，若“至少命中一次”的概率为$\frac{80}{81}$，试求该射手进行一次射击的命中率 .

例 36（小概率事件原理）设随机试验E中事件A出现的概率$\varepsilon(>0)$，则无论ε多么地小，当我们不断地重复做试验时，“A迟早会出现”的概率为1（利用概率的连续性和可加性分别做）.

注：① 虽然小概率事件在一次试验中不太可能发生，但在不断重复该试验时，它却必定迟早会发生；正所谓“只要功夫深，铁杵磨成针”，“智者千虑，必有一失”，“多行不义必自毙”$\cdots$！

② 在生产和生活的许多实际中，人们常不约（言）而同（喻）地认定独立性满足；即使对于系统中看上去有明显相依性的各个环节，也常常近似地假定它们是独立运行，因为“独立性假设”常常可以大大简化概率的计算！

例 37 甲、乙二人轮流射击，首先命中目标者获胜；已知甲的命中率为a，乙的命中率为b，甲先射击，试求“甲（乙）获胜”的概率 .

例 38 甲、乙两选手进行乒乓球单打比赛，已知每局中“甲获胜”的概率为0.6，“乙获胜”的概率为0.4；比赛可采用三局两胜制或五局三胜制，问：何种赛制对甲更有利？

第二章 随机变量及其概率分布

2.1 一维随机变量及其分布函数

概率论是从数量的角度来研究随机现象的统计规律性的；其通过建立一系列的定理和公式，借以更好地描述处理和解决各种与随机现象有关的理论和应用问题 .在处理实际问题时，为全面研究随机试验的结果以揭示随机事件中包含的客观存在的统计规律，常须考虑那些与试验结果联系的某些量；对于那些不表现为数量的结果，也可通过数量处理而得以数量化 .这些量是随着试验结果的变化而变化的，因而是“随机的”.这里的量化过程对应着一个法则，其使样本空间的“样本点”与（实）数集的“数”一一对应起来，这个“随机的数”即为随机变量 .

注：利用随机变量（由俄国的彼得堡学派引入）可以表示随机事件（这是随机事件的第三种表示法），可以将不同的随机事件的规律及其相互关系进行统一处理，且有助于研究随机事件的本质规律；它也极大地丰富了概率论的研究对象与范围，使得微积分等分析工具得以引入从而进一步地拓宽了概率论的应用领域（可以形象地说：随机变量架起了概率论与分析学之间的桥梁）.

举例说明试验结果的数量化处理过程；

引例：(**1**) E_1：掷一枚均匀的硬币；样本空间为 $\Omega=\{\omega_1,\omega_2\}$，其中：$\omega_1$ 一掷出正面，ω_2 一掷出反面；指定 1 与 ω_1 对应，指定 0 与 ω_2 对应，则我们建立起 $\Omega=\{\omega_1,\omega_2\}$ 与 $\{0,1\}$ 之间的一一对应，不妨记之为：Y，即：$Y(\omega_1)=1, Y(\omega_2)=0$.

(**2**) E_2：掷一颗骰子；样本空间为 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_6\}$，其中：$\omega_i$ 一掷出"i"点，$i=1,2,\cdots,6$；指定 i 与 ω_i 对应，则我们建立起 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_6\}$ 与 $\{1,2,\cdots,6\}$ 之间的一一对应，不妨记之为：X，即有：$X(\omega_i)=i, i=1,2,\cdots,6$.

随机变量（直观）定义：设 E 为某随机试验，其样本空间为 Ω；若对于样本空间 Ω 的每个样本点 ω，都有唯一的实数 $X(\omega)$ 与之对应，即：$X(\omega)$ 是定义于 Ω 上取值于 R 的单值实值函数（映射），记之为：$X(\omega):\Omega\to R$，称 $X(\omega)$ 为（实）随机变量（**random variable**），简记为：$r.v.$.

注：① 常用希腊字母 ξ,η,ζ 等或者大写英文（拉丁）字母 X,Y,Z 等（或加以下标）来表示随机变量，而随机变量的取值则用相应的小写字母（若有的话）.

② 对于一个随机试验，我们不仅关心出现了哪些结果，更关心这些结果出现的可能性大小；因此，建立了"样本点"与"实数"之间的对应关系，更须知道随机变量取这些值（解释为何是事件）的概率 .

以下，通过一个简例说明"由随机变量表示的随机事件之间的关系与运算"！

例：E:掷一颗骰子，令 ω_i – 掷出"i"点，$i=1,2,3,4,5,6$；易见，

$\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\}$，记 X 为骰子掷出的点数，易见，$R(X)=\{1,2,3,4,5,6\}$；

直观上：$\{X\le 2\}\subset\{X\le 4\}$；

实际上：$\{X\le 2\}=\{\omega\mid X(\omega)\le 2\}=\{\omega_1,\omega_2\}$，

$\{X\le 4\}=\{\omega\mid X(\omega)\le 4\}=\{\omega_1,\omega_2,\omega_3,\omega_4\}$，$\{\omega_1,\omega_2\}\subset\{\omega_1,\omega_2,\omega_3,\omega_4\}$；

直观上：$\{X\le 4\}-\{X\le 2\}=\{2<X\le 4\}$；

实际上：$\{X\le 4\}-\{X\le 2\}=\{\omega\mid X(\omega)\le 4\}-\{\omega\mid X(\omega)\le 2\}=\{\omega_3,\omega_4\}$，

$\{2<X\le 4\}=\{\omega\mid 2<X(\omega)\le 4\}=\{\omega_3,\omega_4\}$；

直观上：$\overline{\{X\le 2\}}=\{X>2\}$；

实际上：$\overline{\{\omega\mid X(\omega)\le 2\}}=\overline{\{\omega_1,\omega_2\}}=\Omega-\{\omega_1,\omega_2\}=\{\omega_3,\omega_4,\omega_5,\omega_6\}$.

这一切告诉我们：对于由随机变量表示的随机事件的关系与运算，完全可以借助对事件直观上的认识来弄清事件的关系与运算，这也会给实际的运算带来极大的方便！

2.2 离散型随机变量及其分布律

引例：独立地掷两颗骰子，记掷得的点数之和为 X ,试讨论其可能取值情况及相应的概率 .

一般地，一个离散型随机变量（$d.r.v.$）的（概率）分布可表示为：

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix} \quad ,$$

其中$\{x_1, x_2, \cdots, x_k, \cdots\}$为$r.v.X$的取值范围（不妨记之为$R(X)$），而且：

① $P(\{\omega : X(\omega) = x_k, \omega \in \Omega\}) = P(\{X = x_k\}) = P(X = x_k) = p_k$，$k = 1, 2, \cdots$；

② $\sum_{k}^{?} p_k = P(\bigcup_{k} \{\omega : X(\omega) = x_k\}) = P(\Omega) = 1$.

事实上，可以验证：凡是满足上述两条件的一组数$\{p_1, p_2, \cdots, p_k, \cdots\}$（有限个或可列个）必可作为某个离散型随机变量的概率分布（又称分布律或分布列）.

例 1 从$1, 2, 3, 4$中任取一数X，再从$1, 2, \cdots, X$中任取一个数，记之为Y，试求$P(Y = 2)$.

例 2 将一颗骰子连续地掷n次，记 X（Y）为n个点数的最小（大）值，试求X, Y的“概率分布” .

例 3 （常见的离散型分布）

① 退化分布与两点分布 .

②（二项分布）设已做了n重贝努利试验（结果：“成功”和“失败”），每次试验中“成功”的概率为p，令Z表示n次试验中“成功”的次数，试求Z的概率分布 .

③（几何分布）连续独立地掷一枚硬币，记首次掷出正面（不妨称之为“成功”）时已掷的次数为Y（每次掷出正面的概率为p），试讨论其可能取值情况及相应的概率 .

注：几何分布的无记忆性－又称无后效性或永远年青性；几何分布具有无记忆性的根本原因在于每次试验中“成功”的概率不随试验次数而改变 .常有一类产品在使用过程中没有明显损耗，其寿命就具有这种性质 .如：喝水用的玻璃杯—每喝一次水，对杯子几乎没有一点磨损 ，它往往是由于偶然失手才损坏的，而不是在使用中磨碎的，当使用多次后杯子没有破碎，它仍然如同新的一般被使用，此即“无记忆性”.因此无记忆性的实际意义就是“没有明显的损耗”，故服从几何分布的随机变量常用来表示无明显损耗的一类产品的使用寿命 .需注意的是：一个取正整数值的离散型随机变量具有无记忆性的充要条件是它服从几何分布！

历史上，$Poisson$分布是作为二项分布的极限分布出现的 .

注：（$Poisson$定理）在n重$Bernoulli$试验中，记事件A在一次试验中发生的概率为p_n（与试验次数n有关），X为n次试验中A发生的次数；如果当$n \to \infty$时，有$np_n \to \lambda$，则 $\lim\limits_{n \to \infty} P(\{X = k\}) = \lim\limits_{n \to \infty} C_n^k \cdot p_n^k \cdot (1 - p_n)^{n-k} = \dfrac{\lambda^k}{k!} e^{-\lambda}$ ；也即有：当n充分大时，$X \dot{\sim} P(\lambda)$.

例 4（$Poisson$分布）为保证设备正常工作，需要配备一些维修工；如果各台设备发生故障是相互独立的，且每台设备发生故障的概率都是0.01；试在以下情况下，求“设备发生故障而不能及时修理”的概率；

a）若一人包干20台；

b）若三人包干80台；

c）若某工厂有同类设备300台，为了保证设备发生故障而不能及时修理的概率小于0.01，至少应配备多少维修工 ？

2.3 连续型随机变量及其概率密度

引例：向$\left[0,1\right]$区间随机地掷一点，由“等可能性”可知点落入其任一子区间的概率等于其长度（几何概型）；不妨设点的位置（坐标）为 X ，则 X 是一个“随机变量”，$R\left(X\right)=\left[0,1\right]$，$\forall a,b\in\left[0,1\right],P\left(a\le X\le b\right)=P\left(\left\{\omega\middle|a\le X\left(\omega\right)\le b\right\}\right)=\int_a^b 1dx$ $=b-a$；如何去刻划$r.v.X$ 取值的统计规律呢？在上式中，令$a=b$， 则有：$P\left(\left\{\omega:X\left(\omega\right)=a\right\}\right)=P\left(\left\{X=a\right\}\right)=0$ ；由此说明，$r.v.X$ 不同于上一节的离散型随机变量，他们有着本质的区别！

连续型随机变量（**continuous random variable or c.r.v.**）： 设 X 为随机变量，若存在非负函数 $f\left(x\right)$， 使得$\forall a,b\in R,a<b$ ，都有：$P\left(\left\{\omega:a<X\left(\omega\right)\le b\right\}\right)=P\left(\left\{a<X\le b\right\}\right)=$ $\int_a^b f(x)dx$ ；则称 X 为连续型随机变量，简记为：**c.r.v.** ;称 $f\left(x\right)$为 X 的概率密度函数（概率密度、分布密度、密度）（**probability density function**）简记为：**p.d.f.** .

注：① 定义中的条件等式可等价地描述为：

$$\forall a\in R,\ P\left(\left\{\omega:X(\omega)\le a\right\}\right)=P\left(\left\{X\le a\right\}\right)=P\left(\left\{-\infty<X\le a\right\}\right)=\int_{-\infty}^{a}f(x)dx .$$

② 由定义，概率密度函数不是唯一的；这是由于改变其作为被积函数在至多可列个点处的函数值不改变积分值！

③ 若 $f\left(x\right)$是一个概率密度函数，令：$A_n=\left\{\omega:-n<X(\omega)\le n\right\}=\left\{-n<X\le n\right\},n\ge1,$ 易见，事件列$\left\{A_n,n\ge1\right\}\uparrow\Omega$，并且，$\int_{-\infty}^{+\infty}f(x)dx=\lim_{n\to\infty}\int_{-n}^{+n}f(x)dx=\lim_{n\to\infty}P\left(\left\{-n<X\le n\right\}\right)$ $=\lim_{n\to\infty}P\left(A_n\right)=P\left(\lim_{n\to\infty}A_n\right)=P\left(\Omega\right)=1$ ；反之，若函数 $f\left(x\right)$满足:

a） 非负可积；**b**）$\int_{-\infty}^{+\infty}f(x)dx=1$ ；则 $f\left(x\right)$必可作为某连续型随机变量的密度函数 .

④ 若随机变量 X 具有密度 $f\left(x\right)$，则有：$\dfrac{P\left(\left\{x_0-\Delta x<X\le x_0\right\}\right)}{\Delta x}=\dfrac{1}{\Delta x}\int_{x_0-\Delta x}^{x_0}f(x)dx$ ；

若 $f\left(x\right)$在 x_0 点连续，则有：$\lim_{\Delta x\to0}\dfrac{P\left(\left\{x_0-\Delta x<X\le x_0\right\}\right)}{\Delta x}=f(x_0)$；这类似于物理学中的“质量的线（面，体）密度”，故而称 $f\left(x\right)$为密度；但密度不一定是连续函数（密度的 几何意义） .

⑤ 比较：离散型随机变量与连续型随机变量；

a）若 X 为离散型随机变量，则有：

$P\left(\left\{X=a\right\}\right)>0,a\in R\left(X\right)$ ； $P\left(\left\{X=a\right\}\right)=0,a\notin R\left(X\right)$.

b）若 X 为连续型随机变量，且 $X\sim f\left(x\right)$ ，则有

$P\left(\left\{X=a\right\}\right)\le P\left(\left\{a-\dfrac{1}{n}<X\le a+\dfrac{1}{n}\right\}\right)=\int_{a-\frac{1}{n}}^{a+\frac{1}{n}}f(x)dx\to0,n\to\infty$ ； 从而，$\forall a\in R,P\left(\left\{X=a\right\}\right)=0$；也即：连续型随机变量取任意值的概率总为0！进而有：$P\left(\left\{a<X<b\right\}\right)=P\left(\left\{a<X\le b\right\}\right)=P\left(\left\{a\le X<b\right\}\right)=P\left(\left\{a\le X\le b\right\}\right)$.

例 5 设 $r.v.X$ 有 $p.d.f.\ f(x)=\begin{cases}1/3, & 0\le x\le 1\\ 2/9, & 3\le x\le 6\\ 0, & else\end{cases}$，若 $P(X\ge k)=\dfrac{2}{3}$，试确定 k 的取值范围．

例 6 设 $r.v.X\sim f(x)=\begin{cases}Ax, & 1<x<2;\\ B, & 2\le x<3;\\ 0, & 其他;\end{cases}$，且 $P(\{1<X<2\})=P(\{2<X<3\})$，试求常数 A,B 及 $P(\{2<X<4\})$．

例 7 设连续型随机变量 X,Y 具有相同概率密度，且 $X\sim f_X(x)=\begin{cases}{}^3\!/\!_8\, x^2, & 0<x<2;\\ 0, & 其他;\end{cases}$,若已知 $A=\{X>a\}$ 与 $B=\{Y>a\}$ 相互独立，且 $P(A\cup B)={}^3\!/\!_4$，求常数 a．

例 8 设 $c.r.v.X\sim f(x)=\begin{cases}2x, & 0<x<1;\\ 0, & 其他;\end{cases}$，以 Y 表示对 X 的三次独立重复观察中事件 $\left\{X\le\dfrac{1}{2}\right\}$ 出现的次数，试求 $P(\{Y=2\})$．

均匀分布（**Uniform distribution**）：若随机变量 X 具有密度 $f(x)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b;\\ 0, & 其他;\end{cases}$，则称 X 服从 $[a,b]$ 上的均匀分布，记作：$X\sim U[a,b]$．

注：均匀分布的概率背景：几何概型；E：向 $[a,b]$ 上随机地投掷一点，记点的位置（坐标）为 X；易见，$R(X)=[a,b]$，且 $\forall x\in R$，有：

$x<a,F(x)=P(\{X\le x\})=0;a\le x\le b,F(x)=\dfrac{x-a}{b-a};x>b,F(x)=1$．进一步有：$X$ 的概率密度为 $f(x)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b;\\ 0, & 其他;\end{cases}$；反之，若 $X\sim U[a,b]$，则有：$\forall[c,d]\subset[a,b],P(\{X\in[c,d]\})=\dfrac{d-c}{b-a}$，也即 X 取值于 $[a,b]$ 的任一子区间的概率与该区间的长度成正比而与其位置无关，即为几何概型的情形！

例 9 设 $r.v.\xi\sim U[0,5]$，试求“方程 $4x^2+4\xi x+\xi+2=0$ 有实根”的概率．

指数分布（**Exponential distribution**）：若随机变量 X 具有密度 $f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0;\\ 0, & 其他;\end{cases}$，$\lambda>0$ 为常数，则称 X 服从参数为 λ 的指数分布，记作：$X\sim E(\lambda)$（或 $Ex(\lambda),Exp(\lambda)$）．

例 10 ① 设顾客在某银行的窗口等待服务的时间 X（以分钟记）服从参数为 $\dfrac{1}{5}$ 的指数分布，

某顾客在窗口等待服务若超过10分钟他就离开；他一个月要到银行五次，以Y表示一个月内他未等到服务而离开窗口的次数，试求$P\left(Y\geq 1\right)$.

② 某仪器装了三个独立工作的同型号电子元件，其寿命X（以h计）都服从同一指数分布$E\left(\dfrac{1}{600}\right)$；试求此仪器在最初使用的$200h$内，"至少有一个此种电子元件损坏"的概率 .

指数分布的无记忆性：可验证以下命题的等价性：设X为非负连续型随机变量，则

① X服从指数分布；

② $P\left(X>x+y\middle|X>x\right)=P\left(\left\{X>y\right\}\right)$，$x,y>0$.

【2.1（补充） 分布函数】

概率分布（分布列）和概率密度分别是用来刻划离散型随机变量和连续型随机变量取值的统计规律的；为了统一刻划各类随机变量，也为了避免集（合）函数使用和研究的不便；以下引入（概率）分布函数，它只是普通的点函数 .

（概率）分布函数：设$X=X\left(\omega\right)$是一个随机变量，$\forall x\in R$，定义：$F\left(x\right)=P\left(\left\{X\leq x\right\}\right)$，称之为$r.v.X$的（概率）分布函数，简记为：**d.f. (distribution function).**

注：① 由定义，分布函数$F\left(x\right)$是定义于实数集R上的取值于$\left[0,1\right]$上的点函数，它表示事件$\left\{\omega:X\left(\omega\right)\leq x\right\}$发生的概率，也即：$\left\{\omega:X\left(\omega\right)\in(-\infty,x]\right\}$的概率；当右端点$x$发生变化时，$P\left(\left\{X\leq x\right\}\right)$也相应发生变化，故$P\left(\left\{X\leq x\right\}\right)$是右端点$x$的函数 .

② 对于多个随机变量的情形，常常附之以下标，如：$F_X\left(x\right),F_Y\left(y\right),\cdots\cdots$，自变量用相应的小写，这是为了避免混淆 .

③ 随机变量X的分布函数$F\left(x\right)$包含了它的全部概率特征的信息；如：

1）任取数列$\left\{b_n,n\geq 1\right\}$，满足：$b_1>b_2>\cdots>b_n\cdots\to 0$，再令：

$A_n=X^{-1}\left((-\infty,x-b_n]\right)=\left\{X\leq x-b_n\right\}$，$\forall n$，则有$\left\{A_n,n\geq 1\right\}\uparrow A=\left\{X<x\right\}$

从而有：$P\left(\left\{X<x\right\}\right)=P\left(\lim\limits_{n\to\infty}A_n\right)=\lim\limits_{n\to\infty}P\left(A_n\right)=\lim\limits_{n\to\infty}F\left(x-b_n\right)=F\left(x-\right)$（由数列的任意性），$x\in R$.

2）$\forall x\in R$ ，$\left\{X=x\right\}=\left\{X\leq x\right\}-\left\{X<x\right\}$，从而有：

$P\left(\left\{X=x\right\}\right)=P\left(\left\{X\leq x\right\}\right)-P\left(\left\{X<x\right\}\right)=F\left(x\right)-F\left(x-\right)$ ；类似有：

$\forall x\in R$，$\overline{\left\{X\leq x\right\}}=\left\{X>x\right\},\overline{\left\{X\geq x\right\}}=\left\{X<x\right\}$ ， 从而有：

$P\left(\left\{X>x\right\}\right)=1-P\left(\left\{X\leq x\right\}\right)=1-F\left(x\right)$ ；

$P\left(\left\{X\geq x\right\}\right)=1-P\left(\left\{X<x\right\}\right)=1-F\left(x-\right)$.

3）$\forall a,b\in R,a<b$，有如下结论：

$P\left(\left\{a<X<b\right\}\right)=P\left(\left\{X<b\right\}\right)-P\left(\left\{X\leq a\right\}\right)=F\left(b-\right)-F\left(a\right)$ ；

$P\left(\left\{a<X\leq b\right\}\right)=P\left(\left\{X\leq b\right\}\right)-P\left(\left\{X\leq a\right\}\right)=F\left(b\right)-F\left(a\right)$ ；

$P\left(\left\{a\leq X\leq b\right\}\right)=P\left(\left\{X\leq b\right\}\right)-P\left(\left\{X<a\right\}\right)=F\left(b\right)-F\left(a-\right)$ ；

$P\left(\left\{a\leq X<b\right\}\right)=P\left(\left\{X<b\right\}\right)-P\left(\left\{X<a\right\}\right)=F\left(b-\right)-F\left(a-\right)$.

至此，我们知道，无论随机变量在怎样的区间（开的，闭的，半开半闭的，还是有界的，无界的）取值，都可以通过分布函数来刻划！

例 11 已知 $r.v.X$ 的分布函数为 $F\left(x\right)=\begin{cases} 0, x<0; \\ x/2, 0\le x<1; \\ 2/3, 1\le x<2; \\ 11/12, 2\le x<3; \\ 1, x\ge 3; \end{cases}$，试求：$P\left(X>\frac{1}{2}\right)$、$P\left(X<3\right)$、$P\left(1\le X<3\right)$、$P\left(X=3\right)$.

例 12 设随机变量 ξ 的分布函数为 $F\left(x\right)$，试用 $F\left(x\right)$ 表示下列事件的概率：

$$\left\{|\xi|<1\right\},\left\{|\xi-2|<3\right\},\left\{2\xi+1>5\right\},\left\{a\xi+b\le c\right\},\left\{\xi^2\le 4\right\},\left\{\xi^3<8\right\} .$$

例 13 ① 在半径为 R 的圆内任取一点，求此点到圆心距离 X 的分布函数及概率 $P\left(X>\frac{2}{3}R\right)$.

② 在 ΔABC 内任取一点 P，记 X 为点 P 到底边 BC 的距离，试求 X 的分布函数 .

注：① 若 X 为 $d.r.v.$，且有分布列：$P\left(\left\{X=x_i\right\}\right)=p_i, i=1,2,\cdots$，则有：

$$F\left(x\right)=P\left(\left\{X\le x\right\}\right)=P\left(\bigcup_{i:x_i\le x}\left\{X=x_i\right\}\right)=\sum_{i:x_i\le x}P\left(\left\{X=x_i\right\}\right)=\sum_{i:x_i\le x}p_i .$$

例 14 ①设随机变量 X 有分布列：$\begin{pmatrix} X & 0 & 1 & 3 \\ P & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$，试求 X 的（概率）分布函数 .

② 设 $r.v.X$ 的 $d.f.$ 为 $F\left(x\right)=\begin{cases} 0, & x<2; \\ 1/4, & 2\le x<3; \\ \frac{1}{2}, & 3\le x<5; \\ 1, & x\ge 5; \end{cases}$，试求 X 的分布（列、律）.

注：② 若 X 为 $c.r.v.$，且有概率密度 $f\left(x\right)$；则有：

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

若 $f\left(x\right)$ 连续，则（几乎处处）有：$\frac{d}{dx}F(x)=f(x)$.

例 15 ① 设 X 具有密度 $f\left(x\right)=\frac{1}{2}e^{-|x|}, x\in R$，试求 X 的分布函数 .

② 若 X 的分布函数为 $F\left(x\right)=\begin{cases} 0, & x<0; \\ x, & 0\le x<1; \\ 1, & x\ge 1; \end{cases}$，试求 X 的 $p.d.f.$.

分布函数具有如下性质：

① 单调不减性：$\forall a,b\in R.a<b,\{X\le a\}\subset\{X\le b\}$，从而，

$$F(a)=P(\{X\le a\})\le P(\{X\le b\})=F(b).$$

② 右连续性：任取数列$\{b_n,n\ge 1\}$满足：$\forall a\in R$ ，$b_1>b_2>\cdots>b_n\cdots\to a$ ，令

$A_n=\{a<X\le b_n\},n\ge 1$ ，易见随机列 $\{A_n,n\ge 1\}\downarrow\bigcap_{n=1}^{\infty}A_n=\Phi$ ，则

$$0=P\left(\bigcap_{n=1}^{\infty}A_n\right)=\lim_{n\to\infty}P(A_n)=\lim_{n\to\infty}\left[P(\{X\le b_n\})-P(\{X\le a\})\right]=\lim_{n\to\infty}F(b_n)-F(a);$$

由数列$\{b_n,n\ge 1\}$的任意性，即有：$\forall a\in R$，$F(a)=\lim\limits_{x\to a+}F(x)$.

③ 正则性：令$B_n=\{X\le -n\},C_n=\{X\le n\}$，则有事件列：$\{B_n,n\ge 1\}\downarrow\Phi$，$\{C_n,n\ge 1\}\uparrow\Omega$；从而由概率的连续性知：

$$F(-\infty)=\lim_{n\to\infty}F(-n)=\lim_{n\to\infty}P(B_n)=P\left(\lim_{n\to\infty}B_n\right)=P(\Phi)=0 \quad ;$$

$$F(+\infty)=\lim_{n\to\infty}F(n)=\lim_{n\to\infty}P(A_n)=P\left(\lim_{n\to\infty}A_n\right)=P(\Omega)=1 \quad .$$

例 16 ① 设G为曲线$y=2x-x^2$与x轴所围区域，在区域G内任取一点，该点到y轴的距离为ξ，求的分布函数与密度函数 .

② 假设$r.v.\xi$的绝对值不大于1，$P(\{\xi=1\})=2P(\{\xi=-1\})=\dfrac{1}{4}$，在事件$\{-1<\xi<1\}$出现的条件下，$\xi$在$(-1,1)$内任一子区间上取值的条件概率与该区间的长度成正比，试求ξ的分布函数 .

正态（$Normal$）分布：若连续型随机变量X有概率密度为：$f(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},x\in R$，其中$\mu,\sigma(\sigma>0)$为常数，则称$X$服从参数为$(\mu,\sigma^2)$的正态分布，记作：$X\sim N(\mu,\sigma^2)$；若$\mu=0,\sigma=1$，则称$X$服从标准正态分布 .

注：密度$f(x)$具有如下性质：

① 关于直线$x=\mu$对称，此即： $\forall h>0,P(\{\mu-h<X\le\mu\})=P(\{\mu<X\le\mu+h\})$.

② $f(x)$的图像呈倒立的钟形，且在$x=\mu$处取得最大值，曲线以x轴为水平渐近线；对于同样长度的区间，其离μ越远，X落于其上的概率就越小 .

③ $f(x)$在$x=\mu\pm\sigma$处有拐点，且存在任意阶导数；且σ越大（小），曲线越平坦（陡峭）.

可以验证：$\int_{-\infty}^{+\infty}f(x)dx=1$（$Poisson$积分）.

正态分布是自然界中最常见的一种分布，如：测量的误差，炮弹的落点等诸多现象都服从或近似服从正态分布 .一般地，若影响某一数量的指标很多，且每种影响都相互独立，作用很小，即可认为这个数量服从或近似服从正态分布；许多分布都以正态分布来逼近；许多重要的分布，如三大抽样分布：χ^2（卡方）分布，t（学生氏）分布和F（费歇尔）分布都是由正态分布所导出的；同时，正态分布还有许多优良的性质也使其无论是在理论

上，还是在应用上都有重要的作用！

常将标准正态分布的分布函数和密度函数记为：$\Phi(x)$和$\varphi(x)$；

① 若$r.v.X\sim N(\mu,\sigma^2)$，则

$$P(\{X\le\mu-x\})=\int_{-\infty}^{\mu-x}f(t)dt=\int_{x}^{+\infty}f(\mu-t)dt=1-F(\mu+x)\ .$$

特别地，若$r.v.X\sim N(0,1^2)$，则有：$\Phi(-x)=1-\Phi(x)$，$x\in R$.

② 若$r.v.X\sim N(\mu,\sigma^2)$，令 $Y=\dfrac{X-\mu}{\sigma}$， 则有：$\forall y\in R, F_Y(y)=P(\{Y\le y\})=$

$P(\{X\le\mu+\sigma y\})=\dfrac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\mu+\sigma y}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}e^{-\frac{t^2}{2}}dt=\Phi(y)$，此即说明$Y$服从标准正态分布（任何正态分布的概率计算均可转化与标准正态分布有关的概率计算）！

例 17 ① 设$X\sim N(1,0.4^2)$，试求$P(\{X>0\})$和$P(\{0.2<X<1.8\})$.

② 设$r.v.X\sim N(3,2^2)$，试求：

1） $P(2<X\le 5),P(|X|>2)$ ；2）确定c，使得$P(X>c)=P(X<c)$ ；

3）设d满足$P(X>d)\ge 0.9$，d至多为多少？

例 18 ① 已知$\ln X\sim N(1,2^2)$ ，试求 $P\left(\left\{\frac{1}{2}<X<2\right\}\right)$.

② 设$r.v.\xi$的$p.d.f.$为$f(x)=\dfrac{1}{\sqrt{\pi}}e^{-x^2+2x-1},-\infty<x<+\infty$，试求$P(0\le\xi\le 2)$.

例 19 ① 假设数学的考试成绩近似服从正态分布$N(70,10^2)$，按考分从高到低排第100名的成绩恰为60分（其余的不及格），问：按考分从高到低排第20名的成绩约为多少？

② 由学校到火车站有两条路线，所需时间随交通堵塞状况有所变化，若以分钟计算，第一条路线所需时间$\xi_1\sim N(50,10^2)$，第二条路线所需时间$\xi_2\sim N(60,4^2)$，如果要求：

1）在70分钟内赶到火车站；2）在65分钟内赶到火车站；

试问：各应选择哪条路线？

2.4 随机变量函数的分布

设X是一随机变量，$g(\cdot)$是$Borel$函数，如何求$Y=g(X)$的分布呢？

记：$D_y=\{x|g(x)\le y\}$，$-\infty<y<+\infty$；

$$F_Y(y)=P(\{Y\le y\})=P(\{g(X)\le y\})=P(\{X\in D_y\}),$$

即可通过X的分布来导出Y的分布！

例 20 ① 设$X\sim\begin{pmatrix}-1&0&1&2\\ \frac{1}{4}&\frac{1}{8}&\frac{3}{8}&\frac{1}{4}\end{pmatrix}$，试求$Y_1=2X+1,Y_2=X^2,Y_3=\max\{X,1\}=X\vee 1$的分布 .

② 设X有分布：$P(\{X=k\})=\dfrac{1}{2^k}$，$k=1,2,\cdots$；试求$Y=\sin\dfrac{\pi X}{2}$的分布 .

对于连续型随机变量，其 $Borel$ 函数既可是离散型随机变量，也可是连续型随机变量，还可以是混合型随机变量！

例 21 ① 设 $X \sim U(-1,2)$，$Y=\begin{cases}1, & X\ge 0;\\ -1, & X<0;\end{cases}$，试求 Y 的分布．

② 设 $r.v.\xi \sim N(0,1^2)$，$\eta=\xi$ 或 $\eta=-\xi$ 视 $|\xi|\le 1$ 或 $|\xi|>1$ 而定，试求 η 的分布．

③ 设 $r.v.X \sim f(x)=\begin{cases}\dfrac{1}{9}x^2, 0<x<3;\\ 0, 其他;\end{cases}$，且 $Y=\begin{cases}2, X\le 1;\\ X, 1<X<2;\\ 1, X\ge 2;\end{cases}$，试求 Y 的分布函数及 $P(X\le Y)$．

④ 设 $r.v.X \sim E(1)$，$Y=\min\{X,2\}=X\wedge 2$，试求 Y 的分布．

可以验证：设 $c.r.v.X \sim f_X(x)$，若 $y=g(x)$ 是严格单调函数且可导，则 $Y=g(X)$ 是一个连续型随机变量，且有概率密度函数为：$f_Y(y)=f_X[h(y)]\left|h'(y)\right|I_{\{\alpha<y<\beta\}}(y)$，其中：$h(y)$ 是 $g(x)$ 的反函数，$\alpha=\min\{g(-\infty),g(+\infty)\},\beta=\max\{g(-\infty),g(+\infty)\}$，$g(-\infty)=\lim\limits_{x\to-\infty}g(x)$，$g(+\infty)=\lim\limits_{x\to+\infty}g(x)$．

例 22 ① 设 $r.v.X \sim E\left(\dfrac{1}{\theta}\right)$，$\theta>0$ 为参数，试求 $Y=\dfrac{2}{\theta}X$ 的分布．

② 设 $r.v.X \sim E(2)$，试证：$Y_1=e^{-2X}$ 与 $Y_2=1-e^{-2X}$ 均服从 $(0,1)$ 上的均匀分布．

③ 设 $X \sim N(0,1^2)$，则 $Y=-X \sim N(0,1^2)$．

④ 设 $X \sim U\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right)$，求 $Y=\tan X$ 的概率密度函数．

⑤ 设 $\xi \sim f_\xi(x)=2(1-x)I_{\{0<x<1\}}(x)=\begin{cases}2(1-x), & 0<x<1;\\ 0, & 其他;\end{cases}$，求一单调增函数 $h(x)$，使得 $\eta=h(\xi)\sim E(\lambda)$．

⑥ 设 $r.v.X$ 的 $p.d.f.$ 为 $f(x)=\begin{cases}\dfrac{1}{3\sqrt[3]{x^2}}, x\in[1,8];\\ 0, 其他;\end{cases}$，$F(x)$ 为 X 的 $d.f.$，试求随机变量 $Y=F(X)$ 的分布函数．

对于函数非单调的情形，这里不再给出一般性的公式，通过下面的例子来说明之！

例 23 ① 设 $X \sim U(-1,1)$，试求 $Y=X^2$ 的分布．

② 设 $X \sim f_X(x)=\begin{cases}1+x, & -1\le x<0;\\ 1-x, & 0\le x\le 1;\\ 0, & 其他;\end{cases}$，试求 $Y=X^2+1$ 的分布．

③ 设 $X \sim U\left[-\dfrac{\pi}{2},\dfrac{3\pi}{2}\right]$，$Y=X^4$，求 Y 的密度函数．

第三章 随机向量及其分布

3.1 二维随机变量及其联合分布函数

离散型随机变量的概率分布反映了它取各种可能值的概率分配，包含了它的全部概率信息；其描述了 **d.r.v.**取值的统计规律性。实际情况是：我们常常不得不同时考虑多个随机变量 .一个自然的作法是：将来自同一概率空间的多个随机变量放在一起，作为一个随机取值的向量，即随机向量；其取各种可能的向量值的统计规律性自然就是它取各可能向量值的概率！

例 1 一袋中装有 2 只白球和 3 只红球，现不放回地依次抽取 2 球，设 X 表示“第一次取到白球的只数”，Y 表示“第二次取到白球的只数”，试求 (X,Y) 的“联合（概率）分布”.

例 2 将一枚均匀的硬币连续抛掷 3 次，以 X 表示 3 次中出现正面的次数，以 Y 表示 3 次中出现正面与反面之差的绝对值，试写出 (X,Y) 的“联合分布” .

联合分布：设 $d.r.v.X$ 的 $R(X)=\{x_1,x_2,\cdots,x_m\}$，$d.r.v.Y$ 的 $R(Y)=\{y_1,y_2,\cdots,y_n\}$，

(X,Y) 的所有可能取的向量值为 $\{(x_i,y_j):i=1,\cdots,m;\ j=1,\cdots,n\}$ ，且

$$P(\{(X,Y)=(x_i,y_j)\})=P(\{X=x_i,Y=y_j\})=P((X,Y)^{-1}(\{(x_i,y_j)\}))$$

$$=P(X^{-1}(\{x_i\})\bigcap Y^{-1}(\{y_j\}))=p_{ij},i=1,\cdots,m;\ j=1,\cdots,n;$$

称之为 (X,Y) 的联合分布；通常将其表示为如下的矩阵形式：$\begin{pmatrix} X/Y & y_1 & \cdots & y_n \\ x_1 & p_{11} & \cdots & p_{1n} \\ \vdots & \cdots & \cdots & \cdots \\ x_m & p_{m1} & \cdots & p_{mn} \end{pmatrix}$.

例 3 设 $P(\{X\ge 0,Y\ge 0\})=\dfrac{3}{7}$，$P(\{X\ge 0\})=P(\{Y\ge 0\})=\dfrac{4}{7}$，求 $P(\{\max(X,Y)\ge 0\})$.

例 4 设 A,B 为两个事件，且 $P(A)=\dfrac{1}{4}$，$P(B|A)=\dfrac{1}{3}$，$P(A|B)=\dfrac{1}{2}$，令

$$X=\begin{cases}1, & 若A发生;\\ 0, & 否则;\end{cases}\quad Y=\begin{cases}1, & 若B发生;\\ 0, & 否则;\end{cases}$$

试求：① (X,Y) 的概率分布；② $Z=X^2+Y^2$ 的概率分布 .

例 5 袋中有 1 红 2 黑 3 白共 6 个球，现有放回地从袋中取两次，每次取一球，以 X,Y,Z 分别表示两次取到的红、黑、白球的个数，

① 求 $P(X=1|Z=0)$ ；② 求 (X,Y) 的概率分布 .

例 6 设 $r.v.Y\sim E(1)$，且 $X_k=\begin{cases}0, & Y\le k;\\ 1, & Y>k;\end{cases}$，$k=1,2$；试求 (X_1,X_2) 的联合概率分布 .

连续型随机向量： 设 X,Y 是定义于同一概率空间 (Ω,F,P) 上的随机变量，称二维随机向量 (X,Y) 是连续型的（联合连续的），若存在非负可积函数 $f(x,y)$， $(x,y)\in R^2$,使得对于任意的二维矩形域 $B=\{(x,y):a<x\le b,c<y\le d\}$ ，都有

$$P\left(\left\{\left(X,Y\right)\in B\right\}\right)=P\left(\left\{a<X\le b,c<Y\le d\right\}\right)=\iint_B f\left(x,y\right)dxdy \quad ,$$

并称 $f\left(x,y\right)$为$\left(X,Y\right)$的联合（概率）密度（函数）.

注：① 可以验证联合密度 $f\left(x,y\right)$满足：

a) 非负可积；**b)** $\iint_{R^2} f\left(x,y\right)dxdy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f\left(x,y\right)dxdy=1$.

② 若 $f\left(x,y\right)$在$\left(x_0,y_0\right)$点连续，则有

$$\lim_{\Delta x\to 0,\Delta y\to 0}\frac{P\left(\left\{\left(X,Y\right)\in(x_0-\Delta x,x_0]\times(y_0-\Delta y,y_0]\right\}\right)}{\Delta x\Delta y}=f\left(x_0,y_0\right) \quad ,$$

由该式求联合密度的方法称为“微元法” .

③ 事实上，若$B\in\beta^2$（二维 **Borel** 域）是xoy平面上任一区域，则以X,Y为横，纵坐标的随机点落入区域G的概率为：$P\left(\left\{\left(X,Y\right)\in B\right\}\right)=\iint_B f\left(x,y\right)dxdy$ ；并且也可由该式作为上述定义中的条件式 .

例 **7** ① 设$\left(X,Y\right)\sim f\left(x,y\right)=\begin{cases} x^2+\dfrac{xy}{3},(x,y)\in[0,1]\times[0,2]; \\ 0,\text{其他}; \end{cases}$，试求：$P\left(\left\{X\le Y\right\}\right)$，

$P\left(\left\{X+Y\ge 1\right\}\right)$，$P\left(\left\{X=Y\right\}\right)$，$P\left(\left\{X,Y\text{至少有一个大于}\dfrac{1}{2}\right\}\right)$.

② 设$\left(X,Y\right)\sim f\left(x,y\right)=\begin{cases} e^{-y},0<x<y; \\ 0,\text{其他}; \end{cases}$，试求：$P\left(\left\{X+Y\le 1\right\}\right)$.

例 **8** ① 设$\left(X,Y\right)\sim f\left(x,y\right)=\begin{cases} cx^2y,x^2\le y\le 1; \\ 0,\text{其他}; \end{cases}$，试确定常数$c$并求 $P\left(\left\{\left(X,Y\right)\in D\right\}\right)$，

这里，$D:2x^2\le y\le 1$.

② 设$\left(X,Y\right)\sim f\left(x,y\right)=\begin{cases} Ae^{-y},0<x<y; \\ 0,\text{其他}; \end{cases}$，试求常数$A$，并计算$P\left(\left\{X+Y\ge 1\right\}\right)$，

$P\left(\left\{\dfrac{X}{Y}\le\dfrac{1}{2}\right\}\right)$.

例 **9** ① 利用概率思想证明：$\dfrac{1}{\sqrt{2\pi}}\int_{-a}^{a}e^{-\frac{x^2}{2}}dx\le\sqrt{1-e^{-a^2}}$.

② 试举出反例说明X,Y是连续型随机变量，但$\left(X,Y\right)$却不是连续型随机向量；且该情形不适用于离散型随机变量 .

以下将要介绍的联合（概率）分布函数是统一刻划各类随机向量（包括联合离散，联合连续等情形）取（向量）值的概率规律一个重要、有效工具！

联合（概率）分布函数：设$\left(X,Y\right)$为二维随机向量，$\forall\left(x,y\right)\in R^2$，称

$$F\left(x,y\right)=P\left(\left\{X\le x,Y\le y\right\}\right)$$

为(X,Y)的联合（概率）分布函数．

① 若(X,Y)是联合离散的，且有：$P\left(\left\{(X,Y)=(x_i,y_j)\right\}\right)=p_{ij},i=1,2,\cdots;j=1,2,\cdots;$ 则有：$\forall(x,y)\in R^2$，$F(x,y)=P\left(\{X\le x,Y\le y\}\right)=P\left(\bigcup_{i:x_i\le x}\bigcup_{j:y_j\le y}\{X=x_i,Y=y_j\}\right)$

$$=\sum_{i:x_i\le x}\sum_{j:y_j\le y}P\left(\{X=x_i,Y=y_j\}\right)=\sum_{i:x_i\le x}\sum_{j:y_j\le y}p_{ij}.$$

例 10 设(X,Y)的联合分布为：$\begin{pmatrix} X/Y & 1 & 2 \\ 1 & \frac{1}{4} & \frac{1}{2} \\ 3 & \frac{1}{8} & \frac{1}{8} \end{pmatrix}$；求$(X,Y)$的联合分布函数．

② 若(X,Y)是联合连续的，且有：$(X,Y)\sim f(x,y)$，则有：$\forall(x,y)\in R^2$

$$F(x,y)=P\left(\{X\le x,Y\le y\}\right)=P\left(\{(X,Y)\in(-\infty,x]\times(-\infty,y]\}\right)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(s,t)\,dt\,ds.$$

例 11 设$(X,Y)\sim f(x,y)=\begin{cases} ke^{-(3x+4y)}, x,y>0; \\ 0,\text{其他}; \end{cases}$，试求常数$k$、$J.d.f.F(x,y)$ 以及 $P\left(\{0<X\le1,0<Y\le2\}\right)$．

例 12 设$r.v.X\sim f(x)=\begin{cases} 0.5,-1<x<0; \\ 0.25,0\le x<2; \\ 0,\text{其中}; \end{cases}$，令$Y=X^2$，$F(x,y)$为二维随机向量$(X,Y)$的联合分布函数，① 求$Y$的$p.d.f.f_Y(y)$；② 求$F\left(-\frac{1}{2},4\right)$．

3.2 边缘分布

若(X,Y)有联合分布：$\begin{pmatrix} X/Y & y_1 & \cdots & y_n \\ x_1 & p_{11} & \cdots & p_{1n} \\ \vdots & \cdots & \cdots & \cdots \\ x_m & p_{m1} & \cdots & p_{mn} \end{pmatrix}$；则有：

$\forall x_i\in R(X),\{X=x_i\}=\{X=x_i\}\cap\Omega=\{X=x_i\}\cap(\bigcup_{j=1}^{n}\{Y=y_j\})=\bigcup_{j=1}^{n}\{X=x_i,Y=y_j\}$，

$\forall y_i\in R(Y),\{Y=y_j\}=\{Y=y_j\}\cap\Omega=\{Y=y_j\}\cap(\bigcup_{i=1}^{m}\{X=x_i\})=\bigcup_{i=1}^{m}\{X=x_i,Y=y_j\}$；

从而，$p_{i\cdot}=P(\{X=x_i\})=\sum_{j=1}^{n}P(\{X=x_i,Y=y_j\})=\sum_{j=1}^{n}p_{ij}$，

$p_{\cdot j}=P(\{Y=y_j\})=\sum_{i=1}^{m}P(\{X=x_i,Y=y_j\})=\sum_{i=1}^{m}p_{ij}$　　；　　$\{p_{1\cdot},p_{2\cdot},\cdots,p_{m\cdot}\}$

（$\{p_{\cdot 1},p_{\cdot 2},\cdots,p_{\cdot n}\}$），即为$(X,Y)$关于$X(Y)$的边缘（边际）分布，它实际上$X(Y)$就是的概率分布！

注：基于上述，由联合分布求边缘分布，只须将上述联合分布矩阵通过行（列）求和即可；但反之则不可！

例 13 ① 设$\begin{pmatrix} X/Y & -1 & 0 & 2 \\ 0 & 0.1 & 0.2 & 0 \\ 1 & 0.3 & 0.05 & 0.1 \\ 2 & 0.15 & 0 & 0.1 \end{pmatrix}$，试求：$P(\{X\neq 0,Y=0\})$，$P(\{X\leq 0,Y\leq 0\})$，

$P(\{XY=0\})$，$P(\{X=Y\})$，$P(\{|X|=|Y|\})$.

② 设$r.v.X\overset{d}{=}Y$，且$X\sim\begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}$，$P(\{XY=0\})=1$；试求$P(\{X\neq Y\})$.

③ 设$X\sim\begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$，$Y\sim\begin{pmatrix} -1 & 0 & 1 \\ \frac{5}{12} & \frac{1}{4} & \frac{1}{3} \end{pmatrix}$；且$P(\{X<Y\})=0$，$P(\{X>Y\})=\frac{1}{4}$，试求$(X,Y)$的联合分布 .

若(X,Y)具有联合密度$f(x,y)$，易知：$\forall a,b\in R,a<b$，

$$P(\{a<X\leq b\})=P(\{a<X\leq b,-\infty<Y<+\infty\})=\iint\limits_{(a,b]\times(-\infty,+\infty)} f(x,y)dxdy$$

$=\int_a^b\left(\int_{-\infty}^{+\infty} f(x,y)dy\right)dx$ ；由a,b的任意性，可知(X,Y)的分量X的概率密度函数为$\int_{-\infty}^{+\infty} f(x,y)dy$，常称之为$(X,Y)$关于$X$的边缘（边际）密度（函数）；记之为：$f_X(x)$.

同理，可给出(X,Y)关于Y的边缘（边际）密度函数：$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx$. 总之，边缘密度可由联合密度唯一确定！

例 14 ① $(X,Y)\sim f(x,y)=\dfrac{1+\sin x\cdot\sin y}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}$，$-\infty<x,y<+\infty$；试求$(X,Y)$关于$X,Y$的边缘密度$f_X(x),f_Y(y)$.

② 设$(X,Y)\sim f(x,y)=\begin{cases} 2e^{-(x+2y)},x,y>0; \\ 0,\text{其他}; \end{cases}$，试求事件$\{X<3\}$的概率 .

例 15 ① 设$f(x,y)=\begin{cases} x^2+\dfrac{xy}{3},(x,y)\in[0,1]\times[0,2]; \\ 0,\text{其他}; \end{cases}$，

试求X,Y的边缘密度$f_X(x),f_Y(y)$及$P\left(Y<\frac{1}{2}\middle|X<\frac{1}{2}\right)$.

② 设$(\xi,\eta)\sim U(D),D=\left\{(x,y)\middle|0<x^2<y<x<1\right\}$ ，试求：ξ的边缘密度函数及

$P\left(\left\{\eta > \frac{1}{2}\right\}\right)$.

例 16 设 $(\xi,\eta) \sim f(x,y) = \begin{cases} \frac{2}{\pi}\left[1-\left(x^2+y^2\right)\right], & x^2+y^2 \le 1; \\ 0, & 其他; \end{cases}$，试求：

① ξ,η 的边缘密度函数；②“向量 (ξ,η) 落入 $x^2+y^2 \le r^2 (0<r<1)$ 中”的概率 .

例 17 设 $(X,Y) \sim f(x,y) = \begin{cases} 1, & 0<x<1, 0<y<2x; \\ 0, & 其他; \end{cases}$，试求：

① $f_X(x), f_Y(y)$ ；② $Z = 2X - Y$ 的概率密度 $f_Z(z)$ ；③ $P\left(Y \le \frac{1}{2} \middle| X \le \frac{1}{2}\right)$.

3.3 条件分布

条件分布是研究随机变量间“相依关系”的一个有力工具，其广泛被应用于概率统计的各个分支 .

设离散型随机变量 X 和 Y 的取值范围分别为 $\{x_1,x_2,\cdots,x_m,\cdots\}$ 和 $\{y_1,y_2,\cdots,y_n,\cdots\}$，且已知二维离散型随机向量 (X,Y) 的联合分布为：

$P\left(\left\{(X,Y)=(x_i,y_j)\right\}\right) = p_{ij}, i=1,\cdots,m,\cdots; j=1,\cdots,n,\cdots;$ 其边缘分布为：

$p_{i.} = P\left(\{X=x_i\}\right) = \sum_{j\ge1} p_{ij}, i=1,\cdots,m,\cdots;$ $p_{.j} = P\left(\{Y=y_j\}\right) = \sum_{i\ge1} p_{ij}, j=1,\cdots,n,\cdots$ ；

$\forall y_j \in R(Y)$，且 $P\left(\{Y=y_j\}\right)>0$，在事件 $\{Y=y_j\}$ 发生的条件下，考虑 X 的概率分布；也即：条件概率 $P\left(X=x_i \middle| Y=y_j\right)$，$i=1,\cdots,m,\cdots$； 若令

$A_j = \{Y=y_j\}$，$B_i = \{X=x_i\}, i\ge1$，则：

$P\left(X=x_i \middle| Y=y_j\right) = P\left(B_i \middle| A_j\right) = \frac{P\left(A_jB_i\right)}{P\left(A_j\right)} = \frac{p_{ij}}{p_{.j}}, i=1,\cdots,m,\cdots;$ 因此，对于固定的 j，

$\left\{P\left(X=x_i \middle| Y=y_j\right), i=1,\cdots,m,\cdots\right\}$ 给出了一个分布，记为：$\left\{P_{X|Y}\left(x_i \middle| y_j\right), i=1,\cdots,m,\cdots\right\}$.

条件分布：设 (X,Y) 是一个离散型随机向量，若 $P\left(\{Y=y_j\}\right)>0$，则对于固定的 j，

$P_{X|Y}\left(x_i \middle| y_j\right) = P\left(X=x_i \middle| Y=y_j\right) = \frac{p_{ij}}{p_{.j}}, i=1,\cdots,m,\cdots$；称之为 X 在给定 $Y=y_j$ 时（或 $\{Y=y_j\}$ 发生的情况下）的条件分布（列）；若 $P\left(\{X=x_i\}\right)>0$，则对固定的 i，$P_{Y|X}\left(y_j \middle| x_i\right)$ $= P\left(Y=y_j \middle| X=y_i\right) = \frac{p_{ij}}{p_{i.}}$， $j=1,\cdots,n,\cdots$；称之为 Y 在给定 $X=x_i$ 时（或 $\{X=x_i\}$ 发生的情况下）的条件分布（列）；常将以上的条件分布列表示为：

$$X \middle| Y=y_j \sim \begin{pmatrix} x_1 & x_2 & \cdots \\ P_{X|Y}\left(x_1 \middle| y_j\right) & P_{X|Y}\left(x_2 \middle| y_j\right) & \cdots \end{pmatrix},$$

$$Y\left|X=x_i\right.\sim\begin{pmatrix} y_1 & y_2 & \cdots \\ P_{Y|X}\left(y_1|x_i\right) & P_{Y|X}\left(y_2|x_i\right) & \cdots \end{pmatrix}.$$

例 **18** ① 设(X,Y)有如下的联合分布：

$$\begin{pmatrix} (X,Y) & (1,1) & (1,4) & (2,1) & (2,2) & (2,3) & (3,2) \\ P & \frac{2}{27} & \frac{1}{27} & \frac{6}{27} & \frac{6}{27} & \frac{6}{27} & \frac{6}{27} \end{pmatrix};$$

试求：给定$X(Y)=1$时，$Y(X)$的条件分布列 .

② 假设有一枚并不均匀的硬币，出现正面的概率为p，将其独立地掷n次，记Y为掷出正面的次数，首次掷出正面是在第X次，求在n次抛掷中仅掷出一次正面的条件下，X的条件概率分布 .

③ 设$X\sim G(p)$，则$\forall n\geq 1$，$X-n\big|_{X>n}\overset{d}{=}X$.

④ 从$1,2,3,4$中任取一个数，记为X；再从$1,\cdots,X$中任取一个数记为Y，试求(X,Y)的联合分布、Y的分布以及给定$Y=3$时X的条件分布 .

设$(X,Y)\sim f(x,y)$即有：$\forall x\in R,P\left(\{X=x\}\right)=0$；从而$\forall a,b\in R,a<b$，无法直接定义$P\left(a<Y<b\left|X=x\right.\right)$；自然会想到：

$$P\left(a<Y\leq b\left|X=x\right.\right)=\lim_{\delta x\to 0}P\left(a<Y\leq b\left|x-\delta x<X\leq x+\delta x\right.\right);$$

设 $(X,Y)\sim f(x,y),x\in\left\{x\left|f_X(x)>0\right.\right\},\forall y\in R$， 则有：

$$P\left(Y\leq y\left|X=x\right.\right)=\lim_{\Delta x\to 0}P\left(Y\leq y\left|x-\Delta x<X\leq x+\Delta x\right.\right)=\lim_{\Delta x\to 0}\frac{P\left(\{x-\Delta x<X\leq x+\Delta x,Y\leq y\}\right)}{P\left(\{x-\Delta x<X\leq x+\Delta x\}\right)}$$

$$=\lim_{\Delta x\to 0}\frac{\int_{-\infty}^{y}\int_{x-\Delta x}^{x+\Delta x}f(s,t)dsdt}{\int_{x-\Delta x}^{x+\Delta x}f_X(s)ds}=\int_{-\infty}^{y}\frac{f(x,t)}{f_X(x)}dt.$$

条件密度函数：设$(X,Y)\sim f(x,y),x\in\left\{x\left|f_X(x)>0\right.\right\}$，称$\dfrac{f(x,y)}{f_X(x)}$为给定 $X=x$时Y的条件密度函数；同理，$y\in\left\{y\left|f_Y(y)>0\right.\right\}$，称$\dfrac{f(x,y)}{f_Y(y)}$为给定$Y=y$时$X$的条件密度函数。

注：① 以上两个条件密度函数分别记为：$f_{Y|X}\left(y|x\right)$ 和 $f_{X|Y}\left(x|y\right)$；即有：

$$f_{Y|X}\left(y|x\right)=\begin{cases}\dfrac{f(x,y)}{f_X(x)}, & f_X(x)>0\\ 0, & f_X(x)=0\end{cases} \text{ 和 } f_{X|Y}\left(x|y\right)=\begin{cases}\dfrac{f(x,y)}{f_Y(y)}, & f_Y(y)>0\\ 0, & f_Y(y)=0\end{cases};$$

常称之为密度的贝叶斯公式 .

② 由定义，$f(x,y)=f_{X|Y}\left(x|y\right)f_Y(y)$，$f(x,y)=f_{Y|X}\left(y|x\right)f_X(x)$，常称之为密度的乘法公式 .

③ 在联合密度$f(x,y)$中，x与y的地位是“对等”的，它们都是变量，故x与y的变化范围等都并列地标注在$f(x,y)$表达式后面；而在条件密度$f_{X|Y}\left(x|y\right)\left[f_{Y|X}\left(y|x\right)\right]$中，

x 与 y 的地位就不对等了：$\{y|f_Y(y)>0\}\left[\{x|f_X(x)>0\}\right]$ 是 $f_Y(y)>0\left[f_X(x)>0\right]$ 的范围，它是 $f_{X|Y}(x|y)\left[f_{Y|X}(y|x)\right]$ 存在的“前提条件”，必须标在前面，$f_{X|Y}(x|y)\left[f_{Y|X}(y|x)\right]$ 只是 $x[y]$ 的函数，故后面只能标注 $x[y]$ 的变化范围，此时的 $y[x]$ 只能视作“常数” .

例 19 ① 设 $(X,Y)\sim f(x,y)=\begin{cases}3x,0<y<x<1;\\ 0,其他;\end{cases}$，试求给定 $X=x$（$0<x<1$）时，Y 的条件密度函数 $f_{Y|X}(y|x)$.

② 设 $(X,Y)\sim f(x,y)=\begin{cases}\dfrac{1}{y}e^{-y}\cdot e^{-\frac{x}{y}},x,y>0;\\ 0,其他;\end{cases}$，$\forall y>0$，试求给定 $Y=y$ 时，X 的条件密度函数 $f_{X|Y}(x|y)$ 及 $P(X>1|Y=y)$.

③ 设 $(X,Y)\sim f(x,y)=\begin{cases}\dfrac{21}{4}x^2y,x^2\le y\le 1;\\ 0,其他;\end{cases}$，试求条件概率 $P(Y\ge 0.75|X=0.5)$.

例 20 ① 设 $X\sim U(0,1)$，已知 $X=x$（$0<x<1$），$Y\sim U\left(0,\dfrac{1}{x}\right)$，试求 Y 的 $p.d.f.f_Y(y)$.

② 设 ξ 在区间 $[0,1]$ 上随机地取值，当观察到 $\xi=x$ 时，η 在区间 $[x,1]$ 上随机地取值，试求 η 的密度函数 .

③ 设 $\xi\sim f_\xi(x)=\begin{cases}\lambda^2xe^{-\lambda x},x>0;\\ 0,x\le 0;\end{cases}$，$\eta$ 在 $(0,\xi)$ 上均匀分布，试求 η 的密度函数 .

④ 设 $Y\sim f_Y(y)=\begin{cases}5y^4,0<y<1;\\ 0,其他;\end{cases}$，给定 $Y=y$（$0<y<1$）时，X 的条件密度为 $f_{X|Y}(x|y)=\begin{cases}\dfrac{3x^2}{y^3},0<x<y;\\ 0,其他;\end{cases}$，试求 $P(X>0.5)$.

⑤ 试讨论二维均匀分布的条件分布，并试求：设 $(X,Y)\sim U(D),D:x^2+y^2\le R^2$，求：$f_{Y|X}(y|x)$ 以及 $P\left(Y>0\middle|X=\dfrac{R}{2}\right)$ 、$P\left(Y>0\middle|X>\dfrac{R}{2}\right)$.

设 $(X,Y)\sim f(x,y),\forall y\in R$，

$$P(\{Y\le y\})=P(\{-\infty<X<+\infty,Y\le y\})=\int_{-\infty}^{+\infty}\int_{-\infty}^{y}f(x,y)dydx$$

$$=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{y}f_{Y|X}(y|x)dy\right]f_X(x)dx=\int_{-\infty}^{+\infty}P(Y\le y|X=x)f_X(x)dx \quad \textbf{(a)} \;;$$

事实上，类似可得到：$P(\{a<Y\le b\})=\int_{-\infty}^{+\infty}P(a<Y\le b|X=x)f_X(x)dx$ **(b)** ；

更一般地，若 $A\in F$ 且与某连续型随机变量 X 有关，则有：

$$P(A)=\int_{-\infty}^{+\infty}P\left(A\middle|X=x\right)f_X(x)dx \qquad \textbf{(c)} \quad ;$$

上述三式均可称作连续型（广义）全概率公式，其实质是离散型全概率公式的连续版本！

例 21 ① X,Y 为两个随机变量，$Y\sim\begin{pmatrix}0 & 1\\ 0.7 & 0.3\end{pmatrix}$，且给定 $Y=k$ 时，$X\sim N\left(k,1^2\right)$，$k=0,1$；试求 X 的分布 .

② 设 $X\sim\begin{pmatrix}1 & 2\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$，且给定时 $X=k$，$Y\sim U(0,k)$，$k=1,2$；试求 Y 的分布，并求 EY .

例 22 ① 在圆周上任取三点 A,B,C，试求 $\{\Delta ABC\text{为锐角三角形}\}$ 的概率 .

② 设 $r.v.X\sim U(0,1)$，$\forall x\in(0,1)$，当 $X=x$ 时，$Y\sim E\left(\frac{x}{2}\right)$，试求 $P(\{X\geq Y\})$.

③ 设 $X\sim B(n,Y)$，其中 $Y\sim U(0,1)$，试求 X 的分布 .

④ 设 $Y\sim U[2,4]$，且给定 $Y=y$（$2\leq y\leq 4$）时，$X\sim E(y)$，试求 (X,Y) 的 $J.p.d.f.$（联合密度函数）并证明：$XY\sim E(1)$.

⑤ 设 $(X,Y)\sim f(x,y)=\begin{cases}\frac{1}{x}, & 0<y<x<1;\\ 0, & \text{其他};\end{cases}$，求 $f_{Y|X}\left(y|x\right)$ 及 $P\left(X^2+Y^2\leq 1\middle|X=x\right)$，$0<x<1$，并由此求 $P\left(\{X^2+Y^2\leq 1\}\right)$.

⑥ 保险公司一项理赔损失 X 具有分布密度 $f(x)=\frac{3}{8}x^2$，$0\leq x\leq 2$，假定处理损失为 x 的一项理赔需要花费的时间 T（单位：月）服从 $U[x,2x]$，则“处理一项理赔的时间需 3 个月以上”的概率是多少 ？

3.4 随机变量的独立性

随机变量的独立性：设 X,Y 是定义于同一概率空间 (Ω,F,P) 的两个随机变量，且 (X,Y) 的联合分布函数为 $F(x,y)$，X,Y 的边缘分布函数分别为 $F_X(x),F_Y(y)$，若 $\forall(x,y)\in R^2$，$F(x,y)=F_X(x)F_Y(y)$ （**a**），则称 X 与 Y 是（相互）独立的 .

注：① 若 X,Y 是联合离散的，则有（**b**）式：

$$X,Y\text{ 独立}\Leftrightarrow\forall x\in R(X),\forall y\in R(Y),P\left(\{X=x,Y=y\}\right)=P\left(\{X=x\}\right)P\left(\{Y=y\}\right) .$$

由（**a**）式，即有：$\forall x,y\in R,P\left(\{X\leq x,Y\leq y\}\right)=P\left(\{X\leq x\}\right)P\left(\{Y\leq y\}\right)$，

1） 先证 $\forall x,y\in R,P\left(\{X<x,Y\leq y\}\right)=P\left(\{X<x\}\right)P\left(\{Y\leq y\}\right)$；

2） $\forall x\in R$，任取数列 $\{x_n,n\geq 1\}$ 使得 $x_1<x_2<\cdots<x_n<\cdots\to x$，从而

$\bigcup_{n=1}^{\infty}(-\infty,x_n]=(-\infty,x)$ ，即有：$\bigcup_{n=1}^{\infty}\{X\le x_n\}=\{X<x\}$，也即有：

$\bigcup_{n=1}^{\infty}\{X\le x_n,Y\le y\}=\{X<x,Y\le y\}$ ；于是， $P(\{X<x,Y\le y\})=$

$P\left(\bigcup_{n=1}^{\infty}\{X\le x_n,Y\le y\}\right)=\lim_{n\to\infty}P(\{X\le x_n,Y\le y\})=\lim_{n\to\infty}P(\{X\le x_n\})P(\{Y\le y\})$

$=P\left(\bigcup_{n=1}^{\infty}\{X\le x_n\}\right)P(\{Y\le y\})=P(\{X<x\})P(\{Y\le y\})$ ；**1**）－**2**），即得：$\forall x,y\in R,P(\{X=x,Y\le y\})=P(\{X=x\})P(\{Y\le y\})$，

重复以上的\作法，即可得**(b)**式！ 再证：（**b**）$\Rightarrow$（**a**）．$\forall x,y\in R$，$P(\{X\le x,Y\le y\})$

$$=P\left(\bigcup_{i:x_i\le x}\{X=x_i,Y\le y\}\right)=P\left(\bigcup_{i:x_i\le x}\bigcup_{j:y_j\le y}\{X=x_i,Y=y_j\}\right)=\sum_{i:x_i\le x}\sum_{j:y_j\le y}P(\{X=x_i,Y=y_j\})$$

$$=\sum_{i:x_i\le x}\sum_{j:y_j\le y}P(\{X=x_i\})P(\{Y=y_j\})=P(\{X\le x\})P(\{Y\le y\})$$ ；即得（**a**）式！

② 若 X,Y 是联合连续的，则有（**c**）式：

X,Y 独立 $\Leftrightarrow f(x,y)=f_X(x)f_Y(y)$（$f(x,y)$的连续点）．

例 23 ① 设(X,Y)有如下联合分布：

X/Y	0	1
0	$\frac{1}{4}$	b
1	a	$\frac{1}{4}$

，且事件$\{X=0\}$与$\{X+Y=1\}$相互独立，试确定常数a,b； X,Y是否独立？

② 设$X\sim\begin{pmatrix}-1&0&1\\ \frac{1}{4}&\frac{1}{2}&\frac{1}{4}\end{pmatrix}$，$Y\sim\begin{pmatrix}0&1\\ \frac{1}{2}&\frac{1}{2}\end{pmatrix}$，如果$P(X^2=Y^2)=1$，试求$(X,Y)$的联合分布；$X,Y$是否独立？

③ 设(ξ,η)有联合分布：$\begin{pmatrix}\xi/\eta&0&1&2\\ -1&\frac{1}{6}&\frac{1}{9}&\frac{1}{18}\\ 1&\frac{1}{3}&a&b\end{pmatrix}$，试求：$a,b$为何值时，$\xi,\eta$独立？

例 24【分布的再生性（可加性）】

1）设$X\sim B(n,p),Y\sim B(m,p)$，且$X,Y$独立，则$X+Y\sim B(n+m,p)$ ．

2）设$X\sim P(\lambda_1),Y\sim P(\lambda_2)$，且$X,Y$独立，则$X+Y\sim P(\lambda_1+\lambda_2)$；并求给定$X+Y=n$时， X 的条件分布 ．

例 25 ① 设X,Y相互独立，且$P(\{X=1\})=0.3,P(\{X=2\})=0.7$，$Y\sim f_Y(y)$，试求

$Z=X+Y$ 的分布 .

② 设 X,Y 相互独立，且 $X\sim\begin{pmatrix}0 & 2\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$，$Y\sim f_Y(y)=\begin{cases}2y, & 0<y<1;\\ 0, & 其他;\end{cases}$；试求 $P(\{Y\le EY\})$ 及 $Z=X+Y$ 的概率密度 .

③ 设 $(X,Y)\sim U(D)$，$D:0<x<1,x^2<y<\sqrt{x}$；令 $U=\begin{cases}1, & X\le Y;\\ 0, & X>Y;\end{cases}$， **1**）写出 (X,Y) 的联合概率密度；**2**）U,X 是否独立？**3**）试求 $Z=U+X$ 的分布函数 $F(z)$.

例 **26** ① 设 $X\sim f(x)=\frac{1}{2}e^{-|x|},-\infty<x<+\infty$,试证：$X$ 与 $|X|$ 不独立 .

② 设 (X,Y) 的 $J.d.f.$ 为：$F(x,y)=\begin{cases}0, & x<0或y<0;\\ xy, & 0\le x<1,0\le y<1;\\ y, & x\ge1,0\le y<1;\\ x, & 0\le x<1,y\ge1;\\ 1, & x\ge1,y\ge1;\end{cases}$；问：$X,Y$ 是否独立？

例 **27** ① 设 $r.v.X,Y$ 独立，且 $X\sim U[0,1]$，$Y\sim E(1)$，试求 $P(\{Y\le X\}),P(\{X+Y\le1\})$.

② 设 $r.v.X,Y$ 独立，且 $X\sim U[0,1]$，$Y\sim E\left(\frac{1}{2}\right)$，试求“方程 $t^2+2Xt+Y=0$ 有实根”的概率 .

例 **27** ① 在区间 $(0,1)$ 内任取两个数，试求“两数之积小于 0.25”的概率 .

② 从长度为 a 的线段的中点两边随机各选取一点，求“两点间距离小于 $\frac{a}{3}$”的概率 .

例 **28** 设随机向量 (X,Y) 的联合密度为 $f(x,y)=\begin{cases}\frac{1+xy}{4},-1<x,y<1;\\ 0,其他;\end{cases}$，试证：$X,Y$ 不独立，但 X^2,Y^2 是独立地 .

注：① 可以验证：常数或退化的随机变量与任意随机变量必独立；

② 若 X,Y 独立，则对于任意的 $Borel$ 函数 $f(\cdot),g(\cdot)$，必有 $f(X),g(Y)$ 独立 ！

3.5 随机向量函数的分布

设 (X,Y) 是二维随机向量，$g(x,y)$ 是二元 $Borel$ 函数，如何去求随机变量 $Z=g(X,Y)$ 的分布呢 ？易见，$\forall z\in R,F_Z(z)=P(\{Z\le z\})=P(\{g(X,Y)\le z\})$ $=P(\{(X,Y)\in D_z\})$，这里，$D_z=\{(x,y)|g(x,y)\le z\}$.

例 **29** ① 设 $r.v.X,Y$ 具有分布：$X\sim\begin{pmatrix}-1 & 0 & 1\\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4}\end{pmatrix}$，$Y\sim\begin{pmatrix}0 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$；已知 $P(XY=0)=1$，试求 $Z=X\vee Y=\max(X,Y)$ 的分布 .

② 设随机变量 X_1,X_2,X_3,X_4 独立同分布，且 $P(X_i=0)=1-P(X_i=\ )=\ \ 0$.

$i=1,2,3,4$，试求行列式 $X=\begin{vmatrix} X_1 & X_2 \\ X_3 & X_4 \end{vmatrix}$ 的概率分布 .

例 **30** ① 设 $r.v.X,Y$ 独立，且 $P(X=i)=\dfrac{1}{3}$，$i=-1,0,1$；$Y\sim U[0,1)$，记：$Z=X+Y$，

试求： $P\left(Z\le\dfrac{1}{2}\middle|X=0\right)$、$Z$ 的 $p.d.f.f_Z(z)$.

② 设 $r.v.X,Y$ 独立，且 $X\sim N(0,1^2)$，$Y\sim\begin{pmatrix} -1 & 1 \\ \dfrac{1}{4} & \dfrac{3}{4} \end{pmatrix}$，试求 $Z=XY$ 的概率分布 .

③ 设 X,Y 独立，且 $X\sim U(0,1)$，Y 的分布函数为 $F_Y(y)$；令 $Z=\begin{cases} Y, & X\le\dfrac{1}{2}; \\ X, & X>\dfrac{1}{2}; \end{cases}$，试求 Z 的分布 .

若 (X,Y) 为二维连续型随机向量，且 $(X,Y)\sim f(x,y)$，$g(x,y)$ 是二元 $Borel$ 函数，则随机变量 $Z=g(X,Y)$ 有分布：

$$\forall z\in R,F_Z(z)=P(\{Z\le z\})=P(\{g(X,Y)\le z\})=P(\{(X,Y)\in D_z\})=\iint_{D_z} f(x,y)dxdy$$

例 **31** ① 设 $(X,Y)\sim U(D)$，$D=\{(x,y)|0\le x\le 2,0\le y\le 1\}$，试求边长为 X,Y 的矩形面积 S 的概率分布 .

② 设 X,Y 独立同 $N(0,1^2)$ 分布，则 $Z=\sqrt{X^2+Y^2}$ 的分布称为瑞利（ $Rayleigh$ ）分布，试求 Z 的密度函数 .

以下，仅就一些简单情形讨论随机向量函数的分布！

设 $(X,Y)\sim f(x,y)$，$g(\cdot,\cdot)$ 为二元 **Borel** 函数，$Z=g(X,Y)$，考虑 Z 的分布；$\forall z\in R$ ，

$$F_Z(z)=P(\{Z\le z\})=P(\{g(X,Y)\le z\})=\iint_{g(x,y)\le z} f(x,y)dxdy \quad ;$$

① $g(x,y)=x+y,Z=g(X,Y)=X+Y$；② $g(x,y)=x-y,Z=g(X,Y)=X-Y$.

① 求 $Z=X+Y$ 的分布：设 $(X,Y)\sim f(x,y)$，则

$$\forall z\in R,F_Z(z)=P(\{Z\le z\})=P(\{X+Y\le z\})=\iint_{x+y\le z} f(x,y)dxdy$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{z-x} f(x,y)dydx=\int_{-\infty}^{+\infty}\int_{-\infty}^{z} f(x,u-x)dudx \quad ;$$

因此， $X+Y\sim f_{X+Y}(z)=\int_{-\infty}^{+\infty} f(x,z-x)dx=\int_{-\infty}^{+\infty} f(z-y,y)dy$；特别地，若 X,Y 相互独立，则有，$X+Y\sim f_{X+Y}(z)=\int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx=\int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)dy$.

② 求 $Z=X-Y$ 的分布：设 $(X,Y)\sim f(x,y)$，则有，$\forall z\in R$，

$$F_Z(z)=P(\{Z\le z\})=P(\{X-Y\le z\})=\iint\limits_{x-y\le z} f(x,y)\,dxdy=\int_{-\infty}^{+\infty}\int_{x-z}^{+\infty} f(x,y)\,dydx$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{z} f(x,x-u)\,dudx=\int_{-\infty}^{z}\int_{-\infty}^{+\infty} f(x,x-u)\,dxdu \quad ;$$

从而有，$X-Y\sim\int_{-\infty}^{+\infty} f(x,x-z)\,dx=\int_{-\infty}^{+\infty} f(z+y,y)\,dy$；特别地，若 X,Y 独立，则有，

$$X-Y\sim\int_{-\infty}^{+\infty} f_X(x)f_Y(x-z)\,dx=\int_{-\infty}^{+\infty} f_X(y+z)f_Y(y)\,dy \quad .$$

例 32 ① 设 X,Y 独立同 $U(0,1)$ 分布，试求 $Z=X+Y$ 的密度 .

② 设 $(X,Y)\sim f(x,y)=\begin{cases}2-x-y,0<x,y<1;\\ 0,\text{其他};\end{cases}$，试求 $Z=X+Y$ 的密度 .

③ 设 X,Y 独立同 $E(1)$ 分布，试求 $Z=X-Y$ 的密度 .

例 33 ① 设 $X_1,X_2,\cdots,X_n$ 独立同 $E(\lambda)$ 分布，试求 $X_1+X_2+\cdots+X_n$ 的分布 .

② 设 $\{X_n,n\ge1\}$ $i.i.d.$ 且 $X_i\sim E(\lambda)$，记 $N=\min\left\{n\left|\sum_{i=1}^{n}X_i>1,n\ge1\right.\right\}$，试求 N 的分布 .

顺（次）序统计量：设 $X_1,X_{2,}\cdots\cdots,X_n$ 独立同分布且有分布函数 $F(\cdot)$，对每个 $\omega\in\Omega$，将 $X_1(\omega),X_2(\omega),\cdots\cdots,X_n(\omega)$ 从小到大排序：$X_{(1)}(\omega)\le X_{(2)}(\omega)\le\cdots\cdots\le X_{(n)}(\omega)$ ；这样，即可得到 n 个新的随机变量 $X_{(1)},X_{(2)},\cdots\cdots,X_{(n)}$ ，称它们为 $X_1,X_{2,}\cdots\cdots,X_n$ 的顺（次）序统计量；其中，$X_{(1)}$（$X_{(n)}$）称为最小（大）顺（次）序统计量，$X_{(i)}$ 称为第 i 个顺（次）序统计量（举例说明定义）.

易见，$X_{(1)}=\min\{X_1,X_{2,}\cdots\cdots,X_n\}$，$X_{(n)}=\max\{X_1,X_{2,}\cdots\cdots,X_n\}$，且有：

$$F_{X_{(n)}}(x)=P(\{X_{(n)}\le x\})=P(\{X_1\le x,X_2\le x,\cdots,X_n\le x\})=\prod_{i=1}^{n}P(\{X_i\le x\})=[F(x)]^n$$

$$F_{X_{(1)}}(x)=P(\{X_{(1)}\le x\})=1-\prod_{i=1}^{n}P(\{X_i>x\})=1-[1-F(x)]^n \quad .$$

例 34 设 $X_1,X_2,\cdots,X_n$ 独立，若 $Y=\max(X_1,X_2,\cdots,X_n)$，$Z=\min(X_1,X_2,\cdots,X_n)$，试求 $Y-Z$ 的分布；这里，$X_i\sim U(0,1),i=1,2,\cdots,n$.

第四章 随机变量的数字特征

4.1 随机变量数学期望

已经知道，离散型随机变量的概率分布完全刻划了随机变量取值的概率（统计）规律，但是常常无须对其描述得那么精细（有时是不能）；从而我们希望用一个或多个实数来描述随机变量取值的某种（些）特征 。

所谓随机变量的数字特征，即指联系于它的概率分布的某个（些）数，如：平均值，最大值等，其反映了随机变量某个（些）方面的特征 . 数学期望（**expectation or expected value**）（又称均值[**mean**]）即是这样一个数，它实际是加权平均概念的推广 .

引例：已知某班的数学成绩（按5分制）统计如下：$\begin{pmatrix} \text{分数} & 1 & 2 & 3 & 4 & 5 \\ \text{人数} & 0 & 5 & 8 & 9 & 9 \end{pmatrix}$；易见该班平均数学成绩为：$\bar{x}=1\times\frac{0}{31}+2\times\frac{5}{31}+3\times\frac{8}{31}+4\times\frac{9}{31}+5\times\frac{9}{31}$；换个角度考虑此问题：现从该班随机地选出一名学生,记其数学成绩为 X ,则其有如下的概率分布；$\begin{pmatrix} X & 1 & 2 & 3 & 4 & 5 \\ P & \frac{0}{31} & \frac{5}{31} & \frac{8}{31} & \frac{9}{31} & \frac{9}{31} \end{pmatrix}$；从而可以认为随机变量 X 的平均值即是它的所有可能取值的加权平均,权重为取相应值的概率;以后我们常常称之为“概率平均值”．

数学期望：设 X 为概率空间 (Ω,F,P) 上的一个离散型随机变量，其概率分布为：

$P(\{X=x_n\})=p_n,n=1,2,\cdots\cdots$；若级数 $\sum\limits_{n\geq1}x_np_n$ 绝对收敛,即：$\sum\limits_{n\geq1}|x_n|p_n$ 收敛，则随机变量 X 的(数学)期望定义为 $\sum\limits_{n\geq1}x_np_n$，并记之为 $E(X(\omega))$，简记为 $E(X)$ 或 EX．

注：① 定义中要求级数绝对收敛，是为了保证求和不受求和次序的影响．

② 一个随机变量的数学期望（若存在）已不再是随机变量，而是一个确定的实数－随机因素已被加权平均掉了；另外，若 X 为仅取非负整值的随机变量，则有：

$$EX=\sum_{n\geq1}nP(\{X=n\})=\sum_{n\geq1}\sum_{k=1}^{n}P(\{X=n\})\overset{\text{交换求和次序}}{=\!=\!=}\sum_{k\geq1}\sum_{n\geq k}P(\{X=n\})$$

$$=\sum_{k\geq1}P(\{X\geq k\})=\sum_{n\geq1}P(\{X\geq n\})\ .$$

例 1 ① 设 $d.r.v.X$ 有分布列：$P\left(\left\{X=(-1)^{j+1}\frac{3^j}{j}\right\}\right)=\frac{2}{3^j},j=1,2,\cdots$， 试求 EX．

② 若 $r.v.X\sim B(n,p)$，$r.v.Y\sim Geo(p)$，$r.v.Z\sim P(\lambda)$，试求 EX 和 EY ,EZ．

③ 设 $X\sim f(x)=\begin{cases}2^{-x}\cdot\ln2, & x>0;\\ 0, & x\leq0;\end{cases}$，对 X 进行独立重复地观测，直到第二个大于3的观测值出现时停止，记 Y 为观测的次数，试求 Y 的分布及 EY．

例 2 ① 设 $d.r.v.X$ 有分布列：$X\sim\begin{pmatrix}-1 & 0 & 1 & 2\\ \frac{1}{4} & \frac{1}{6} & \frac{1}{3} & \frac{1}{4}\end{pmatrix}$； 试求 $Y=X^2,Z=X+1$ 的数学期望 EY 和 EZ．

② 设 $d.r.v.X$ 有分布列：$P(\{X=k\})=\frac{1}{2^{k+1}},k=0,1,\cdots$ ，求 $Y=\cos(\pi X)$ 的数学期望．

注:（一维离散型表示性定理）设 X 为概率空间 $(\Omega，F,P)$ 上的一个离散型随机变量，$g(\cdot)$ 为一元 $Borel$ 函数，令 $Y=g(X)$，则有 Y 的数学期望（若存在）为：

$$EY=\sum_{j:y_j\in R(Y)} y_j P\left(\{Y=y_j\}\right)=\sum_{j:y_j\in R(Y)} y_j P\left(\bigcup_{i:g(x_i)=y_j}\{X=x_i\}\right)$$

$$=\sum_{j:y_j\in R(Y)}\sum_{i:g(x_i)=y_j} g(x_i)P\left(\{X=x_i\}\right)=\sum_{i:x_i\in R(X)} g(x_i)P\left(\{X=x_i\}\right)=E\left[g(X)\right] .$$

为简化数学期望的计算与讨论，以下考虑其性质；设 X 为 (Ω,F,P) 上一随机变量，可以验证：

① $|E(X)|\le E(|X|)$ （若存在）.

② 若 $\exists a,b\in R$, 使得 $a\le X\le b$ ，则 $a\le EX\le b$.

③【线性性质】设 $X_1,X_2,\cdots,X_n$ 是 n 个存在期望的随机变量，$c_0,c_1,\cdots,c_n$ 是 $n+1$ 个实数，则 $E\left(c_0+c_1X_1+\cdots+c_nX_n\right)=c_0+c_1EX_1+\cdots+c_nEX_n$.

由该性质，自然有：

（**a**）常数作为一退化的随机变量其期望即其本身 .

（**b**）对于多维随机向量，即便不知其联合分布，仅知其边缘分布，亦可通过该性质求出其和的数学期望 .

例 **3** ① 求连续独立地掷100颗骰子所得点数之和的数学期望 .

② 将 n 只球独立地放入 M 只盒子中，若每只球放入各只盒子是等可能的，问：平均有多少只盒子有球？

③ 某仪器由 A,B,C 三个元件组成，A,B,C 发生故障的概率分别为0.1，0.2和0.3，问：某时刻平均有多少元件正常工作？

④ 一袋中装有60只黑球和40只红球，现从中任取20只，则平均取到多少只红球？

已经知道离散型随机变量的数学期望即是其所有可能取值的概率平均值；对于连续型随机变量，其密度 $f(x)$ 与 dx 的作用相当于离散型随机变量的分布列 p_i，受此启发：利用微元法的思想，用 $f(x)dx$ 替代 p_i，用积分号替代求和号来定义连续型随机变量的数学期望。

【分析】设连续型随机变量 X 的值域 $R(X)$ 为有限区间 $(a,b]$ 在此区间任意插入 $n-1$ 个分点 $x_1,x_2,\cdots,x_{n-1}$，使得 $a=x_0<x_1<x_2<\cdots<x_{n-1}<x_n=b$， 将 $(a,b]$ 分割成 n 个子区间，其中第 i 个子区间为 $(x_{i-1},x_i]$，$i=1,2,\cdots,n$；相应有：

$$P\left(\{X\in(x_{i-1},x_i]\}\right)=\int_{x_{i-1}}^{x_i} f(x)dx=f(\xi_i)(x_i-x_{i-1}),\xi_i\in(x_{i-1},x_i] \;;$$

若 x_i-x_{i-1} 充分小，则有 $P\left(\{X\in(x_{i-1},x_i]\}\right)\approx f(x_i)(x_i-x_{i-1})$，$i=1,2,\cdots,n$；此时可将 X 近似视作离散型随机变量，且有分布：$X\dot{\sim}\begin{pmatrix}\xi_1 & \xi_2 & \cdots & \xi_n\\ f(\xi_1)\Delta x_1 & f(\xi_2)\Delta x_2 & \cdots & f(\xi_n)\Delta x_n\end{pmatrix}$，

其中：$\Delta x_i=x_i-x_{i-1},i=1,2,\cdots,n$；故有：$EX\approx\sum_{i=1}^{n}\xi_i f(\xi_i)\Delta x_i$ ；注意到，n 越大，和式作为期望的近似值精确度就越高；故可定义：$EX=\lim_{n\to\infty}\sum_{i=1}^{n}\xi_i f(\xi_i)\Delta x_i=\int_a^b xf(x)dx$ ；

去掉最初关于随机变量值域的局限，即有：$EX=\int_{-\infty}^{+\infty}xf(x)dx$.

数学期望：设 X 是一个连续型随机变量，且有密度 $f(x)$，如果 $\int_{-\infty}^{+\infty}|x| f(x)dx<\infty$，则称 $\int_{-\infty}^{+\infty}xf(x)dx$ 为 X 的（数学）期望（均值），记之为：$E(X)$ 或 EX；若 $\int_{-\infty}^{+\infty}|x| f(x)dx=\infty$，则称 X 的数学期望不存在 .

例 4 ① 设 $r.v.X \sim f(x)=\dfrac{1}{\pi(1+x^2)}$，$-\infty<x<+\infty$；试判断 EX 是否存在？

② 试推导均匀分布、指数分布与正态分布的期望 .

例 5 设 $r.v.X \sim f(x)=\begin{cases} a+bx^2, & 0\le x\le 1; \\ 0, & \text{其他}; \end{cases}$，且 $EX=\dfrac{2}{3}$，试求 a,b .

例 6 ① 设 $r.v.X$ 的 $p.d.f.f(x)$ 满足：$f(\mu+x)=f(\mu-x)$，$\forall x\in(-\infty,+\infty)$，其中 μ 为常数，$\int_{-\infty}^{+\infty}|x| f(x)dx$ 收敛；试证明：$EX=\mu$.

② 设 $X \sim f(x)=\begin{cases} 1+x, & -1\le x<0; \\ 1-x, & 0\le x\le 1; \\ 0, & \text{其他}; \end{cases}$，试求 EX .

③ 设 $X \sim f(x)=\dfrac{1}{2}e^{-|x-a|}$，$-\infty<x<+\infty$，试求 EX .

例 7 设 G 为曲线 $y=2x-x^2$ 与 x 轴所围区域，在区域 G 内任取一点，该点到 y 轴的距离为 ξ，求 $E\xi$.

注：（连续型一维表示性定理）设 X 为一个连续型随机变量，$g(\cdot)$ 为一元 $Borel$ 函数，令 $Y=g(X)$，则有 Y 的数学期望（若存在）为：$EY=E\left[g(X)\right]=\int_{-\infty}^{+\infty}g(x)f_X(x)dx$；这里，$X \sim f_X(x)$.

例 8 ① 设 $X \sim U(-1,2)$，$Y=\begin{cases} 1, & X\ge 0; \\ -1, & X<0; \end{cases}$，试求 EY .

② 设 $r.v.X \sim E(1)$，$Y=\min\{X,2\}=X\wedge 2$，试求 EY .

③ 设随机变量 $X \sim N(\mu,\sigma^2)$，则有：$E(|X-EX|)=\sqrt{\dfrac{2}{\pi}}\sigma$.

例 9 ① 游客乘电梯从底层到电视塔顶层观光。电梯于每个整点的第 5 分钟、第 25 分钟和第 55 分钟从底层起行，假设一游客在早八点的第 X 分钟到达底层候梯处，且 X 在 $[0,60]$ 上服从均匀分布，试求该游客的平均等候时间 .

② 国际市场每年对我国某商品的需求量是随机变量 X（吨），且 $X \sim U[2000,4000]$；已知每售出一吨，可挣得外汇 3 千元，但若售不出去，则每吨需支付存储费及其他损失 1 千元，问：需组织多少货源，才能使国家期望收益最大？

注：（二维离散型与连续型表示性定理）若 $g(x,y)$ 是二元 $Borel$ 函数，则随机向量 (X,Y) 的 $Borel$ 函数在以下情形下有：

① (X,Y)有联合分布：$P\left(\left\{(X,Y)=\left(x_i,y_j\right)\right\}\right)=p_{ij}$，$i=1,2,\cdots,m$；$j=1,2,\cdots,n$；

（m,n也可为无穷大），则 $E\left[g(X,Y)\right]$（若存在）

$$=\sum_{i=1}^{m}\sum_{j=1}^{n}g\left(x_i,y_j\right)\cdot P\left(\left\{(X,Y)=\left(x_i,y_j\right)\right\}\right)=\sum_{i=1}^{m}\sum_{j=1}^{n}g\left(x_i,y_j\right)\cdot p_{ij}\quad .$$

② $(X,Y)\sim f(x,y)$，则

$E\left[g(X,Y)\right]$（若存在）$=\iint_{R^2}g(x,y)\cdot f(x,y)dxdy$.

例 **10** ① 设ξ,η独立同分布，且$\xi(\eta)\sim\begin{pmatrix}1&2&3\\ \frac{1}{3}&\frac{1}{3}&\frac{1}{3}\end{pmatrix}$；若$X=\xi\vee\eta,Y=\xi\wedge\eta$，试求

$EX,EY,E(XY)$.

② 设(X,Y)有联合分布：$\begin{pmatrix}X/Y&1&2\\1&0.25&0.32\\2&0.08&0.35\end{pmatrix}$，试求：$E\left(X^2+Y\right)$.

例 **11** ① 设$(X,Y)\sim f(x,y)=\begin{cases}\dfrac{3}{2x^3y^2}, & \dfrac{1}{x}<y<x,x>1;\\ 0, & 其他;\end{cases}$，试求$EY,E\left(\dfrac{1}{XY}\right)$.

② 设$(X,Y)\sim f(x,y)=\begin{cases}3x,0<y<x<1;\\ 0,其他;\end{cases}$，试求：$E(XY),EY,E\left(X^2\right)$.

③ 设$(X,Y)\sim f(x,y)=\begin{cases}\dfrac{x\left(1+3y^2\right)}{4},0<x<2,0<y<1;\\ 0,其他;\end{cases}$，试求$E\left(\dfrac{Y}{X}\right)$.

例 **12** ① 设$(X,Y)\sim U(D)$，其中：$D=\left\{(x,y)\middle|0\le x\le 1,1-x\le y\le 1\right\}$；试求

$Z=X+Y$的数学期望（多种方法）.

② 一商店经销某种商品，每周进货的数量X与顾客对该种商品的需求量Y是相对独立的随机变量，且都服从$U[10,20]$分布；商店每售出一单位商品可得利润1千元；若需求量超过进货量，可从其他商店调剂供应，这时每单位商品可得利润5百元；试计算此商店经销该商品一周的平均利润 .

可以由密度的卷积公式或特征函数验证：独立的正态随机变量的线性组合仍是正态随机变量；这里，不再详述 .

例 **13** ① 设X,Y独立同$N\left(0,1^2\right)$分布，试求$E\left[\max(X,Y)\right]$.

② 设X,Y独立同$N\left(\mu,\sigma^2\right)$分布，试求$E\left[\max(X,Y)\right],E\left[\min(X,Y)\right]$.

4.2 随机变量的方差

已经知道离散型随机变量的数学期望定义为它的概率加权平均值，而以此作为随机变量的某种特征这种作法是否“最为合适”，并且又该如何理解这里的“最为合适”？

设 X 是一个随机变量，且 EX 存在记之为 μ ；对于任一实数 a ，$X-a$ 表示 X 与 a 的离差（误差），$\left(X-a\right)^2$ 表示平方离差（误差），$E\left(X-a\right)^2$ 表示平均平方离差（误差）；$E\left(X-a\right)^2=E\left(X-\mu+\mu-a\right)^2=E\left(X-\mu\right)^2+\left(\mu-a\right)^2\geq E\left(X-\mu\right)^2$ ；等号成立当且仅当 $\mu=a$ ，也即：$\min\limits_{a\in R}E\left(X-a\right)^2=E\left(X-\mu\right)^2$； 这就意味着：若用一个“实数”去代表一个“随机变量”，它的数学期望（若存在）即是使得平均平方误差最小的那个实数；而此时的平均平方误差 $E\left(X-\mu\right)^2$ 也恰好刻划了随机变量在其均值附近取值的分散程度，以后记之为：DX （或 $D(X)$、$VarX$ ），称之为<u>方差</u> ．方差的算术平方根则称为均方差或标准差 ．

注：① DX 的量纲不同于 X,EX ，但 $\sqrt{DX}$ 却与其相同！

② $DX=E\left(X-EX\right)^2=\cdots=EX^2-\left(EX\right)^2$；很多场合，我们还可以方便地通过这个公式而不是定义来计算方差！由一维表示性定理：

（**a**）若 $X-d.r.v.$，且 $P\left(\left\{X=x_i\right\}\right)=p_i,i=1,2,\cdots$；则

$$DX=\sum_{i=1}^{n}\left(x_i-EX\right)^2\cdot p_i=\sum_{i=1}^{n}x_i^2\cdot p_i-\left(EX\right)^2 \quad .$$

（**b**）若 $X-c.r.v.$，且 $X\sim f\left(x\right)$；则

$$DX=\int_{-\infty}^{+\infty}\left(x-EX\right)^2\cdot f\left(x\right)dx=\int_{-\infty}^{+\infty}x^2\cdot f\left(x\right)dx-\left(EX\right)^2 \quad .$$

例 **14** ① 设 $X\sim P\left(\lambda\right),Y\sim G\left(p\right)$，$Z\sim B\left(n,p\right)$，试求 DX,DY,DZ ．

② 设 $X\sim U\left[a,b\right],Y\sim E\left(\lambda\right),Z\sim N\left(\mu,\sigma^2\right)$，试求 DX,DY,DZ ．

例 **15** ① 设 $r.v.X$ 有 $d.f.F\left(x\right)=\begin{cases}\dfrac{1}{2}e^x, & x<0;\\ \dfrac{1}{2}, & 0\leq x<1;\\ 1-\dfrac{1}{2}e^{-\frac{1}{2}(x-1)}, & x\geq 1;\end{cases}$ ，试求 DX ．

② 某类电话的呼唤时间 T 满足：$P\left(\left\{T>t\right\}\right)=a\cdot e^{-\lambda t}+\left(1-a\right)e^{-\mu t},t>0$，$0\leq a\leq 1$，$\lambda,\mu>0$；试求 ET,DT ．

例 **16** ① 设 $\left(X,Y\right)\sim U\left(D\right)$，其中 D 是以 $\left(0,1\right),\left(1,0\right),\left(1,1\right)$ 为顶点的三角形区域，试求 $D\left(X+Y\right)$ ．

② 设 $\left(X,Y\right)\sim f\left(x,y\right)=\begin{cases}8xy, & 0\leq y\leq x\leq 1;\\ 0, & \text{其他};\end{cases}$ ，试求 $DX,DY,E\left(X+Y\right)^2$ ．

注：① 可以验证：随机变量 X 的方差为零当且仅当 X 以概率 **1** 取值为某个常数（数学期望）；也即：$D(X)=0\Leftrightarrow P\left(\left\{X=EX\right\}\right)=1$，或 $X=EX,a.s.$；这是由于：

$$\overline{\{X=EX\}}=\{X\neq EX\}=\{|X-EX|>0\}=\bigcup_{n=1}^{\infty}\left\{|X-EX|\geq\frac{1}{n}\right\};$$

再由概率函数的连续性以及切比雪夫不等式即可得证！

② 若 DX 存在，则有：$D(X+c)=DX, D(aX+b)=D(aX)=a^2DX$.

③ $D(X\pm Y)=DX+DY\pm 2E\left[(X-EX)(Y-EY)\right]$ ；特别地，若 X,Y 独立，
$D(X\pm Y)=DX+DY$ ，$D(aX+bY)=a^2DX+b^2DY$.

例 **17** ① 设 $X\sim B(n,p)$，试求 EX, DX .

② 有 n 个人参加一宴会，进门时他们将各自帽子放在一起，结束时每人随机取一顶帽子；设 ξ 表示恰好取到自己帽子的人数，试求 $E\xi, D\xi$.

4.3 随机变量的协方差与相关系数

【柯西不等式（**Cauchy-Schwarz**）】已经见过的柯西不等式版本有：

① $\left(a_1^2+a_2^2+\cdots+a_n^2\right)\left(b_1^2+b_2^2+\cdots+b_n^2\right)\geq\left(a_1b_1+a_2b_2+\cdots+a_nb_n\right)^2$.

② $f(x),g(x)\in R[a,b], \int_a^b f^2(x)dx\int_a^b g^2(x)dx\geq\left[\int_a^b f(x)g(x)dx\right]^2$.

③ $\alpha,\beta\in V, |(\alpha,\beta)|\leq\|\alpha\|\|\beta\|$，这里，$V$ 是向量空间，$(\cdot,\cdot),\|\cdot\|$ 分别是向量的内积与模 .

④（概率版本）设 $E(X^2),E(Y^2)<+\infty$，则 $\left[E(XY)\right]^2\leq E(X^2)E(Y^2)$.

注：① 设 X 为随机变量，则 $DX=0\Leftrightarrow P(\{X=EX\})=1$（证明）.

② 由上述①可得到柯西（内积）不等式等号成立的条件：

$\forall t\in R, E\left[(tX+Y)^2\right]\geq 0$，$E(X^2)t^2+2E(XY)t+E(Y^2)\geq 0$，从而，

$\left[2E(XY)\right]^2-4E(X^2)E(Y^2)\leq 0$，等号成立当且仅当存在唯一的实数 t_0，使得

$E\left[(t_0X+Y)^2\right]=0$，$E(t_0X+Y)=0$，$D(t_0X+Y)=0$，从而有

$$P\left(\{t_0X+Y=E(t_0X+Y)\}\right)=P\left(\{t_0X+Y=0\}\right)=1,$$

（置于向量空间【线性相关，内积】的范畴来解释等号成立的充要条件）.

协方差：设 (X,Y) 为二维随机向量，若 $E|(X-EX)(Y-EY)|<\infty$，则称

$E\left[(X-EX)(Y-EY)\right]$ 为 随机变量 X 与 Y 的协方差（**Covariance**），记作：

$Cov(X,Y)$，即：$Cov(X,Y)=E\left[(X-EX)(Y-EY)\right]$.

注：由期望的线性性质，即有：$Cov(X,Y)=E(XY)-EXEY$ ；由上述定义，我们有：

① 若 X,Y 是联合离散的，且有联合分布列：

$P\left(\{(X,Y)=(x_i,y_j)\}\right)=p_{ij}, i=1,2,\cdots; j=1,2,\cdots$；则

$$Cov(X,Y)=\sum_i\sum_j (x_i-EX)(y_j-EY)p_{ij} \quad .$$

② 若 X,Y 是联合连续的，且有联合分布密度 $f(x,y)$ ；则

$$Cov(X,Y)=\iint_{R^2}(x-EX)(y-EY)f(x,y)dxdy \quad .$$

若 $Cov(X,Y)=0$，称 X,Y 不相关或零相关！

通常，协方差 $Cov(X,Y)$ 有如下一些性质：

①（对称性）$Cov(X,Y)=Cov(Y,X)$；　② $Cov(X,X)=DX$ ；

③ $Cov(aX,bY)=abCov(X,Y)$ ；　④ $Cov(X+c,Y+d)=Cov(X,Y)$ ；

⑤ $Cov(X,Y+Z)=Cov(X,Y)+Cov(X,Z)$；⑥ $Cov(X,a)=0,\ \in a \quad co$ ；

⑦ $D(X\pm Y)=DX+DY\pm 2Cov(X,Y)$ ，并且更一般地，

$$D\left(\sum_{i=1}^n a_iX_i\right)=\sum_{i=1}^n a_i^2DX_i+2\sum_{1\le i<j\le n}a_ia_jCov(X_i,X_j) \quad .$$

例 18 ① 设 (X,Y) 有联合分布：$\begin{pmatrix} X/Y & 0 & 1 & 2 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} \\ 1 & 0 & \frac{1}{3} & 0 \\ 2 & \frac{1}{12} & 0 & \frac{1}{12} \end{pmatrix}$，试求：$Cov(X-Y,Y)$.

② 设 $X\sim U[-1,1]$，$Y=g(X)$，其中：$g(x)=\begin{cases}1, & x\ge 0; \\ -1, & x<0;\end{cases}$，试求：$Cov(X,Y)$.

③ 设 $(X,Y)\sim U(D)$，其中 $D=\left\{(x,y)\,\middle|\,x^2+y^2\le 1\right\}$，试证：$X,Y$ 既不相关也不独立 .

例 19 ① 设 $X\sim f(x)$（偶函数），且 $E\left(|X|^3\right)<\infty$，试证：$X$ 与 $Y=X^2$ 不相关也不独立 .

② 设 $(X,Y)\sim f(x,y)=\begin{cases}\frac{1}{2}, & -1\le x\le 1, |y|\le|x|; \\ 0, & \text{其他};\end{cases}$，试求 $D(2X-3Y+4)$.

注： 协方差的数值虽然一定程度上反映了 X 与 Y 之间的相互关系，但它仍然受 X,Y 本身数值（量纲）大小的影响，再如：直观上，kX 与 kY（k 为常数）之间的联系和 X 与 Y 之间的联系并无二致，但 $Cov(kX,kY)=k^2Cov(X,Y)$，协方差却增加了 k^2 倍；为了克服以上不足，先对随机变量"标准化"．所谓"标准（化）随机变量"，即指 $r.v.X$ 满足 $EX=0,DX=1$ ；对于任何随机变量 X ，若 $EX=\mu,DX=\sigma^2$ ，只须令：$Y=\dfrac{X-EX}{\sqrt{DX}}=\dfrac{X-\mu}{\sigma}$ ，则有 $EY=0,DY=1$，于是 ，则称 Y 是 X 的标准化，这也是 σ 被称为"标准差"的一个原因 .

（线性）相关系数：设(X,Y)是一个二维随机向量，若DX,DY存在，则称

$$Cov\left(\frac{X-EX}{\sqrt{DX}},\frac{Y-EY}{\sqrt{DY}}\right)=E\left[\left(\frac{X-EX}{\sqrt{DX}}\right)\left(\frac{Y-EY}{\sqrt{DY}}\right)\right]=\frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DX}}$$

为X与Y的（线性）相关系数，记之为：$\rho,\rho_{X,Y},\rho(X,Y)$.

注：① 相关系数消除了计量单位对随机变量的影响，因此它比协方差更能精确地刻划随机变量之间的关系 .

② 经过标准化处理的$r.v.X^*=\dfrac{X-EX}{\sqrt{DX}}$是把原分布中心$EX$移至原点，不使分布中心偏左或偏右，然后缩小或扩大坐标轴，使分布不致过疏或过密；在排除这些干扰后，原随机变量的一些性质就会显现出来，故标准化处理技术在概率统计中会经常使用 .

由定义，相关系数具有如下性质：

①（对称性）$\rho(X,Y)=\rho(Y,X)$；　② $\rho(X,Y)=\rho(aX,bY),ab>0$　；

③ $\rho(X,Y)=\rho(X+c,Y+d)$　；　④ $|\rho(X,Y)|\le 1$　.

注：性质④可由柯西不等式加以验证：$|\rho(X,Y)|=\left|Cov\left(\dfrac{X-EX}{\sqrt{DX}},\dfrac{Y-EY}{\sqrt{DY}}\right)\right|=$

$$\left|E\left[\left(\frac{X-EX}{\sqrt{DX}}\right)\left(\frac{Y-EY}{\sqrt{DY}}\right)\right]\right|\le\sqrt{E\left(\frac{X-EX}{\sqrt{DX}}\right)^2E\left(\frac{Y-EY}{\sqrt{DY}}\right)^2}=1$$

，其等号成立的条件为等价于柯西不等式等号成立的条件，其也可通过方差的性质得到验证 .

可以验证以下四个命题等价：① $Cov(X,Y)=0$ ；　② $\rho(X,Y)=0$ ；

③ $E(XY)=EXEY$ ；　④ $D(X\pm Y)=DX+DY$.

注：① 若$r.v.X,Y$的方差存在，其相关系数为ρ_{XY}，则：$|\rho_{XY}|=1\Leftrightarrow X$与$Y$以概率1线性相关，即存在非零实数$a$与实数$b$，使得$P(\{Y=aX+b\})=1$ ；易见，

$|\rho_{XY}|=1\Leftrightarrow\left|E\left(\dfrac{X-EX}{\sqrt{DX}},\dfrac{Y-EY}{\sqrt{DY}}\right)\right|=\sqrt{E\left(\dfrac{X-EX}{\sqrt{DX}}\right)^2E\left(\dfrac{Y-EY}{\sqrt{DY}}\right)^2}\Leftrightarrow$ 存在非零实数a和b，使得 $P\left(\left\{\dfrac{Y-EY}{\sqrt{DY}}=a\dfrac{X-EX}{\sqrt{DX}}\right\}\right)=1\Leftrightarrow P(\{Y=aX+b\})=1$ ，这里，$b=EY-\dfrac{aEY\sqrt{DY}}{\sqrt{DX}}$.

② $|\rho_{XY}|=1$时，即除去一个零概率事件外，X与Y之间存在着一个线性关系；$\rho_{XY}=1$时，称X与Y完全正相关；$\rho_{XY}=-1$时，称X与Y完全负相关；$\rho_{XY}=0$时，称X与Y不（零）相关；因此，相关系数是描述随机变量间线性关系强弱的一个数字特征，确切地应称为线性相关系数 .

③ 独立与不相关的比较：若 $r.v.X,Y$ 独立 $\Rightarrow E(XY)=EXEY\Rightarrow Cov(X,Y)=0\Rightarrow$ $\rho(X,Y)=0$；即：独立必不相关；反之则不成立！这是由于X,Y不相关，只是说明X,Y之间“不存在”线性关系，但却可能存在别的函数关系，此时就可能不再相互独立 .

例 20 ① 设$(X,Y)\sim U(D)$，$D=[0,2]\times[0,1]$；记：$U=\begin{cases}0, & X\le Y;\\ 1, & X>Y;\end{cases}$，

$V=\begin{cases}0, & X\le 2Y;\\ 1, & X>2Y;\end{cases}$；求$\rho(U,V)$.

② 设$(\xi,\eta)\sim f(x,y)=\begin{cases}2-x-y, & 0<x,y<1;\\ 0, & \text{其他};\end{cases}$，试求：$E\xi,E\eta,D\xi,D\eta,Cov(\xi,\eta)$ $,\rho(\xi,\eta)$.

例 21 设$\xi\sim P(\lambda)$，$\eta\sim U[\lambda,\lambda+2]$，且$\rho(\xi,\eta)=-\dfrac{1}{2}$；令$X=\dfrac{\xi}{3}+\dfrac{\eta}{2}$，试求$EX,DX$.

例 22 ① 设$r.v.X\sim U\left(-\dfrac{1}{2},\dfrac{1}{2}\right)$，$Y=\cos X$，则$Y$与$X$有（非线性）函数关系，则$X,Y$不（线性）相关，即无线性关系 .

② 设 $r.v.X\sim N(0,1^2)$，$Y\sim\begin{pmatrix}-1 & 1\\ \dfrac{1}{2} & \dfrac{1}{2}\end{pmatrix}$，且$X,Y$独立，令$Z=XY$，试证：$Z\sim N(0,1^2)$，且 X,Z不相关也不独立 .

例 23 ① 设$r.v.X_1,X_2,\cdots,X_n$中任意两个的相关系数都是ρ，试证：$\rho\ge-\dfrac{1}{n-1}$.

② 设二维随机向量(X,Y)满足：$EX=EY=0$，$DX=DY=1$，$Cov(X,Y)=\rho$，试证：

$$E\left[\max\left(X^2,Y^2\right)\right]\le 1+\sqrt{1-\rho^2}\ .$$

例 24 ① 将一枚均匀硬币重复掷n次，记X,Y分别为正面朝上、反面朝上的次数，试求X,Y的协方差与相关系数 .

② 设随机变量X,Y满足$D(2X+Y)=0$，试求X,Y的相关系数ρ_{XY} .

例 25 设$DX,DY>0$，$\rho=\rho(X,Y)$，则 $\min\limits_{a,b\in R}E\left[Y-(aX+b)\right]^2=DY\cdot(1-\rho^2)$；由此亦可见相关系数$\rho$刻划了随机变量$X,Y$的线性关系的强弱 .

二元（维）正态分布：若$(X,Y)\sim f(x,y)=$

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1{}^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2{}^2}\right]\right\}$$，其中，$-\infty<x,y<+\infty(\sigma_1,\sigma_2,>0,\mu_1,\mu_2\in R)$；则称$(X,Y)$服从参数为$(\mu_1,\mu_2;\sigma_1{}^2,\sigma_2{}^2;\rho)$的二元（维）正态分布，记作：$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2{}^2;\rho)$.

注：二元（维）正态分布有如下的典型分解式：

① $$f(x,y)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1{}^2}}\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}e^{-\frac{\left[y-\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right)\right]^2}{2\sigma_2{}^2(1-\rho^2)}}$$；

② $f(x,y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}e^{-\frac{\left[x-\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right)\right]^2}{2\sigma_1^2(1-\rho^2)}}$.

由以上典型分解式，易得二元正态分布的性质：若$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1,\sigma_2;\rho)$，则

① 两个边缘分布：$X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$.

② 两个条件分布：$Y\mid X=x\sim N\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),(1-\rho^2)\sigma_2^2\right)$ ；

$X\mid Y=y\sim N\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),(1-\rho^2)\sigma_1^2\right)$ ；由上述 ②，可知，

$E(X|Y=y)=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)$ ， $E(Y|X=x)=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$.

注：设$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1,\sigma_2;\rho)$，讨论$\rho$的意义；$X,Y$的相关系数为：

$$\frac{Cov(X,Y)}{\sqrt{DXDY}}=\frac{E(XY)-\mu_1\mu_2}{\sigma_1\sigma_2}=\frac{E\left[E(XY|X)\right]-\mu_1\mu_2}{\sigma_1\sigma_2}=\frac{E\left[XE(Y|X)\right]-\mu_1\mu_2}{\sigma_1\sigma_2}$$

$=\cdots\cdots=\rho$ ；若令二元正态密度中的$\rho=0$，即有：$f(x,y)=f_X(x)f_Y(y)$，从而X,Y独立；故在二元正态分布场合，独立与不相关等价！

例 26 ① 设随机向量(X,Y)服从二元正态分布 $f(x,y)=\frac{1}{2\pi ab}e^{-\frac{1}{2}\left(\frac{1}{a^2}x^2+\frac{1}{b^2}y^2\right)}$ ，试求

“(X,Y)取值于椭圆$\frac{1}{a^2}x^2+\frac{1}{b^2}y^2=k^2$内”的概率 .

② 设随机向量 $(X,Y)\sim N(\mu_1,\mu_2;a^2,b^2;\rho)$，求“ (X,Y)取值于椭圆

$\frac{1}{a^2}(x-\mu_1)^2-\frac{2\rho}{ab}(x-\mu_1)(y-\mu_2)+\frac{1}{b^2}(y-\mu_2)^2=k^2$ 内”的概率 .

例 27 ① 若随机向量$(X,Y)\sim f(x,y)=ke^{-(ax^2+2bxy+cy^2)},-\infty<x,y<+\infty$，在什么条件下，$X$ 与Y相互独立?

② 试问在何条件下，函数 $f(x,y)=k\exp\left\{-\left(ax^2+2bxy+cy^2\right)\right\}$为某二维随机向量的联合密度函数?

4.4 矩与协方差矩阵（见教材）

第五章 大数定律与中心极限定理

5.1 大数定律

马尔可夫（**Markov**）不等式

（一）若非负随机变量X 的期望EX 存在，则$\forall\varepsilon>0$ ，

① 若 X 是离散型随机变量且有分布列：$P\left(\{X=x_i\}\right)=p_i, i=1,2,\cdots$，

$$EX=\sum_i x_i P\left(\{X=x_i\}\right)\ge\sum_{i:x_i\ge\varepsilon}x_iP\left(\{X=x_i\}\right)\ge\sum_{i:x_i\ge\varepsilon}\varepsilon P\left(\{X=x_i\}\right)\ge\varepsilon P\left(\bigcup_{i:x_i\ge\varepsilon}\{X=x_i\}\right)$$

$$=\varepsilon P\left(\{X\ge\varepsilon\}\right)$$ ；也即：$P\left(\{X\ge\varepsilon\}\right)\le\dfrac{EX}{\varepsilon}$.

② 若 X 是连续型随机变量且有密度函数 $f(x)$，

$$EX=\int_{-\infty}^{+\infty}xf(x)dx=\int_0^{+\infty}xf(x)dx\ge\int_{\varepsilon}^{+\infty}xf(x)dx\ge\int_{\varepsilon}^{+\infty}\varepsilon f(x)dx=\varepsilon P\left(\{X\ge\varepsilon\}\right)$$；也即：

$$P\left(\{X\ge\varepsilon\}\right)\le\frac{EX}{\varepsilon}$$.

（二）若随机变量 X 满足：$E\left(|X|^k\right)<+\infty$ ，则 $\forall\varepsilon>0$，即有：

$$P\left(\left\{|X|^k\ge\varepsilon^k\right\}\right)\le\frac{E\left(|X|^k\right)}{\varepsilon^k}$$ ，也即：$$P\left(\{|X|\ge\varepsilon\}\right)\le\frac{E\left(|X|^k\right)}{\varepsilon^k}$$.

切比雪夫（**Chebyshev**）不等式：将马尔可夫不等式中的 $|X|^k$ 置换成 $(X-c)^2$，即有：

$$P\left(\{|X-c|\ge\varepsilon\}\right)\le\frac{E\left(|X-c|^2\right)}{\varepsilon^2}$$ ；

它有如下常见的形式：

① $P\left(\{|X-c|\ge\varepsilon\}\right)\le\dfrac{E\left((X-c)^2\right)}{\varepsilon^2}$.

② 取 $c=EX$ ，即有：$P\left(\{|X-EX|\ge\varepsilon\}\right)\le\dfrac{E\left((X-EX)^2\right)}{\varepsilon^2}=\dfrac{DX}{\varepsilon^2}$.

③ 取 $c=EX$ ， $\varepsilon=k\sigma=k\sqrt{DX}$ ，即有：$P\left(\{|X-EX|\ge k\sigma\}\right)\le\dfrac{1}{k^2}$.

注：切比雪夫不等式常用来：① 证明切比雪夫大数定律；② 估算概率的上、下界 .

例 1 ① 设 $r.v.X$ 满足：$EX=11$，$DX=9$，试由切比雪夫不等式估计 $P(2<X<20)$.

② 设 $r.v.X,Y$ 满足：$EX=EY=2$，$DX=1, DY=4$，且 $\rho_{XY}=0.5$，试由切比雪夫不等式估计 $P(|X-Y|\ge6)$.

例 2 设 $r.v.X\sim f(x)=\begin{cases}\dfrac{x^m}{m!}e^{-x}, & x>0;\\ 0, & x\le0;\end{cases}$，$m\in\mathrm{N}$，试证：$P\left(0<X<2(m+1)\right)\ge\dfrac{m}{m+1}$.

极限定理包括大数定律（法则）（弱大数定律与强大数定律）和中心极限定理（局部中心极限定理与积分中心极限定理）. 通俗地说，凡描述在一定的条件下随机序列的前若干项的算术平均值收敛于其均值的算术平均值的一类定理都可统称为大数定理；凡描述在一定的条件下，大量随机变量之和的极限分布是正态分布的一类定理都可统称为中心极限定理！

依概率收敛：设$\{X_n, n\ge 1\}$为概率空间(Ω, F, P)上的随机序列，若存在随机变量X，使得：$\forall \varepsilon>0, \lim\limits_{n\to\infty} P\left(\{|X_n - X|\ge\varepsilon\}\right)=0$，或等价地$\lim\limits_{n\to\infty} P\left(\{|X_n - X|<\varepsilon\}\right)=1$，则称随机序列$\{X_n, n\ge 1\}$依概率收敛于随机变量$X$；记作：$\lim\limits_{n\to\infty} X_n = X(P)$，$P-\lim\limits_{n\to\infty} X_n = X$ 或 $X_n \xrightarrow{P} X$；若$X_n \xrightarrow{P} X$，即有：$X_n - X \xrightarrow{P} 0$.

注：$X_n \xrightarrow{P} X$也即$\forall\varepsilon>0$，n充分大时，事件$\{|X_n - X|\ge\varepsilon\}$发生的概率很小（收敛为$0$），这是在概率意义下的收敛性；因此，$n$很大时，有很大把握保证$X_n$很接近于$X$.

例 3 ① 设$\{\xi_n, n\ge 1\}$独立同$U[0,a]$分布，令$X_n = \max\{\xi_1, \xi_2, \cdots, \xi_n\}, n\ge 1$，则$\{X_n, n\ge 1\}$依概率收敛于$a$，即：$X_n \xrightarrow{P} a$.

② 设$X_1, X_2, \cdots, X_n, \cdots, i.i.d.$，其$p.d.f.$为$f(x) = e^{-(x-a)} I_{\{x>a\}}(x) = \begin{cases} e^{-(x-a)}, & x>a; \\ 0, & x\le a; \end{cases}$，令$Y_n = \min\{X_1, X_2, \cdots, X_n\}, n\ge 1$，试证：$\{Y_n, n\ge 1\}$依概率收敛于$a$，即：$Y_n \xrightarrow{P} a$.

③ 设随机变量序列$\{X_n, n\ge 1\}$独立同分布，且$EX_n = \mu$，$DX_n = \sigma^2$存在，若$\forall n\ge 1$，令$Y_n = \dfrac{2}{n(n+1)}\sum\limits_{k=1}^{n} kX_k$，即有随机序列$\{Y_n, n\ge 1\}$依概率收敛为$\mu$，也即：$Y_n \xrightarrow{P} \mu$.

切比雪夫大数定律： 设$\{X_n, n\ge 1\}$为$i.i.d$随机序列，且$EX_i = \mu, DX_i = \sigma^2$，记$S_n = \sum\limits_{i=1}^{n} X_i, \overline{X} = \dfrac{1}{n} S_n$，则$\forall\varepsilon>0, \lim\limits_{n\to\infty} P\left(\{|\overline{X} - \mu|\ge\varepsilon\}\right)=0$，即：$\overline{X} \xrightarrow{P} \mu$；有时，也将该定律的条件写为：$\{X_n, n\ge 1\}$为独立随机序列，若存在常数$c$，使得$\forall n, DX_n \le c$，即$X_n$的方差一致有界，则$\forall\varepsilon>0, \lim\limits_{n\to\infty} P\left(\left\{\left|\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i - \dfrac{1}{n}\sum\limits_{i=1}^{n} EX_i\right|\ge\varepsilon\right\}\right)=0$；若随机序列$\{X_n, n\ge 1\}$满足该式，则称它服从（弱）大数定律！

注：当试验次数趋于无穷时，$\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$依概率收敛于其期望，这也是对算术平均值稳定性的较确切的解释；它说明大数次重复试验下所呈现的客观规律，故称大数定律！

贝努利（**Bernoulli**）大数定律：设n_A表示在n重 **Bernoulli** 试验中事件A发生的频数，且每次试验A发生的概率为p，则$\forall\varepsilon>0$，有：$\lim\limits_{n\to\infty} P\left(\left\{\left|\dfrac{n_A}{n} - p\right|\ge\varepsilon\right\}\right)=0$，即：$\dfrac{n_A}{n} \xrightarrow{P} p$.

注：频率依概率收敛于概率；这也说明在大数次重复试验下，用频率去估计概率是可行的 .

以上的定律都是假定随机变量的二阶矩存在且方差一致有界的前提下成立大数定律；然而有些前提却是不必要的，

辛钦大数定律：设$\{X_n, n\ge 1\}$为$i.i.d.$随机序列，且$EX_n = \mu$，则

$$\forall\varepsilon>0,\lim_{n\to\infty}P\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|\ge\varepsilon\right\}\right)=\quad，即：\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}\mu\quad.$$

例 4 ① 设$\{X_n,n\ge1\}$为独立的随机序列，且$P\left(X_n=\pm\sqrt{\ln n}\right)=\dfrac{1}{2}$，$n=1,2,\cdots$；则 $\{X_n,n\ge1\}$服从大数定律 .

② 设$\{X_n,n\ge1\}$为独立随机序列，且$P\left(X_n=\pm2^n\right)=\dfrac{1}{2^{2n+1}}$，$P(X_n=0)=1-\dfrac{1}{2^{2n}}$，$n=1,2,\cdots$；则$\{X_n,n\ge1\}$服从大数定律 .

例 5 ① 设$\{X_n,n\ge1\}$为独立的随机序列，且$X_n\sim P\left(\sqrt{n}\right)$，$n=1,2,\cdots$；试问：$\{X_n,n\ge1\}$是否服从大数定律？

② 设随机序列 $\{X_n,n\ge1\}$ 独立同分布，且有分布函数 $F(x)=\dfrac{1}{2}+\dfrac{1}{\pi}\arctan\dfrac{x}{a}$ ，$-\infty<x<+\infty$；试问：辛钦大数定律对此随机序列是否适用？

例 6 设随机序列$\{X_n,n\ge1\}$独立同$E(2)$分布，则当$n\to\infty$时，$\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i,\dfrac{1}{n}\sum_{i=1}^{n}X_i^2$，$\dfrac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ 分别依概率收敛于什么 ？

5.2 中心极限定理

依分布收敛：设$F_n(x),n\ge1,F(x)$分别为$X_n,n\ge1,X$ 的分布函数，若对于$F(x)$的任一连续点x，有 $\lim_{n\to\infty}F_n(x)=F(x)$，则称随机序列$\{X_n,n\ge1\}$依分布收敛于$X$，$F(x)$为$\{F_n(x),n\ge1\}$的极限分布函数 .

例 7 设 $\forall n\ge1,X_n\sim\begin{pmatrix}\frac{1}{n}&\frac{2}{n}&\cdots&1\\\frac{1}{n}&\frac{1}{n}&\cdots&\frac{1}{n}\end{pmatrix}$ ，则 $X_n\xrightarrow{d}X\sim U(0,1)$.

例 9 设 $\{X_n,n\ge1\}$ 独立同 $\begin{pmatrix}0&1\\\frac{1}{2}&\frac{1}{2}\end{pmatrix}$ 分布，$\forall n\ge1$，$Y_n=\sum_{k=1}^{n}\dfrac{X_k}{2^k}$；则$\{Y_n,\forall n\ge1\}$依分布收敛于$Y$，且 $Y\sim U(0,1)$.

例 9 设 $X_1,X_2,\cdots,X_n,\cdots,i.i.d.$ ，且$X_i\sim f(x)=\begin{cases}2x,&0<x<1\\0,&else\end{cases}$ ；$\forall n\ge1$，记：$Z_n=n\left[1-\max(X_1,X_2,\cdots,X_n)\right]$的分布函数为$F_n(x)$，试证：$\lim_{n\to\infty}F_n(x)=F(x)$，$-\infty<x<+\infty$，其中$F(x)$是参数为2的指数分布函数 ！

列维－林德伯格中心极限定理：设$\{X_n, n\geq 1\}$为$i.i.d.$随机序列，且$EX_n=\mu, DX_n=\sigma^2$，$n\geq 1$，则$\forall x\in R$，有

$$\lim_{n\to\infty} P\left(\left\{\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}\leq x\right\}\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x)\quad ;$$

也即：$\dfrac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$的极限分布是标准正态分布；此时也称随机序列服从中心极限定理！

注：定理表明：独立同分布的随机变量之和的标准化变量（规范和）依分布收敛于标准正态变量！定理既验证了“只要独立同分布的随机变量方差存在，无论其原来是何分布，其极限分布均是正态分布”，从而在理论上支持了正态分布的重要性，初步说明了为什么实际应用中会经常遇到正态分布；又提供了一种“计算独立同分布随机变量和的分布的近似方法”，从而实际应用时十分简洁有效；只要和式中相加各项的个数充分地大，即可不必关心每个随机变量原先为何分布，都可用正态分布逼近！

将该定理应用到 **Bernoulli** 试验的背景，即有：

德莫佛－拉普拉斯中心极限定理：设在n重 **Bernoulli** 试验中，事件A在每次试验中出现的概率为p $(0<p<1)$，n_A为n次试验中A出现的次数，则，

$$\forall x\in R, \lim_{n\to\infty} P\left(\left\{\frac{n_A-np}{\sqrt{np(1-p)}}\leq x\right\}\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x)\quad .$$

注：大数定律和中心极限定理都是描述随机变量的“和”的分布的极限过程，但它们却有质的区别：大数定律只刻划了“频率稳定于概率”，“平均值稳定于期望值”，即随机序列算术平均值的取值发展的趋向，属于定性的描述；而中心极限定理肯定地是在某些条件下，随机变量之和服从正态分布，即给出了过程的定量的描述，其在实用上往往具有更大的价值（大样本统计推断的理论基础）.

例 **10** 设随机序列$\{X_n, n\geq 1\}$独立同参数为$\frac{1}{2}$的$0-1$分布，若

$$\lim_{n\to\infty} P\left(\frac{C\sum_{i=1}^{n}\left(X_{2i}-X_{2i-1}\right)}{\sqrt{n}}\leq x\right)=\Phi(x)$$，试求常数 C .

例 **11** 用概率论的方法证明：$\lim_{n\to\infty}\left(1+n+\frac{n^2}{2!}+\cdots+\frac{n^n}{n!}\right)e^{-n}=\frac{1}{2}$.

例 **12** ① 某餐厅每天接待400名顾客，假设每位顾客的消费额（元）服从$(20,100)$上的均匀分布，且顾客的消费额是相互独立的，试求：

（**i**）该餐厅每天的平均营业额；

（**ii**）“该餐厅每天的营业额在平均营业额± 760内”的概率 .

② 设某种福利彩票的奖金额 X 由摇奖决定，其分布列为：

$$X \sim \begin{pmatrix} 5 & 10 & 20 & 30 & 40 & 50 & 100 \\ 0.2 & 0.2 & 0.2 & 0.1 & 0.1 & 0.1 & 0.1 \end{pmatrix},$$

若一年中要开出 300 个奖，试问需要多少奖金总额，才有 95% 的把握能够发放奖金？

例 13 ① 独立重复地对某物体的长度 l 进行 n 次测量，假设每次测量的结果 X_i 服从正态分布 $N\left(l, 0.2^2\right)$；记 $\overline{X}$ 为 n 次测量结果的算术平均值，为保证有 95% 的把握使平均值与实际值 l 的差异小于 0.1，试问至少需要测量多少次？

② 设有 2500 个同一年龄段和同一社会阶层的人参加了某保险公司的人寿保险，假设在一年中每个人死亡的概率为 0.002，每个人在年初向保险公司缴纳保费 1200 元，而在死亡时保险受益人可以从保险公司领到保险金 200000 元，问：

（**i**）“保险公司亏本”的概率是多少？

（**ii**）“保险公司获利不少于 1000000 元”的概率是多少？

例 14 ① 一复杂系统由 100 个相互独立工作的部件组成，每个部件正常工作的概率为 0.9；已知整个系统中至少有 85 个部件正常工作，系统才能正常工作，试求“系统正常工作”的概率 **.**

② 某车间有同型号的机床 200 台，在一小时内每台机床约有 70% 的时间是工作的；假定各机床工作是相互独立的，工作时每台机床要消耗电能 $15kW$，问：至少需要多少电能，才可以有 95% 的可能性保证此车间正常生产？