随机过程

马啸

2025 年 9 月 16 日

目录

第 0 章

概率论回顾

§0.1 概率论的基本概念

0.1.1 随机模型三要素

概率论是研究随机现象并揭示其规律的学科。为此，我们需要建立一个称之为随机试验的概率模型。一个随机试验（或称随机模型）由下面三个要素组成：

1. 样本空间（sample space）是一个集合，由所有可能的试验结果构成，通常记作 Ω。样本空间可以是有限的，也可以是无限的。样本点可以是非数值类的，也可以是数值类的，此时样本点可以是标量，也可以是向量，甚至无穷序列。

2. 事件（event) 是 Ω 的子集，而所有事件构成的类 $\mathcal{F}$ 须满足

2.1 $\Omega \in \mathcal{F}$（必然事件）；

2.2 $A \in \mathcal{F} \ \Rightarrow\ \bar{A} \in \mathcal{F}$（补运算封闭），这里“$\bar{A}$”表示 A 的补集；

2.3 $A_i \in \mathcal{F} \ \Rightarrow\ \bigcup_i A_i \in \mathcal{F}$（可列并运算封闭）。

由此可推出不可能事件 $\varnothing \in \mathcal{F}$，可列交运算封闭，即事件的“加（并）、减（补）、乘（交）”运算封闭。换句话讲，事件类是根据研究问题而定义的子集类。与集合论里的术语稍有不同的是，我们用和事件与积事件分别表示集合的并与交。当两个事件对应的子集不相交时候，我们说两个事件是不相容的。

3. 概率是一个映射 $P : \mathcal{F} \to (-\infty，+\infty)$，满足

3.1 $P(\Omega) = 1$（规范性）；

3.2 $P(A) \geqslant 0$（非负性）；

3.3 $P\left(\biguplus_i A_i\right) = \sum_i P(A_i)$（可列可加性，$\biguplus_i$ 表示两两不相容的 A_i 的并）。

由 3.2、3.3 可以推出 $P(\varnothing) = 0$。此处，P 类似归一化的“计数、长度、面积、体积”等测度。

【例题 0.1】 抛一枚硬币，根据实际情况，可能选择如下模型。

模型 1

（1）样本空间：$\Omega = \{$正，反$\}$。

（2）事件类：$\mathcal{F}$ =\{$\varnothing$，\{正\}，\{反\}，\{正，反\}\}。

（3）概率：$P(\{$正$\}) = \frac{1}{2}$，$P(\{$反$\}) = \frac{1}{2}$。

模型 2

（1）样本空间、事件类同模型 1。

（2）概率：$P(\{$正$\}) = \frac{1}{4}$，$P(\{$反$\}) = \frac{3}{4}$。

模型 3

（1）样本空间：$\Omega=${正，反，立}。

（2）事件类、概率从略。

【例题 0.2】 掷一均匀骰子

（1）样本空间 $\Omega=\{1，2，3，4，5，6\}$。

（2）取 $A=\{1，3，5\}$。$\{\Omega，A，\varnothing\}$ 是事件类吗？

（3）$\{\Omega，A，\bar{A}，\varnothing\}$ 是事件类吗？

（4）$\mathcal{F}=2^{\Omega}$ 是事件类吗？

（5）对于有限样本空间，我们通常取所有子集构成的类为事件类，即任意一个子集都视作事件。若有限样本空间大小为 $|\Omega|$，问共有多少子集？

（6）若样本空间为有限样本空间，则概率的定义通常以定义单点集的概率为基本方式。若骰子是均匀的，则定义

$$P(\{1\})=P(\{2\})=\cdots=P(\{6\})=\frac{1}{6}。$$

注：我们可以用 $P(1)$ 简单记概率 $P(\{1\})$，不过要注意概率 P 是一个广义的函数，其定义域是事件类，而自变量是事件。我们也可以用 $P(A)$ 来表示事件 A 的概率。同一样本空间及事件类上可以定义不同的概率函数。此时可以用 $Q(A)$ 等表示事件 A 的概率。

【例题 0.3】 一个箱子里有 M 个可区分的球，从中取出 n 个，分析如下试验的样本空间的大小。

（1）有放回：有序，无序；

（2）不放回：有序，无序。

【例题 0.4】 把 n 个球扔进 M 个箱子，分析如下试验的样本空间的大小。

（1）球可区分，箱子可区分，不限箱内球数；

（2）球不可区分，箱子可区分，不限箱内球数；

（3）球可区分，箱子可区分，箱内球数不超过 1；

（4）球不可区分，箱子可区分，箱内球数不超过 1；

（5）球不可区分，箱子不可区分，箱内球数不限。

0.1.2 概率的基本运算律

在概率中，我们有以下运算法则：

1. **加法**：对于事件 A，B，

- $P(A\cup B)=P(A)+P(B)-P(AB)$；$P(A\cup B)\leq P(A)+P(B)$。
- 若 A 与 B 不相容，则 $P(A\cup B)=P(A)+P(B)$（分类）。
- **容斥定理**：若事件集 A_1，$\cdots$，A_n 为有限集，则有

$$P\left\{\bigcup_{i=1}^{n}A_i\right\}=\sum_{i=1}^{n}P\{A_i\}-\sum_{1\leq i<j\leq n}P\{A_i\cap A_j\}+\sum_{1\leq i<j<k\leq n}P\{A_i\cap A_j\cap A_k\}-\cdots+(-1)^{n-1}P\{A_1\cap\cdots\cap A_n\}。$$

2. **乘法**：

- 条件概率定义：设 A 与 B 为样本空间 Ω 中的两个事件，其中 $P\{A\}>0$。那么在事件 A 发生的条件下，事件 B 发生的**条件概率**为：

$$P(B|A)=\frac{P(AB)}{P(A)}；$$

- $P(AB)=P(B|A)P(A)$；若 A 与 B **相互独立**，则 $P(AB)=P(A)P(B)$；
- 链式法则：$P(A_1A_2\ldots A_n)=P(A_1)P(A_2|A_1)\cdots P(A_n|A_1A_2...A_{n-1})$；

- **多事件独立的定义**：设有 A_1，A_2，...，$A_n(n \geq 2)$ 个事件，如果其中任意 k $(2 \leq k \leq n)$ 个事件的积事件的概率都等于各事件概率之积，即

$$P(A_{i_1}A_{i_2}\cdots A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j}),$$

对于所有 $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ 均成立，则称它们相互独立。

3. **减法**：

- $P(\bar{A}) = 1 - P(A)$；
- $P(A - B) = P(A) - P(AB)$。

4. 全概率公式：若 B_i 是样本空间 Ω 的划分，则有 $P(A) = \sum\limits_i P(B_i)P(A|B_i)$（分类、分步）。

5. 贝叶斯公式：$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum\limits_j P(B_j)P(A|B_j)}$。

贝叶斯分析在工程中有广泛的应用背景。我们可以把 B_i 视作引起 A 发生的原因，$P(B_i)$ 称之为先验概率，而 $P(B_i|A)$ 是后验概率，反映事件 A 对于 B_i 概率的影响。条件概率 $P(A|B_i)$ 通常称为似然概率，是“由因及果”，而贝叶斯公式计算的后验概率是“由果及因”，是概率推断的重要工具。在实际应用中，我们通常可以借助图模型来分析复杂的概率问题，概率树可以用来描述“分类”与“分步”:“分类”用“分支”,“分步”用“分节”。从根节点（全集，或者说“必然事件”）开始做划分，随着树的生长，划分得越来越细。每个子节点是父节点的子事件，所有子节点划分（不重不漏）它们的父节点。根节点的概率是 1。每条分支有个权重，对应给定父节点条件下相应子节点发生的概率。因此，从一个节点出发的各个分支的权重之和为 1。某个节点的概率是从根节点到达该节点的路径的概率，而该概率是边概率的连乘积。一个子树的概率是该子树所有叶子对应的概率的和，这个子树的概率是等于该子树的根节点的概率的。在概率树中，一个节点的概率既可以由其父亲节点通过一次乘法算得（由上至下），也可以由其儿子节点通过加法算得（由下至上）。有些问题在题目中描述比较直接，有些需要自己提炼，整理出“概率树”，不同的“概率树”对应不同的概率模型。

【例题 0.5】 在 0 到 200 的整数中，求 $A=\{$能被 3 整除的数$\}$，$B=\{$能被 5 整除的数$\}$，和 $C=\{$能被 7 整除的数$\}$ 并集的大小。

【例题 0.6】 设 $P(A(B\cup C))=0.3$，$P(\bar{A})=0.6$，$P(\bar{A}\bar{B}\bar{C})=0.1$，求 $P(B\cup C)$。

【例题 0.7】令 A_1，A_2，...和 B_1，B_2，...分别为两个事件序列。若 $\lim_{n\to\infty} P(A_n)=1$ 且 $\lim_{n\to\infty} P(B_n)=1$，证明：$\lim_{n\to\infty} P(A_n\cap B_n)=1$。

【例题 0.8】 给定 B，$P(B)>0$，则 $P(*|B)$ 也是概率。

（1）特别地，$P(A|B)+P(\bar{A}|B)=1$。

（2）$P(A|B)+P(A|\bar{B})$ 等于 1 吗？讨论 $P(A|B)$ 与 $P(A)$ 的大小关系。

【例题 0.9】 掷两个骰子，记：事件 $A=\{$第一个骰子点数为奇数$\}$；事件 $B=\{$第二个骰子点数为奇数$\}$；事件 $C=\{$两个骰子点数和为奇数$\}$；则 A，B，C 两两独立，但不互相独立。

【例题 0.10】 考虑一个概率空间。记 A，B，C 是三个事件。

（1）举例说明事件 A 与（B 与 C 的和事件）独立，但 A 与 B 不独立，与 C 也不独立。

（2）设 A，B 两个事件的概率均大于 3/4。证明 A 与 B 的积事件的概率大于 1/2。

【例题 0.11】 现有三个袋子，编号为 $i=1,2,3$，已知第 i 个袋子有 b_i 个黑球和 w_i 个白球，现在随机的选择一个袋子，然后在这个袋子中随机的选择一个球，选中黑球的概率是多少？若已知选取的是黑球，如何推断黑球是从哪一个袋里取出来的？

【例题 0.12】 Alice 向 Bob 发送字符 0 或 1，发送 0 的概率为 p_0，传输过程中字符有概率 p_e 被翻转，求当 Bob 收到的字符为 1 时，Alice 发送字符为 0 的概率。

【例题 0.13】 老师在三个信封中分别装入三道数学测试题，一道是线性代数的，而另两道是随机过程的。为方便起见，学生甲随机选定一个信封，不妨记作 A。然后，助教打开剩下两个信封中的一个，记作 B。拆开后发现其中是随机过程的试题。此时助教问同学甲：要不要选做另一个信封（不妨记作 C）的题目，请分析同学甲有偏好的选择及依据。

【例题 0.14】 老师准备从五个学生 A、B、C、D、E 中选三个代表班级参加随机过程课程竞赛。

（1）A 被选中的概率是多少？

（2）老师公布了其中一个参赛同学：E。此时 A 被选中的概率是多少？

（3）在老师未公布之前，A 向助教打听自己有无入选。助教按照纪律要求不能告知 A 是否入选，但告知 E 入选了。此时，A 入选的概率是多少？

§0.2

随机变量及其分布函数

0.2.1 基本概念

【定义 0.15】 随机变量 X 不是（自）变量，而是样本空间 Ω 到实数集 $\mathbb{R}$ 的映射，即 $X:\Omega\to\mathbb{R}$，满足：对于任意实数，$\{\omega\in\Omega|X(\omega)\le x\}$ 是一个事件。为简便起见，我们把该事件记为 $\{X\le x\}$。

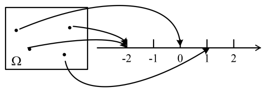


图 1: 随机变量

例如，抛一枚硬币，概率模型如例 0.1.1 模型 1。我们可以定义随机变量 $X(\text{正})=1$，$X(\text{反})=0$。引入随机变量后，样本点就有了数值映像。这有助于我们利用数学分析的工具揭示一些随机现象的统计规律。但需要注意的是，不同的样本点可以映射到相同的数值。

【例题 0.16】 掷一枚均匀的骰子，定义 X 为骰子面上的点数。此时，样本空间 $\Omega=\{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$。回答如下问题。

（1）取 $A=\{1,\ 3,\ 5\}$。$\mathcal{F}_0=\{\Omega,\ A,\ \varnothing\}$ 是事件类吗？$\mathcal{F}_1=\{\Omega,\ A,\ \bar{A},\ \varnothing\}$ 是事件类吗？$\mathcal{F}=2^{\Omega}$ 是事件类吗？

（2）若我们考察的概率模型以 $\mathcal{F}_1$ 为事件类，X 是随机变量吗？若我们考察的概率模型以 $\mathcal{F}$ 为事件类，X 是随机变量吗？

（3）定义 $Y=X \bmod 2$，讨论上一个问题。

注：以 $\mathcal{F}_1$ 为事件类，X 不是随机变量。因为存在 x_0 使得 $\{\omega\in\Omega|X(\omega)\le x_0\}\notin\mathcal{F}_1$。如 $x_0=2$，$\{X\le 2\}=\{1,\ 2\}\notin\mathcal{F}_1$。以 $\mathcal{F}$ 为事件类，X 是随机变量。而不论以 $\mathcal{F}_1$ 还是 $\mathcal{F}$ 为事件类，Y 都是随机变量。另外也要注意，上面的讨论与骰子是否均匀关系不大。换句话讲，一个从样本空间 Ω 到实数轴 R 的映射 X 是否是随机变量，仅与事件类 $\mathcal{F}$ 的定义有关。

设 X 是一个随机变量。由于 $\{X\le x\}$ 是事件，我们可以定义 $F_X(x)=P(X\le x), x\in\mathbb{R}$, 称之为随机变量 X 的累积分布函数（cumulative distribution function，CDF）。当 x 在 $(-\infty,\ \infty)$ 变化时，累积分布函数反映了随机变量取值在实数轴上的分布情况。如果 X 取值有限或者可列，则称为离散型随机变量，我们可以定义概率质量函数 $P_X(x)=P(X=x)$，$x\in\mathcal{X}$（$\mathbb{R}$ 的有限子集或可列子集）满足 $P_X(x_i)\geqslant 0$ 且 $\sum_i P_X(x_i)=1$。如果 $F_X(x)$ 连续且几乎处处可微，我们称 X 为连续型随机变量，并定义概率密度函数 $f_X(x)$，$x\in\mathcal{X}$（$\mathbb{R}$ 上的区间 $\mathcal{X}$）满足 $f_X(x)\geqslant 0$ 和 $\int_{\mathbb{R}} f_X(x)\mathrm{d}x=1$。图2、图3与图4分别给出了离散型随机变量的 PMF、连续型随机变量的 PDF 与一般随机变量的 CDF 的示意图。

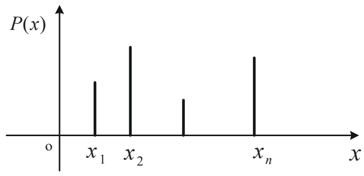


图 2: 离散型随机变量的 PMF (又称分布律)，“谱线”

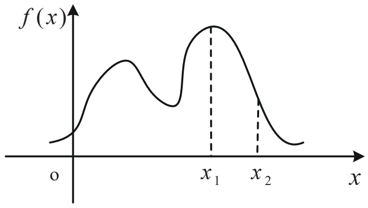


图 3: 连续型随机变量的 PDF，“谱密度”

一个随机变量 X 的累积分布函数 $F_X(x) \triangleq P(X \leqslant x)$，$x \in \mathbb{R}$ 满足如下性质：

1. 对离散型随机变量 X，$F_X(x) = \sum\limits_{x_i \leqslant x} P_X(x_i)$ 是阶梯函数，其中 $P_X(x)$，$x \in \mathcal{X}$ 是 X 的概率质量函数。

2. 对连续型随机变量 X，$F_X(x) = \int_{y \leqslant x} f_X(y)\mathrm{d}y$ 是连续函数，其中 $f_X(x)$，$x \in \mathcal{X}$ 是 X 的是概率密度函数。

3. $F_X(x)$ 是单调递增（未必严格单调）函数，且

$$\lim_{x \to -\infty} F_X(x) = 0, \qquad \lim_{x \to +\infty} F_X(x) = 1。$$

4. $F_X(x)$ 的间断点至多可列个；若在 x_1 处有跳跃，则跳跃的高度恰好是概率 $P_X(x_1)$。

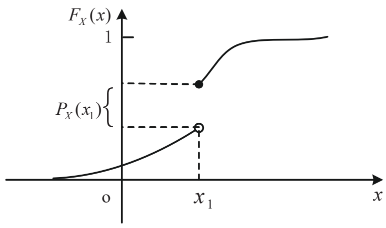


图 4: 随机变量的 CDF

以上讨论的只限于一个随机变量的情况，如果随机试验的结果需要同时考虑两个或两个以上的随机变量的情况，则可以定义复随机变量，随机向量，随机序列，随机过程等。

0.2.2 常见的随机变量

常见的离散随机变量的概率质量函数分述如下，其中 $0 < p < 1$，$\lambda > 0$ 是参数。

1. 两点分布：随机变量 X 取 0 与 1 两个值，其分布律是 $P_X(1) = p$，$P_X(0) = 1 - p$。

2. 二项分布：设 n 是正整数，记 $B(n;p)$ 为二项分布，其分布律是 $P_X(m) = \binom{n}{m} p^m (1-p)^{n-m}$，$0 \leq m \leq n$。

3. 几何分布：随机变量 X 取正整数，其分布律是 $P_X(m) = p(1-p)^{m-1}$，$m \geq 1$。

4. 泊松分布：随机变量 X 取非负整数，其分布律是 $P_X(n) = \frac{\lambda^n \exp(-\lambda)}{n!}$，$n \geq 0$。

常见的连续随机变量的概率密度函数分述如下，其中 $\lambda > 0$、$\sigma > 0$、μ 是参数。

1. 指数分布：随机变量 X 的密度函数 $f_X(x) = \lambda \exp(-\lambda x)$，$x \geq 0$。

2. 高斯分布：随机变量 X 的密度函数是 $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-(x-\mu)^2}{2\sigma^2})$，$x \in \mathbb{R}$。

3. 均匀分布：给定 $a > 0$。随机变量 X 的密度函数是 $f_X(x) = \frac{1}{a}$，$0 \leq x \leq a$。

注：密度函数的定义域与累积分布函数的定义域一致，均是整个实数轴。在给出密度函数的时候，通常只给出函数的支撑集（support set), 即函数值不为 0 的自变量所构成的子集合。在支撑集之外，函数值默认为 0。

我们称 $F_{X^n}(x^n) = P(X_1 \leq x_1，X_2 \leq x_2，\cdots，X_n \leq x_n)$ 为 n 维随机向量的联合分布函数（joint distribution function）。它表示事件 $X_1 \leq x_1$，$X_2 \leq x_2$，$\cdots$，$X_n \leq x_n$ 同时出现的概率。联合分布函数亦称多维分布函数。一般地，多维随机变量（包括随机序列、随机过程）可借助联合分布 $F_{X^n}(\underline{x}^n) = P(X_1 \leqslant x_1，X_2 \leqslant$

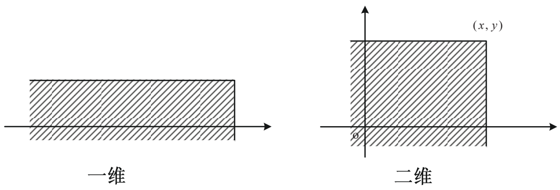


图 5: 分布函数

x_2，$\cdots$，$X_n \leqslant x_n$) 来刻画，其中 x^n 表示（x_1，x_2，$\cdots$，x_n)。图5给出了分布函数的示意图，即试验结果对应的随机变量取值落在阴影部分的概率，而这个阴影部分随着 $(x，y)$ 的变化可以“扫描”整个平面。

对于二维离散随机变量 $(X，Y)$，我们有以下分布律：

- 联合分布律：$P_{X,\ Y}(x，y)$，$x \in \mathcal{X}$，$y \in \mathcal{Y}$。
- 条件分布律：$P_{X|Y}(x|y) = \frac{P_{X,\ Y}(x，y)}{P_Y(y)}$，$P_{Y|X}(y|x) = \frac{P_{X,\ Y}(x，y)}{P_X(x)}$。
- 边缘分布律：$P_X(x) = \sum\limits_y P_{X,\ Y}(x，y)$，$P_Y(y) = \sum\limits_x P_{X,\ Y}(x，y)$。

【定义 0.17】 设 X 是随机变量，而 $\mathbf{g}(x)$ 是一个普通函数，则 $Y = \mathbf{g}(X)$ 也是一个随机变量，即

$$Y : \Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}$$

【例题 0.18】投掷两个均匀的骰子。随机变量 X_i，$i \in \{1，2\}$ 表示第 i 个骰子的点数，$X_i \in \{1，2，3，4，5，6\}$。令 $X = X_1 + X_2$，求：

（1）$P(X = 2|X_1 = 6)$；

（2）$P(X_1 = 2|X_2 = 2)$；

（3）$P(X_1 = 1|X = 2)$。

【例题 0.19】 甲乙两人网聊。甲向乙随机发送一个数字 $X \in \{0，1，2\}$，概率依次为 $\frac{1}{2}$，$\frac{1}{4}$，$\frac{1}{4}$。乙收到数字 $Y \in \{0，1，2\}$。但传输过程中有可能出错。设错误模型是 $Y = X + Z \bmod 3$，其中 $Z \in \{0，1，2\}$ 的分布律是 $\frac{7}{8}$，$\frac{1}{16}$，$\frac{1}{16}$。讨论乙如何根据接收的 Y 推断发送的 X，并给出具体的准则。

【例题 0.20】 设 X 是随机变量，则 $Y = X^2$，$Z = P_X(X)$，$L = -\log P_X(X)$ 均是随机变量。他们的分布律如下所示。注：随机变量“被函数”之后，“谱线”的数目可能减少，高度可能增加。

X	-1	0	1	2
$P_X(x)$	0.2	0.3	0.2	0.3

$Y = X^2$	0	1	4
$P_Y(y)$	0.3	0.4	0.3

$Z = P_X(X)$	0.2	0.3
$P_Z(z)$	0.4	0.6

$L = -\log P_X(X)$	2.32	1.74
$P_L(l)$	0.4	0.6

【例题 0.21】 已知 X 的密度函数，Y 的密度函数。若 X，Y 独立，求 $Z = X + Y$ 的密度函数。

【例题 0.22】 随机变量 X，分布函数为 $F_X(x)$。求以下随机变量的分布函数。

（1）分布函数为 $F_X(x)$ 的 n 个独立同分布随机变量中的最大值。

（2）分布函数为 $F_X(x)$ 的 n 个独立同分布随机变量中的最小值。

（3）以上 (1) 和 (2) 中定义的随机变量之差。假设 X 的密度函数为 $f_X(x)$。

§0.3 数字特征

【定义 0.23】 设离散型随机变量 X 的概率质量函数为 $P_X(x_k) \triangleq p_k$，$k = 1，2，\cdots$。若级数 $\sum\limits_{k=1}^{\infty} x_k p_k$ 绝对收敛，则 X 的数学期望定义为：

$$\mathbf{E}(X) = \sum_{k=1}^{\infty} x_k p_k。$$

设连续型随机变量的概率密度函数为 $f(x)$，若积分 $\int_{-\infty}^{\infty} xf(x)\mathrm{d}x$ 绝对收敛，则 X 的数学期望定义为：

$$\mathbf{E}(X)=\int_{-\infty}^{\infty} xf(x)\mathrm{d}x。$$

而一般随机变量 X 的数学期望若存在的话，可以表示为以下的 Stieltjes（斯蒂尔杰斯）积分：

$$\mathbf{E}(X)=\int_{-\infty}^{\infty} x\mathrm{d}F_X(x)。$$

其中，$F_X(x)$ 为 X 的累积分布函数。

随机变量的期望 $\overline{X}=\mathbf{E}(X)$，有时也记作 $\mathbf{E}[X]$，通常被认为是一个随机变量的一个“典型值”。我们有如下命题。

【命题 0.24】 对于任意非负的随机变量 X，$\mathbf{E}[X]=\int F_X^c(y)dy$，其中，$F_X^c(x)=1-F_X(x)=P(X>x)$。

证明. 证明从略。（见拓展阅读材料） □

【定义 0.25】 设 X 是一个随机变量，如果 $D(X)\triangleq\mathbf{E}\left([X-\mathbf{E}(X)]^2\right)$ 存在，则称其为 X 的**方差**。

【命题 0.26】 设 X，Y 是随机变量，$z=g(x)$ 是一个普通函数。则当所涉及的数学期望存在时，我们有

1. $\mathbf{E}(X+Y)=\mathbf{E}(X)+\mathbf{E}(Y)$（不管 X 与 Y 是否独立）;
2. 若 X 与 Y 独立，则 $\mathbf{E}(XY)=\mathbf{E}(X)\mathbf{E}(Y)$;
3. 若 X 与 Y 独立，则 $D(X+Y)=D(X)+D(Y)$;
4. $\mathbf{E}(Y)=\mathbf{E}(\mathbf{E}(Y|X))$;
5. $\mathbf{E}(Z)=\mathbf{E}(g(X))$。

证明. 证明从略。 □

注：给定 $X=x$，$\mathbf{E}(Y|X=x)$ 是一个数，而 $\mathbf{E}(Y|X)$ 可以看作是随机变量 X 的函数。随机变量的函数的期望公式不只是简单地等量替换，要注意区分期望所对应的变量。

【定义 0.27】 一个随机变量的矩生成函数（MGF，moment generating function）定义为

$$g_X(r)=\mathbf{E}(e^{rX})=\int_{-\infty}^{\infty} e^{rx}\mathrm{d}F_X(x),$$

其中 r 是实变量。

注：矩生成函数 MGF 的定义形式上是以无穷积分给出的，其在 $r=0$ 时收敛的，且 $g_X(0)=1$。一般情况下，$g_X(0)$ 的收敛区域是实数轴的一个区间。

【定理 0.28】 设 $g_X(r)$，$r\in I\subseteq\mathbb{R}$ 是随机变量的矩生成函数，则

（1）$\mathbf{E}(X^k)=\frac{\mathrm{d}^k g_X(r)}{\mathrm{d}r^k}|_{r=0}$，对于 $k\geq 0$。

（2）设 X_1，X_2，…，X_n 是 n 个独立随机变量，$S_n=\sum_{i=1}^n X_i$，则

$$g_{S_n}(r)=\prod_{i=1}^n g_{X_i}(r)$$

证明. 证明从略。 □

【例题 0.29】 概率模型中的事件 A 可以定义一个二元随机变量 $\mathbb{I}_A$，称之为 A 的示性变量，满足 $\mathbb{I}_A(\omega)=1$，$\omega\in A$，而 $\mathbb{I}_A(\omega)=0$，$\omega\notin A$。反之，一个二元随机变量必定是某个事件 A 的示性变量。我们有

$$P_{\mathbb{I}_A}(0)=1-P(A);\quad P_{\mathbb{I}_A}(1)=P(A)。$$

$$\mathbf{E}[\mathbb{I}_A]=P(A);\quad \sigma_{\mathbb{I}_A}=\sqrt{P(A)(1-P(A))}。$$

注：概率是期望，期望是积分；积分是期望，期望是概率。

【例题 0.30】 期末某校评优，甲、乙、丙、丁、戊 5 名同学分别获得“德、智、体、美、劳”单项奖状。不巧的是，辅导员发放奖状时完全随机放乱了。请问，5 名同学中平均意义上有几人拿到了与自己匹配的奖状？

§0.4 尾部概率

在实际应用中，一方面，事件的概率未必有简易的计算方法；另一方面，很多情况下非平凡的概率界也可以用来指导系统的设计。下面介绍一些重要的不等式。

1. Markov（马尔科夫）不等式（若均值有限，则尾部概率线性递减趋于 0）

设非负随机变量 X 的期望 $\mathbf{E}(X)$ 的值有限，则对于 $y>0$， $P(X\geq y)\leq\frac{\mathbf{E}(X)}{y}$。

证明：

$$\mathbf{E}(X)=\int_0^{+\infty}x\mathrm{d}F_X\geqslant\int_y^{+\infty}x\mathrm{d}F_X\geqslant y\int_y^{+\infty}\mathrm{d}F_X=yP(X\geqslant y)。$$

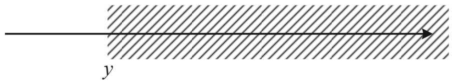


图 6: Markov（马尔科夫）不等式

2. Chebyshev（切比雪夫）不等式（若方差有限，则尾部概率平方递减趋于 0）

对于任意 $\delta>0$，$P(|X-\mathbf{E}(X)|\geqslant\delta)\leqslant\frac{\sigma_X^2}{\delta^2}$，其中 σ_X^2 表示 X 的方差。

证明：对 $(X-\mathbf{E}(X))^2$ 应用 Markov 不等式，

$$\begin{aligned}&P(|X-\mathbf{E}(X)|\geqslant\delta)\\=\;&P(|X-\mathbf{E}(X)|^2\geqslant\delta^2)\\\leqslant\;&\frac{\mathbf{E}(X-\mathbf{E}(X))^2}{\delta^2}=\frac{\sigma_X^2}{\delta^2}。\end{aligned}$$

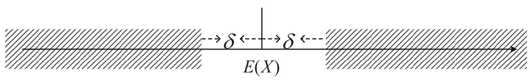


图 7: Chebyshev（切比雪夫）不等式

3. Chernoff（切诺夫）界（若矩生成函数有限，则尾部概率指数递减趋于 0）

$$P(X\geqslant y)\leqslant\mathbf{E}(e^{sX})e^{-sy},\quad 对所有s\geqslant 0。$$

证明：对 e^{sX} 应用 Markov 不等式，

$$P(X\geqslant y)=P\left(e^{sX}\geqslant e^{sy}\right)\leqslant\frac{\mathbf{E}(e^{sX})}{e^{sy}},\quad s\geqslant 0。$$

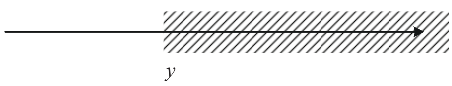


图 8: Chernoff（切诺夫）界

类似地，$P(X\leqslant y)\leqslant\mathbf{E}(e^{sX})e^{-sy}$，　对所有 $s\leqslant 0$。

上述几个不等式也可以借助积分的性质（非负函数的积分非负）来证明，如图9所示。为此，对于 $y>0$，定义示性函数（indicator function）

$$Z(\omega)=\begin{cases}1, & X(\omega)\geqslant y\\0, & 其他\end{cases}$$

又 $\Pr\{X \geqslant y\} = \Pr\{Z = 1\} = \mathbf{E}(Z)$，可得

$$Z \leqslant \frac{X}{y} \Rightarrow \Pr\{X \geqslant y\} = \mathbf{E}(Z) \leqslant \frac{\mathbf{E}(X)}{y}$$
$$Z \leqslant \frac{X^2}{y^2} \Rightarrow \Pr\{|X| \geqslant y\} = \mathbf{E}(Z) \leqslant \frac{\mathbf{E}(X^2)}{y^2}$$
$$Z \leqslant \frac{e^{sX}}{e^{sy}} \Rightarrow \Pr\{X \geqslant y\} = \mathbf{E}(Z) \leqslant \frac{\mathbf{E}(e^{sX})}{e^{sy}}.$$

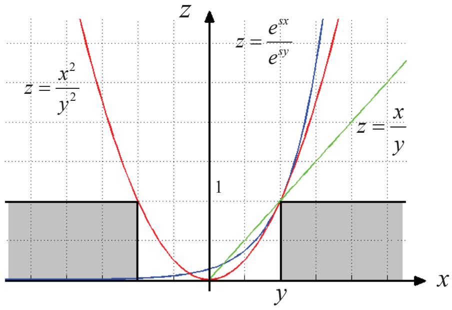


图 9: 基于积分性质的尾部概率不等式的证明示意图

§0.5
几种典型概率问题分析

0.5.1 采样模型：设有罐子盛有 b 个黑球，w 个白球，均匀随机取一个球，记录球的颜色

1. 有放回采样：放回取出的球；
2. 无放回采样：不放回取出的球；
3. Polya 采样：放回取出的球，并同时添加 r 个同颜色的球。

可以看出，若 $r = 0$，则是有放回采样，若 $r = -1$，则是无放回采样。

研究的主要问题是，连续不断做上述试验，采样所得的球的颜色分布。例如，

【例题 0.31】 记 B_k 是第 k 次取到黑球的事件，W_k 是第 k 次取到白球的事件。求：

（1）$P(B_2|W_1)$；

（2）$P(B_1B_2)$；

（3）定义一个随机变量 X_t 如下。若第 t 次取到的是黑球，令 $X_t = 1$，否则 $X_t = 0$。说明 $\boldsymbol{X}^m = (X_1, X_2, \cdots, X_m)$ 与 $\tilde{\boldsymbol{X}}^m = (\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_m)$ 是同分布的，这里 $\tilde{\boldsymbol{X}}^m$ 是 $\boldsymbol{X}^m$ 的置换版本。

0.5.2 球箱模型：独立均匀随机地投 m 个球到 n 个箱子，需研究以下问题

【例题 0.32】 **生日悖论/碰撞问题：** 把 n 个球放进 M 个箱子，有一个箱子内球数超过 1 的概率。

【例题 0.33】 **集优惠券/空箱问题：** 有 n 类不同的优惠券，买 1 件商品可以随机获得其中某类 1 张，问：平均需要购买多少商品才能集齐？

【例题 0.34】 **最大负载问题：** 若 $m = n$，则最“满”的箱子里有多少球？

0.5.3 独立掷硬币多次，正面朝上的概率为 p

【例题 0.35】 用语言描述两点分布、二项分布、几何分布。给出它们的概率质量函数、数学期望与矩生成函数。

【例题 0.36】 观察到 n 次正面朝上时，共掷了多少次？

0.5.4 概率树，概率转移图，贝叶斯推断

【例题 0.37】 一个家庭有两个孩子，设男孩、女孩的机会均等。问：

（1）已知老大是女孩，老二也是女孩的概率；

（2）已知有一个是女孩，有两个女孩的概率。

【例题 0.38】 有三个文件柜，一个文件有可能放在某个文件柜里。设快速翻阅一个会有该文件的文件柜，发现该文件的概率为 α。现翻阅文件柜 A 后，未发现。问文件在 A 中的概率。

【例题 0.39】 单项选择题设几个选项合适？

【例题 0.40】 试着解释临床检查中的“假阳性与假阴性”。一个阳性者真有病的概率受哪些因素影响？

0.5.5 数学期望

【例题 0.41】 **非负随机变量的数学期望的几何意义：**

（1）$X \geq 0$，$\mathbf{E}X$；

（2）$Y = X^2$，$\mathbf{E}Y$；

（3）$X \geq 0$，$Y = \min(1,\ X)$，说明 $\mathbf{E}Y = P(X \geq U)$，其中 U 是独立于 X, 取值于 $[0,\ 1]$ 区间的均匀随机变量；

（4）非负整数变量 X，$\mathbf{E}X = \sum\limits_{i=1}^{\infty} P\left(X \geq i\right)$。

【例题 0.42】 **条件期望：**

（1）$\mathbf{E}(X|Y)$ 的意义；

（2）$\mathbf{E}X = \mathbf{E}(\mathbf{E}(X|Y))$；

（3）掷一个均匀骰子，点数记为 N，再掷 N 次，点数记为 X_1，X_2，....，X_N，求 $\mathbf{E}(\sum X_i)$。

【例题 0.43】 考虑一个正整数随机变量 Y，其累积分布函数 $F_Y\left(y\right) = 1 - \frac{2}{(y+1)(y+2)}$，求：

（1）Y 的质量分布函数；

（2）$\mathbf{E}Y$；

（3）设 X 是正整数随机变量，且 $P_{X|Y}\left(x|y\right) = \frac{1}{y}$，$1 \leq x \leq y$，求 (a) $\mathbf{E}(X|Y = y)$；(b) X 的概率质量函数 $P_X(x)$，$\mathbf{E}X$。

0.5.6 对称性的应用

【例题 0.44】 箱子里最初有 1 个白球，1 个黑球。每次取出一个，然后放回，同时添 1 个同色球。n 次试验后，猜测箱内白球数目的分布并证明。

【例题 0.45】 掷均匀硬币 10 次，求下面事件的概率：

（1）正面朝上的次数 = 反面朝上的次数；

（2）正面朝上的次数多于反面朝上的次数；

（3）至少有连续 4 次正面朝上。

【例题 0.46】 有 r 个参赛人，其中参赛人 $i(i = 1,\ \ldots,\ r)$ 在开始时有 $n_i(n_i > 0)$ 个单位（财富）。在每一阶段参赛人中的两个被选中比赛，赢者从输者那里得到一个单位。任何参赛人，当他的财富减少到 0 时就退出，如此继续，直至某个参赛人占有所有单位的 $n = \sum_{i=1}^{r} n_i$ 个单位为止，此参赛人就是胜利者。假定相继比赛的结果是独立的，而且在每次比赛中两个参赛人等可能地获胜，求参赛人 i 是胜利者的概率。

0.5.7 $Monte\ Carlo$ 仿真

【例题 0.47】 讨论积分、事件概率与数学期望的关系，并给出仿真计算方法。

【例题 0.48】 由均匀随机变量 $(0,\ 1)$ 产生随机变量 $X \sim F_X\left(x\right)$ 的一般方法。

【例题 0.49】 用均匀硬币序列产生一个二元变量 X，使得 $P_X(1) = \frac{1}{\sqrt{e}}$，$P_X(0) = 1 - \frac{1}{\sqrt{e}}$。

【例题 0.50】 产生 n 维球内均匀分布的点。

0.5.8 概率应用

【例题 0.51】 设两个多项式 $f(x)$，$g(x)$，编写一段程序计算乘积 $h(x)=f(x)g(x)$。给出一种测试所写程序正确与否的方法。

【例题 0.52】 设 m，n 是两个大整数。一个函数 $F:\{0，1，\ldots，n-1\}\to\{0，1，\ldots，m-1\}$ 满足 $F((x+y)\bmod n)=(F(x)+F(y)) \bmod m$。该函数可以查表计算，即 $F(x)$，$0\le x\le n-1$ 是已知的，但被随机修改了，设错的位置占比 1/5。设计一种算法，可靠计算 $F(x)$，$0\le x\le n-1$。

§0.6 课后作业

1. 从 [1，1 000 000] 范围中随机抽取一个数。请运用容斥定理计算这个数能被 4，6 和 9 中一个或多个整除的概率。

2. 有一枚均匀的硬币和一枚两面都是头像（正面）的硬币，以相同概率从这两枚硬币中随机选择一枚并投掷。已知投掷结果是出现正面，那么投掷的是两面均是头像的硬币的概率是多少。

3. 连续地抛掷一枚均匀的硬币。

（1）求抛掷的前四次是下列情况的概率：

H，H，H，H。

T，H，H，H。

（2）求模式 T，H，H，H 出现在模式 H，H，H，H 之前的概率。

4. 甲乙两人比赛，规定只要中间有一人赢得了 n 局，比赛立即结束。假定比赛在两人间公平进行，即每人赢得一局比赛的概率都是 1/2，与其他不同局的结果无关。那么比赛结束时，失败一方已经赢得 k 局的概率是多少？

5. 投掷 10 枚标准的六面体骰子，假定投掷每枚骰子是独立的。它们的点数之和能被 6 整除的概率是多少？

6. 对 n 个人进行核酸检测。每个人可以单独检测，但费用过高。合并检测可以减少费用。把 k 个人的样本合起来同时分析，如果检测结果呈阴性，对这 k 个人的组，这一次检测就好了。如果检测结果呈阳性，则这 k 个人需要再进行单独检测，因此这 k 个人需要进行 $k+1$ 次检测。假定我们产生了 n/k 个不同的组，每组 k 个人（k 能整除 n），并用合并法进行检测。假设对于独立检测，每个人呈阳性的概率为 p。

（1）对 k 个人的合并样本，检测呈阳性的概率是多少？

（2）需要检测的期望次数是多少？

（3）描述如何求最优的 k 值。

（4）给出一个不等式，说明对什么样的 p 值，合并检测比每个人单独检测更好。

7. 设 A、B 是两个事件。证明如下示性变量之间的关系并回答有关问题。

（1）$\mathbb{I}_\Omega=1$，$\mathbb{I}_\phi=0$；

（2）$\mathbb{I}_{\bar{A}}=1-\mathbb{I}_A$；

（3）$\mathbb{I}_{A\cup B}=\max(\mathbb{I}_A,\mathbb{I}_B)$，$\mathbb{I}_{AB}=\min(\mathbb{I}_A,\mathbb{I}_B)$；

（4）$\mathbb{I}_A+\mathbb{I}_B \bmod 2$ 对应的事件是什么？$\mathbb{I}_A\mathbb{I}_B$ 呢？

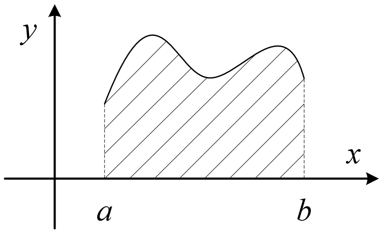
8. 在《万里归途》电影里有一个情景，穆夫塔刁难宗大伟，发起“轮盘赌”。现设枪中子弹数服从概率为 $\frac{1}{2}$ 的 0-1 两点分布。请细化概率模型，分析随着“轮盘赌”进行，枪中无子弹的概率是如何变化的？具体地，无子弹的初始概率为 $\frac{1}{2}$。选择至少两个概率模型（必要时可以改变两轮之间的规则），分析第 $i\ge 1$ 枪后无子弹的概率。

§0.7 拓展阅读

一、积分、期望、蒙特卡洛仿真

我们讨论非负函数或者非负随机变量。在此前提下，分析一些概念的几何意义，有助于直观理解。

一个函数 $y=f(x)$ 在一个给定区间 $[a,b]$ 上的定积分，即是这个曲线下方的面积。



从定义上讲，可以分成四步来描述这个积分过程。

第一步，把区间 $[a,b]$ 分成很多小区间，每个区间的长度记为 Δx；

第二步，在每个区间内取一个点，计算每一个区间对应“梯形”的近似面积 $f(x)\Delta x$；

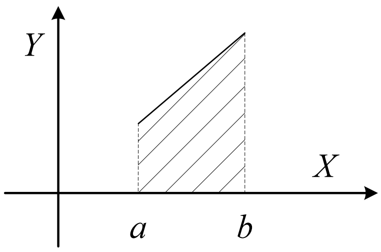
第三步，对这些面积求和；

第四步，当区间越分越细时，上述和的极限是积分 $\int_a^b f(x)\mathrm{d}x$。

一个非负随机变量 X 的数学期望 $\mathbf{E}(X)$ 是概率加权和。对于离散型随机变量而言，$\mathbf{E}(X)=\sum xP(x)$；对于连续型随机变量而言，$\mathbf{E}(X)=\int xf(x)\mathrm{d}x$。这里的 $P(x)$ 是概率质量函数，$f(x)$ 是概率密度函数。对于一般的随机变量 X，其累积分布函数记作 $F_X(x)$，则数学期望定义为

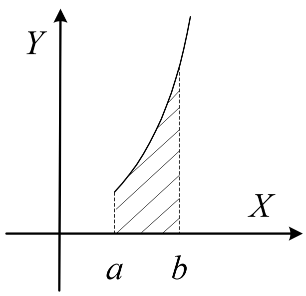
$$\mathbf{E}(X)=\int x\mathrm{d}F_X(x).$$

这个形式是黎曼-斯蒂尔杰斯（Riemann-Stieltjes）积分，与常见的黎曼积分相比，其定义也可以描述为四个步骤，但梯形的面积近似用 $f(x)\cdot\Delta F_X(x)$ 来取代，也就是说，横坐标的“长度”不是线性的，而是由 $F_X(x)$ 来定义。这样，我们可以认为 $\mathbf{E}(X)$ 是如下阴影部分的“面积”，其中 a,b 分别是 X 的最小值与最大值。



再次强调，这部分面积的计算是按照 $F_X(x)$ 的尺度计算的。

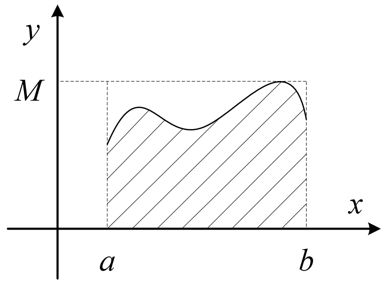
设 $Y=X^2$，则 $\mathbf{E}(Y)$ 对应如下阴影部分的面积。



一个事件 A 的概率记作 $P(A)$，其可以表示成一个数学期望 $\mathbf{E}(I_A)$，其中 I_A 表示 A 的示性函数。当一个样本点 $\omega \in A$ 时，$I_A(\omega) = 1$；否则 $I_A(\omega) = 0$。由于 $I_A(\omega)$ 是定义在样本空间 Ω 上的函数，所以，当样本空间是多维时，$I_A(\omega)$ 的图示就不方便了，这也表明 $I_A(\omega)$ 更具一般性。

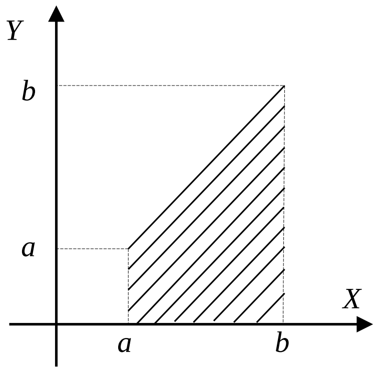
蒙特卡洛仿真可以很方便地估算 $P(A)$ 的值。基本方法是，产生大量的样本点 ω，统计 $\omega \in A$ 发生的频率即可。蒙特卡洛仿真可以估算 $P(A)$，而 $P(A)$ 是示性函数的积分，因而，蒙特卡洛仿真也可以用来计算积分。

第一种情况，常规积分 $\int_a^b f(x)\mathrm{d}x$ 的仿真方法。



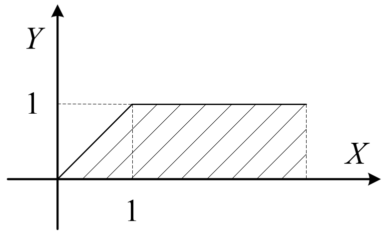
产生 $[a,b] \times [0,M]$ 区域中的均匀分布，统计落在阴影部分的频率，用该频率乘以矩形部分面积 $M(a-b)$ 来估计积分值。

第二种情况，$\mathbf{E}(X) = \int_a^b x\mathrm{d}F_X$。



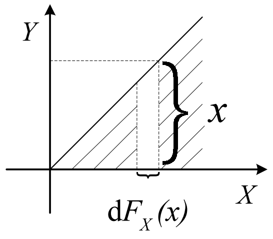
产生 X 的样本，同时产生 $[a,b]$ 之间的均匀分布 Y，统计 (X,Y) 落在阴影部分的频率。此时，矩形“面积”为 b，因为矩形在 X 方向上的“长度”是用 F_X 来计算的，长度为 1。再用统计出的频率乘以矩形部分面积 b 来估计积分值。

再看一个例子。设 X 是非负随机变量，说明 $\mathbf{E}(\min(1,X)) = \Pr\{X \geq U\}$，其中 U 与 X 相互独立，且服从 $[0,1]$ 区间上的均匀分布。我们可以看出，$Y = \min(1,X)$ 的图形如下，

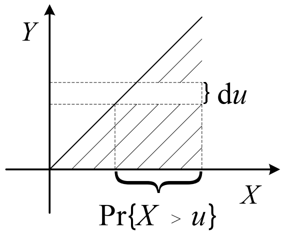


因此，$\mathbf{E}(Y)$ 即是上图阴影部分的面积，这个面积的计算可以归为产生 (X,U)，统计频率即可，其意义就是 $\Pr\{X \geq U\}$。

最后，我们再来解释一下非负随机变量的一个重要公式 $\mathbf{E}(X) = \int_0^{+\infty} F_X^c(x)\mathrm{d}x$，即 X 的均值等于其累积分布补函数的积分。我们已经知道，$\mathbf{E}(X)$ 是如下阴影部分的概率。



方法一



方法二

阴影部分的概率有两种计算方法。

一是“先纵后横”，给定 x，其对应横坐标“长度”是 $\mathrm{d}F_X(x)$，高度是 x。因而纵向小矩形面积是 $x\mathrm{d}F_X(x)$，横向累积即期望 $\mathbf{E}(X)=\int x\mathrm{d}F_X$。

二是“先横后纵”，给定 u，这个 u 对应的高度是 $\mathrm{d}u$，宽度是 $\Pr\{X>u\}$，所以 $\mathbf{E}(X)=\int \Pr\{X>u\}\mathrm{d}u$。

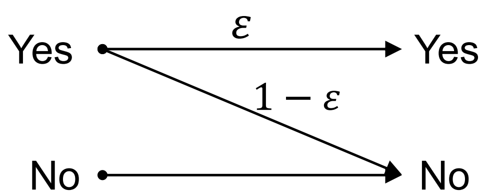
二、随机判决算法

一般地，随机判决算法分为两类：蒙特卡洛算法和拉斯维加斯（Las Vegas）算法。

- 蒙特卡洛算法：有限时间内停下，但输出有可能是错的；
- Las Vegas 算法：输出总是正确的，但运行时间可能无界，不过平均所需时间是有界的。

蒙特卡洛算法又分为单边错误与双边错误两种情况：

1. 单边错误的蒙特卡洛算法

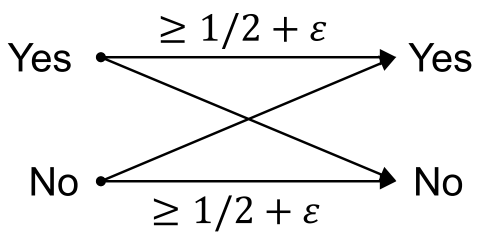


若输出“Yes”，则一定是“Yes”，若输出“No”，则不确定。当输出是“No”的时候，可以再启动算法（采用不同的随机种子）。若 t 次算法之后，得到的总是“No”，则断言“No”，此时错误的概率 $\leq (1-\epsilon)^t$。若要使这个概率 $\leq \delta$，得到

$$(1-\epsilon)^t \leq \delta,$$

则 $t=\mathcal{O}\left(\log(1/\delta)\right)$。若上述算法是多项式时间完成，称为“RP”（Randomized polynomial time，随机多项式时间）。

2. 双边错误的多项式时间蒙特卡洛算法，称之为 BBP（Bounded-error probability polynomial time）。



可以运行算法多次，按照大数逻辑选择结果输出。

$$\sum_{k=0}^{t/2}\binom{t}{k}(1/2+\epsilon)^k(1/2-\epsilon)^{t-k}=\sum_{k=0}^{t/2}\binom{t}{k}(1/2-\epsilon)^t\left(\frac{1/2+\epsilon}{1/2-\epsilon}\right)^k \quad \leq \sum_{k=0}^{t/2}\binom{t}{k}(1/2+\epsilon)^{t/2}(1/2-\epsilon)^{t/2}.$$

多项式时间 Las Vegas 算法又称为 ZPP（Zero-error probabilistic polynomial time）。若有 Las Vegas 算法 A，执行平均时间 $EX = T$。我们可以执行算法，比如，$3T$ 次，则由 Markov 不等式，失败概率 $P(X \geq 3T) \leq 1/3$。这样的算法输出“Yes”总是对的。

注：P⊆RP⊆BPP，RP⊆NP。

第 1 章

独立同分布随机变量序列

§1.1 基本概念

简言之，一个随机过程可以看作是样本空间到实变量函数空间的一个映射，即对于一个试验结果 ω，有一个函数 $X(t，\omega)$ 为样本函数或者样本路径；而给定 t，$X(t，\omega)$ 是随机变量，如图1.1示例。

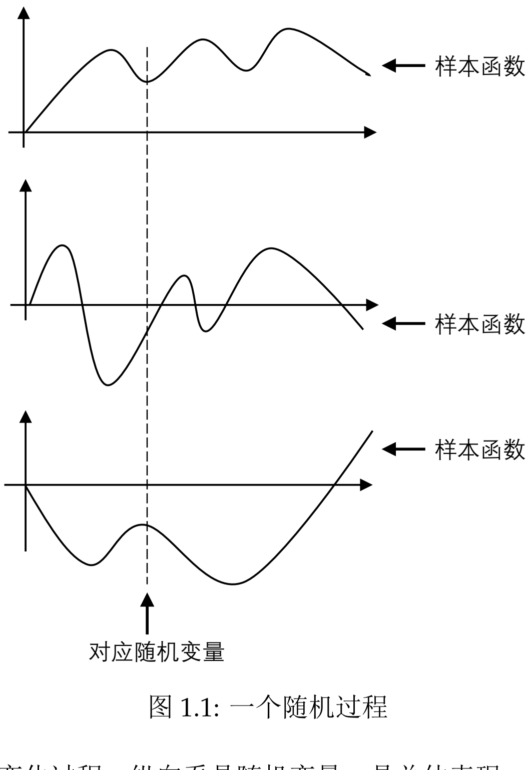


图 1.1: 一个随机过程

横向看是样本函数，随时间变化过程。纵向看是随机变量，是总体表现。

随机过程大致可分为时间离散、时间连续两类，而每一类又可以根据性质进一步细分。刻画一个随机过程，原则上要给出任意时刻对应的联合分布，即，对于任意给定的 n 个时刻，$-\infty < t_1 < t_2 < \ldots < t_n < +\infty$，我们能描述随机向量 $(X(t_1)，X(t_2)，\ldots，X(t_n))$ 的联合分布律。这对于一般随机过程是不可行的。

我们这章考虑如下随机过程，称之为独立同分布（i.i.d.，independently identically distributed）随机变量序列。

（1）时间是离散的,即 $X(t)$ 中的 t 仅取正整数值（有时取非负整数值),可以记为 X_1，X_2，$\ldots$，X_n，$\ldots$

（2）任意 n 个时刻，$t_1 < t_2 < \ldots < t_n$，

$$F_{X_1,\ X_2,\ \ldots,\ X_n}(x_1,\ x_2,\ \ldots x_n) = \prod_{j=1}^{n} F_{X_j}(x_j),$$

即满足独立性。

（3）任意 n，有 $F_{X_n}(x) = F_{X_1}(x)$，$x \in \mathbb{R}$，即同分布。

【例题 1.1】 设 X_1，X_2，...，X_n，... 是独立同分布 (i.i.d.) 连续随机变量，密度函数是 $f_X(x)$。对于 $n \geq 2$，若 $X_n > X_i$ 对于所有 $i < n$ 成立，称 X_n 是记录。求：

（1）X_2 是记录的概率；

（2）X_n 是记录的概率；

（3）前 m 次试验中平均多少次记录；

（4）N_1 是第一次记录产生的时刻，问 $P\{N_1 > n\}$，$n \geq 2$。

（5）证明 $\mathbf{E}N_1 = \infty$。

§1.2 随机变量的极限

设 A_1，A_2，$\ldots$，A_n，$\ldots$ 是事件序列，我们可以考虑如下概率极限。

1. 当 $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq \ldots$，即 A_n，$n \geq 1$ 是不减事件列时，则

$$P\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \lim_{n\to\infty} P\{A_n\}。$$

   上述右端的极限是存在的，因为 $P(A_n)$，$n \geq 1$ 是单调数序且有界。该性质的证明用到可列可加性与减法律，因 $\bigcup\limits_{n=1}^{\infty} A_n = A_1 + A_2 \backslash A_1 + A_3 \backslash A_2 + \ldots$。

2. 由性质 1，我们可以证明当 $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$，即 A_n 是不增事件列时，则

$$P\left\{\bigcap_{n=1}^{\infty} A_n\right\} = \lim_{n\to\infty} P\{A_n\}。$$

上述两种情形，我们可以定义事件的极限事件。若 $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq \ldots$，我们令 $A = \bigcup\limits_{n=1}^{\infty} A_n$，称之为 A_n，$n \geq 1$ 的极限事件，记为 $A_n \uparrow A$。若 $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$，我们令 $A = \bigcap\limits_{n=1}^{\infty} A_n$，称之为 A_n，$n \geq 1$ 的极限事件，记为 $A_n \downarrow A$。上述两种特殊情况下的极限事件相应于数列极限中的单调数列的极限。类比于数列，我们可以定义事件序列的上极限事件与下极限事件。

- 上极限事件 $\limsup A_n = \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k$。
- 下极限事件 $\liminf A_n = \bigcup\limits_{n=1}^{\infty} \bigcap\limits_{k=n}^{\infty} A_k$。

上极限事件 $\limsup A_n$ 表示的是有无穷多 A_n 发生的事件，而下极限事件 $\liminf A_n$ 表示的是只有有限个 A_n 不发生的事件。上极限事件 $\limsup A_n$ 又记作 A_n，i.o.，我们有以下引理。

【引理 1.2】Borel-Cantelli 引理: 设 A_1，A_2，$\ldots$，A_n，$\ldots$ 是事件序列,若 $\sum_i P(A_i) < \infty$,则 $P(A_i，\text{i.o.}) = 0$。若 $\sum_i P(A_i) = \infty$ 且 A_1，A_2，$\ldots$，A_n，$\ldots$ 独立，则 $P(A_i，\text{i.o.}) = 1$。

下面回顾随机变量的四种收敛类型。

【定义 1.3】 如果对 F 的任意连续点 $x \in \mathbb{R}$，都有

$$\lim_{n\to\infty} F_n(x) = F(x),$$

称随机变量序列 X_1，X_2，... 依分布收敛 (convergence in distribution) 于随机变量 X，其中 F_n 和 F 分别是随机变量 X_n 和 X 的累积分布函数。

【定义 1.4】 如果对任意 $\varepsilon > 0$，都有

$$\lim_{n\to\infty} \Pr\{|X_n - X| > \varepsilon\} = 0.$$

称随机变量序列 X_n 依概率收敛 (convergence in probability) 于随机变量 X。

【定义 1.5】 称数列随机变量序列 X_n 几乎处处收敛或具有概率 1 强收敛于 X，意味着

$$\Pr\{\lim_{n\to\infty} X_n = X\} = 1。$$

【定义 1.6】 如果

$$\lim_{n\to\infty} \mathbf{E}[(X_n - X)^2] = 0,$$

其中 **E** 表示期望运算符，称数列随机变量序列 X_n 均方收敛收敛于 X。

举例说明**依概率收敛但不以概率 1 收敛**：

【例题 1.7】 考虑一个随机试验，其输出结果 ω 均匀分布于 $[0，1)$ 区间。定义如下无穷二叉树：

（1）根节点是 $[0，1)$；

（2）若区间 $[a，b)$ 是一个节点，则其左儿子是 $[a，\frac{a+b}{2})$，右儿子是 $[\frac{a+b}{2}，b)$。

把树中的所有节点对应的区间依顺序排列，离根近的层排在前面，同一层的从左到右排。这些区间依次编号为 0，1，2，…，n，…，定义 X_n 是所对应区间的示性函数。如：

$$X_0(\omega) = 1，\omega \in [0，1);$$

$$X_1(\omega) = \begin{cases} 1 & , \quad \omega \in [0，\frac{1}{2}); \\ 0 & , \quad \omega \in [\frac{1}{2}，1); \end{cases}$$

$$X_2(\omega) = \begin{cases} 0 & , \quad \omega \in [0，\frac{1}{2}); \\ 1 & , \quad \omega \in [\frac{1}{2}，1)。 \end{cases}$$

这些收敛性的关系如图1.2所示。

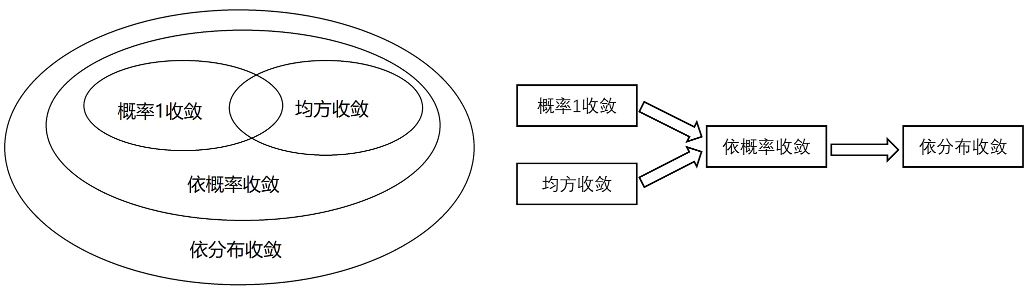


图 1.2: 收敛性关系图

§1.3 大数定律及中心极限定理

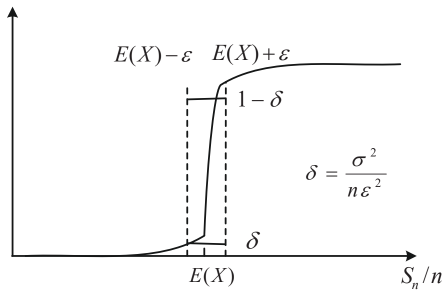
在试验不变的条件下，重复试验多次，随机事件的频率接近于它的概率，这就是大数定律。定理叙述如下。

【定理 1.8】 设 X_1，X_2，...，X_n，... 是独立同分布的，且具有数学期望 $\mathbf{E}(X) = \mu$ 和方差 σ^2，则对于任意 $\varepsilon > 0$，序列 $S_n = \sum_{k=1}^{n} X_k$ 满足（Chebyshev 不等式）：

$$P(|S_n/n - \mu| \geqslant \varepsilon) \leqslant \sigma^2/n\varepsilon^2。$$

这个定理表明：

1. 统计平均在概率意义上趋于集合平均，即 $S_n/n \xrightarrow{P} \mathbf{E}(X)$；
2. 随着 n 的增大，S_n/n 的分布函数趋于一个阶跃函数，阶跃点在 $\mathbf{E}(X)$。



【定理 1.9】 设 X_1，X_2，...，X_n，... 是独立同分布的，且具有数学期望 $\mathbf{E}(X) = \mu$，对于任意 $\varepsilon > 0$，$S_n = \sum_{i=1}^{n} X_i$ 则满足

$$P\{\lim_{n\to\infty} |S_n/n - \mu| \geqslant \varepsilon\} = 0。$$

【例题 1.10】 设 X 是指数分布，$P\{X > x\} = \exp(-x)$。取一样本 x，放入信封 A；再以 1/2 的概率计算 $y = 2x$ 或 $y = x/2$，把 y 放入信封 B。请问，选 A 还是选 B 能得到更大的数。
A 与 B 信封若混淆了。请问，选定一个之后，要不要换？

在客观实际中，一些现象受到许多相互独立的随机因素的综合影响，如果每个因素所产生的作用都很微小时，总的影响可以看作是服从正态分布的。这时，我们可以应用中心极限定理。

【定理 1.11】 设 X_1，X_2，...，X_n 是独立同分布的随机变量，具有数学期望 $\mathbf{E}(X)$ 和方差 σ^2，则序列 $S_n = \sum_{k=1}^{n} X_k$ 满足：

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mathbf{E}(X)}{\sqrt{n}\sigma} \leqslant y\right) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx。$$

用语言描述即，$\frac{S_n - n\mathbf{E}(X)}{\sqrt{n}\sigma}$ 的分布函数趋于正态分布函数（注意，前者可能没有密度函数）。

§1.4 课后作业

1. 记 $\mathbb{F}_2 = \{0,1\}$ 是二元域，$\mathbb{F}_2^n$ 是其上的线性空间。一个二元向量 $\boldsymbol{v} \in \mathbb{F}_2^n$ 的汉明重量 $W_H(\boldsymbol{v})$ 定义为 $\boldsymbol{v}$ 中 1 的个数，设 $\boldsymbol{v}_1$，$\boldsymbol{v}_2$，$\cdots \boldsymbol{v}_n$，$\cdots$ 是独立同分布的二元随机列向量，汉明重量是 1，给定 $\boldsymbol{u} \in \mathbb{F}_2^n$ 是一个行向量。讨论 $\boldsymbol{uv}_1$，$\boldsymbol{uv}_2$，$\cdots \boldsymbol{uv}_n$，$\cdots$ 的分布律，其中 $\boldsymbol{uv}$ 是行向量与列向量的积，其运算在 $\mathbb{F}_2$ 中进行。

2. 设 X_1，X_2，$\ldots$，X_n，$\ldots$ 是 Bernoulli 过程，$P_X(1) = \frac{1}{4}$，$P_X(0) = \frac{3}{4}$。

（1）求 $\mathbf{E}X$ 和 $\mathbf{D}X$。

（2）令 $S_n = \sum_{i=1}^n X_i$，$Y_n = \frac{S_n - n\mathbf{E}X}{n}$，$Z_n = \frac{S_n - n\mathbf{E}X}{\sqrt{n\mathbf{D}X}}$。取 $n = 1$，2，4，8，16，32，分别画出 S_n，Y_n，Z_n 的累积分布函数随 n 变化的趋势，体会大数定律与中心极限定理的内涵。

3. 假设 $\{X; i \geq 1\}$ 是独立同分布的二元随机变量。设 $PX_i = 1 = \delta$，$PX_i = 0 = 1-\delta$。令 $S_n = X_1 + \cdots + X_n$。让 m 是一个任意但固定的正整数。计算以下内容并解释

（1）$\lim_{n\to\infty} \sum_{i:n\delta-m\leq i\leq n\delta+m} \Pr\{S_n = i\}$；

（2）$\lim_{n\to\infty} \sum_{i:0\leq i\leq n\delta+m} \Pr\{S_n = i\}$；

（3）$\lim_{n\to\infty} \sum_{i:n(\delta-1/m)\leq i\leq n(\delta+1/m)} \Pr\{S_n = i\}$。

§1.5
拓展阅读

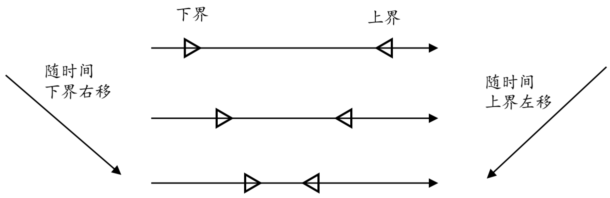
数列的上下极限与集合列的上下极限

设 a_1，a_2，$\ldots$，a_n，$\ldots$ 是一个数列，A_1，A_2，$\ldots$，A_n，$\ldots$ 是一个集合列，我们对比一下上下极限的定义。

给定 n，我们称 a_n，a_{n+1}，a_{n+2}，$\ldots$ 为序列 $\{a_i\}$，$i \geq 1$ 的尾部序列。类似地，A_n，A_{n+1}，A_{n+2}，$\ldots$ 为序列 $\{A_i\}$，$i \geq 1$ 的尾部序列。对于数列，我们可以定义其尾部列的上、下确界，即 $m_n = \inf\{a_i，i \geq n\}$，$M_n = \sup\{a_i，i \geq n\}$。显然，$m_n \leq a_i \leq M_n$，$i \geq n$。这是说尾部序列落在了 $[m_n，M_n]$ 之间。有趣的是，m_n 随着 n 增大而递增，M_n 随着 n 增大而递减，这其实是由尾部序列逐渐“变短”而得到的结论。因此，m_n，$n \geq 1$ 与 M_n，$n \geq 1$ 均是有极限的，分别记作上极限 $\overline{\lim} x_n$ 和下极限 $\underline{\lim} x_n$。

类比的，对于尾部集合序列，其“下确界”$L_n = \bigcap\limits_{k=n}^{\infty} A_k$，“上确界”$U_n = \bigcup\limits_{k=n}^{\infty} A_k$。这里“上”与“下”是对集合包含所定义的序而言的。下确界是指“最大”的下界，即 $L_n \subseteq A_k$，$k \geq n$，且若有 $B \subseteq A_k$，$k \geq n$，则 $B \subseteq L_n$。而上确界是指“最小”的上界，即 $U_n \supseteq A_k$，$k \geq n$，且若有 $B \supseteq A_k$，$k \geq n$，则 $B \supseteq U_n$。可以看到 L_n 是递增的，U_n 是递减的。所有 A_k，$k \geq n$ 都满足 $L_n \subseteq A_k \subseteq U_n$。$L_n$ 的极限即 A_n，$n \geq 1$ 的下极限，记作 $\underline{\lim} A_n$，U_n 的极限即 A_n，$n \geq 1$ 的上极限，记作 $\overline{\lim} A_n$。

不论是数列还是集合（事件）列，我们可以用下图示意尾部列的范围。



极限存在当且仅当上下极限相等。

参考文献