第五讲：概率的公理化体系

Email: xzhangseu@seu.edu.cn

东南大学 quad 数学学院

代数 (algebra) 或域 (field)

定义 0.1 (集合的代数 (algebra)) 设 $\mathcal{A}$ 是由 Ω 的某些子集所构成的集合类，称 $\mathcal{A}$ 为集合 Ω 上的代数，如果，

1. $\Omega \in \mathcal{A}, \varnothing \in \mathcal{A}$;
2. 若 $A \in \mathcal{A}$，则 $\overline{A} := \Omega - A \in \mathcal{A}$;
3. 若 $A, B \in \mathcal{A}$，则 $A \cup B \in \mathcal{A}$.

例 0.1 设 $\Omega = \{a, b, c, d\}$，则集族 $\mathcal{A} = \{\varnothing, \Omega, \{a\}, \{b, c, d\}\}$ 是代数.

例 0.2 (有限集上的代数) 设 Ω 为有限集，则集族 2^Ω (即 Ω 的所有子集构成的集合) 是代数. 例如，当 $\Omega = \{a, b, c\}$ 时，

$$\mathcal{A} = \{\varnothing, \Omega, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$$

是代数.

注 0.1

- 代数是"集合的集合"，即其本身是集合，其中的元素仍然是集合.
- 代数对于集合的求补、有限交和有限并运算封闭。
- 由于代数仅对集合的有限交并封闭，若样本空间 Ω 是无限集，将会导致它将无法包含一些重要的集合 (特别是无限集合).

事件 σ 域定义及其性质

定义 0.2 (事件 σ 域或 σ- 代数) 设 $\mathcal{F}$ 为 Ω 的某些子集构成的集合类，称 $\mathcal{F}$ 为 Ω 上的事件 σ 域或 σ- 代数，如果

1. $\Omega \in \mathcal{F}$;
2. 若 $E \in \mathcal{F}$, 则 $\overline{E} := \Omega - E \in \mathcal{F}$;
3. 若 $E_i \in \mathcal{F}, i = 1, 2, \cdots$, 则 $\cup_{i=1}^{\infty} E_i \in \mathcal{F}$.

定理 0.1 若 $\mathcal{F}$ 为 Ω 上的 σ-代数，则

- $\emptyset \in \mathcal{F}$;
- 若 $E_k \in \mathcal{F}, k = 1, 2, \cdots$, 则 $\cap_{k=1}^{\infty} E_k \in \mathcal{F}$;
- 若 $E_1, \cdots, E_n \in \mathcal{F}$, 则 $\cap_{k=1}^{n} E_k \in \mathcal{F}$ 且 $\cup_{k=1}^{n} E_k \in \mathcal{F}$;
- 若 $E_1, E_2 \in \mathcal{F}$, 则 $E_1 - E_2 \in \mathcal{F}$.

例 0.3 (代数而非 σ-代数) 设 Ω 是无限集合, 集族 $\mathcal{A} \subset 2^{\Omega}$ ，满足

$$A \in \mathcal{A} \Longleftrightarrow \# A < \infty \text{ 或} \# \overline{A} < \infty,$$

即 $\mathcal{A}$ 中的集合或者自身是有限集，或者其补集是有限集. 容易验证, $\mathcal{A}$ 是代数, 但不是 σ-代数.

例 0.4 (可数集上 σ-代数) 设样本空间 Ω 为可数集, $\mathcal{F}$ 为其上的 σ-代数, 且 $\forall a \in \Omega$, 有 $\{a\} \in \mathcal{F}$, 则 $\forall A \subset \Omega$, 有 $A \in \mathcal{F}$.

例 0.5 设 $\mathcal{F}$ 为实数集 $\mathbf{R}$ 上的 σ-代数. 若对 $\forall x \in \mathbf{R}$, 都有 $(-\infty, x] \in \mathcal{F}$, 则 $\mathcal{F}$ 包含 $\mathbf{R}$ 上的任意区间.

证明: 仅以 $[a, b)$ 为例. 注意到:

$$(-\infty, a) = \bigcup_{k=1}^{\infty} \left(-\infty, a - \frac{1}{k}\right] \in \mathcal{F}, \quad (x, \infty) = \overline{(-\infty, x]} \in \mathcal{F}, \forall x$$

$$[b, \infty) = \bigcap_{k=1}^{\infty} \left(b - \frac{1}{k}, \infty\right) \in \mathcal{F} \quad [a, b) = \overline{(-\infty, a) \cup [b, \infty)} \in \mathcal{F}$$

σ-代数例子及如何构造

- $\mathcal{F} := \{\emptyset, \Omega\}$;
- $\mathcal{F} := \{A : A \subset \Omega\}$;
- $\mathcal{F} := \{\emptyset, A, \bar{A}, \Omega\}$;
- 如何由 $\{A, B\}$(假设 A, B 为非平凡事件，互不包含且不互为对立事件) 扩充成事件 σ 域？

定理 0.2 假设 $\mathcal{C}$ 为 Ω 的子集组成的非空集类，则存在唯一 σ-代数 $\mathcal{C}_0$ 使得:

(1) $\mathcal{C} \subset \mathcal{C}_0$;

(2) 若 $\mathcal{G}$ 为任一 σ-代数，且 $\mathcal{C} \subset \mathcal{G}$, 则 $\mathcal{C}_0 \subset \mathcal{G}$.

证明:

- 若 $\mathcal{F}_t, t \in T$ 为一族 σ-代数，则 $\cap_{t\in T}\mathcal{F}_t$ 仍为 σ-代数;
- 令 $C^{\#} := \{\mathcal{B} : \mathcal{B}\text{为}\sigma - \text{代数}, \mathcal{C} \subset \mathcal{B}\}$;
- 令 $\mathcal{C}_0 := \cap_{\mathcal{B}\in C^{\#}}\mathcal{B}$, 则易证 $\mathcal{C}_0$ 即为定理所求之 σ-代数.

注 0.2

- $\mathcal{C}_0 := \cap_{\mathcal{B}\in C^{\#}}\mathcal{B}$ 是包含集类 $\mathcal{C}$ 的最小 σ-代数;
- 一般也称 $\mathcal{C}_0$ 为集类 $\mathcal{C}$ 的生成 σ-代数，后面我们常用 $\sigma(\mathcal{C})$ 表示 $\mathcal{C}_0$.

例 0.6 令 $\Omega := \mathbb{R}$, 考虑

$$\mathcal{A}_1 = \{(-\infty, x) | x \in \mathbb{R}\},\ \mathcal{A}_2 = \{(a, b) | -\infty < a < b < \infty\},$$

则 $\sigma(\mathcal{A}_1) = \sigma(\mathcal{A}_2)$

解: 注意到，对任意实数 x, 均有

$$(-\infty, x) = \cup_{n=1}^{\infty}(x - n, x).$$

因此 $(-\infty, x) \in \sigma(\mathcal{A}_2)$, 从而 $\mathcal{A}_1 \subset \sigma(\mathcal{A}_2)$, 进而可得 $\sigma(\mathcal{A}_1) \subset \sigma(\mathcal{A}_2)$.
反之，注意到 $[a, b) = (-\infty, b) - (-\infty, a) \in \sigma(\mathcal{A}_1)$ 以及

$$\{a\} = \cap_{n=1}^{\infty}[a, a + \frac{1}{n}) \in \sigma(\mathcal{A}_1),$$

因此，$(a, b) = [a, b) - \{a\} \in \sigma(\mathcal{A}_1)$, 即 $\mathcal{A}_2 \subset \sigma(\mathcal{A}_1)$, 从而

$$\sigma(\mathcal{A}_2) \subset \sigma(\mathcal{A}_1).$$

例 0.7 设 $\Omega = \mathbb{R}$, 令

$$
\begin{aligned}
\mathcal{C}_1 &= \{(-\infty, x) : -\infty < x < +\infty\} \\
\mathcal{C}_2 &= \{(a, b] : -\infty \le a \le b < +\infty\} \\
\mathcal{C}_3 &= \{[a, b] : -\infty < a \le b < +\infty\} \\
\mathcal{C}_4 &= \{[a, b) : -\infty < a \le b \le +\infty\} \\
\mathcal{C}_5 &= \{(a, b) : -\infty \le a \le b \le +\infty\} \\
\mathcal{C}_6 &= \{A \subset \mathbb{R} : A\text{为开集}\} \\
\mathcal{C}_7 &= \{A \subset \mathbb{R} : A\text{为闭集}\}
\end{aligned}
$$

则以上 7 个集类的生成 σ-代数均相同，称以上集类的生成 σ-代数中的元素为 Borel 集，并记作 $\mathcal{B}(\mathbb{R})$, 即

$$\sigma(\mathcal{C}_1) = \cdots = \sigma(\mathcal{C}_7) := \mathcal{B}(\mathbb{R}).$$

其他集类

定义 0.3 设 $\mathcal{C}$ 为 Ω 的某些子集构成的非空集类,

- 称 $\mathcal{C}$ 为 π 类, 如果 $\mathcal{C}$ 关于有限交运算封闭, 即

$$A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C};$$

- 称 $\mathcal{C}$ 为代数, 如果 $\mathcal{C}$ 关于有限交与取余运算封闭, 即

$$A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}, \bar{A} \in \mathcal{C};$$

- 称 $\mathcal{C}$ 为单调类, 如果 $\mathcal{C}$ 关于单调极限运算封闭, 即

$$A_n \in \mathcal{C}, A_n \uparrow A \text{ 或} A_n \downarrow A \Rightarrow A \in \mathcal{C}$$

- 称 $\mathcal{C}$ 为 λ 类, 如果 $\Omega \in \mathcal{C}$ 并且 $\mathcal{C}$ 对真差运算及上升极限运算封闭, 即
  - $\Omega \in \mathcal{C}$;
  - $A, B \in \mathcal{C}, B \subset A \Rightarrow A - B \in \mathcal{C}$;
  - $A_n \in \mathcal{C}, n \geq 1, A_n \uparrow A \Rightarrow A \in \mathcal{C}$

定理 0.3 设 $\mathcal{C}$ 为 Ω 的某些子集构成的非空集类, 则

1. 若 $\mathcal{C}$ 为 $\sigma-$ 代数, 则 $\mathcal{C}$ 一定是:π 类, 代数,λ 类, 单调类;
2. 若 $\mathcal{C}$ 既是代数又是单调类, 则 $\mathcal{C}$ 一定是 $\sigma-$ 代数;
3. 若 $\mathcal{C}$ 既是 π 类又是 λ 类, 则 $\mathcal{C}$ 一定是 $\sigma-$ 代数.

定义 0.4 与生成 $\sigma-$ 代数类似, 我们可以定义包含集类 $\mathcal{C}$ 的最小单调类 $m(\mathcal{C})$ 及最小 λ 类 $\lambda(\mathcal{C})$, 分别称为由集类 $\mathcal{C}$ 生成的单调类与 λ 类.

定理 0.4 (单调类定理) 设 $\mathcal{C},\mathcal{G}$ 为 Ω 中的集类且 $\mathcal{C}\subset\mathcal{G}$, 则

1. 若 $\mathcal{C}$ 为一代数, 则 $m(\mathcal{C})=\sigma(\mathcal{C})$;
2. 若 $\mathcal{C}$ 为一 π 类, 则 $\lambda(\mathcal{C})=\sigma(\mathcal{C})$;
3. 若 $\mathcal{C}$ 为代数而 $\mathcal{G}$ 为单调类, 则 $\sigma(\mathcal{C})\subset\mathcal{G}$;
4. 若 $\mathcal{C}$ 为 π 类而 $\mathcal{G}$ 为 λ 类, 则 $\sigma(\mathcal{C})\subset\mathcal{G}$.

如何应用单调类定理

- 问题: 如何证明某 $\sigma-$ 代数 $\mathcal{F}$ 中的所有元素都具有某种性质 (P)?
- 步骤:
  - $\mathcal{G} := \{A \in \mathcal{F} : A$具有性质$(P)\}$;
  - 找到某一集类 $\mathcal{C}$ 为代数或 π 类且具有以下性质:

$$\mathcal{C} \subset \mathcal{G}, \quad \sigma(\mathcal{C}) = \mathcal{F};$$

  - 证明 $\mathcal{G}$ 为单调类或 λ 类;
  - 由单调类定理可知:

$$\mathcal{F} = \sigma(\mathcal{C}) \subset \mathcal{G} \subset \mathcal{F}$$

即

$$\mathcal{G} = \mathcal{F}$$

从而 $\mathcal{F}$ 中所有的元素都具有某种性质 (P).

概率的公理化定义

定义 0.5 设 $\mathcal{F}$ 为样本空间 Ω 上的 σ-代数，如果定义在 $\mathcal{F}$ 上的集函数 $P(\cdot)$ 具有:

1. 非负性: $P(A) \geq 0, \forall A \in \mathcal{F}$;
2. 规范性: $P(\Omega) = 1$;
3. 可列可加性：对 $\mathcal{F}$ 中的任何两两互不相容事件列 $\{A_n\}$ 有

$$P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$$

则称 $P(\cdot)$ 为 $\mathcal{F}$ 上的概率测度，而 $P(A)$ 称为事件 A 的概率，$(\Omega, \mathcal{F}, P)$ 称为概率空间.

如何定义概率

- 概率定义在 $\sigma-$ 代数 $\mathcal{F}$ 上, 但是由于 $\mathcal{F}$ 中集合的复杂性, 直接对每个集合赋予概率存在困难;
- 首先在 $\mathcal{F}$ 的子族 $\mathcal{A} \subset \mathcal{F}$ 上定义概率 (一般 $\mathcal{A}$ 中的集合较为简单), 然后再将其扩张到 $\mathcal{F}$ 上;
- 如何选择恰当的 $\mathcal{A}$, 使得 $\mathcal{F} = \sigma(\mathcal{A})$?

例 0.8(有限集合上的概率) 设样本空间 $\Omega=\{a,b,c,d\}$, σ-代数 $\mathcal{F}$ 由集族 $\mathcal{A}$ 生成, 即 $\mathcal{F}=\sigma(\mathcal{A})$, $\quad \mathcal{A}=\{\{a\},\{a,b\},\{a,b,c\},\Omega\}$. 定义

$$P(\{a\})=\frac{1}{4},P(\{a,b\})=\frac{1}{2},P(\{a,b,c\}=\frac{3}{4}).$$

例 0.9 令 $\Omega=\{a,b,c,d\}$, $\mathcal{F}=2^{\Omega}$, $\mathcal{A}=\{\{a,b\},\{a,c\},\{b,d\},\{c,d\}\}$.

- 容易看出,$\sigma(\mathcal{A})=2^{\Omega}$;
- 定义 P_1 和 P_2 为

$$P_1(\{a\})=P_1(\{d\})=\frac{1}{6},\quad P_1(\{b\})=P_1(\{c\})=\frac{1}{3}$$
$$P_2(\{a\})=P_2(\{d\})=\frac{1}{3},\quad P_2(\{b\})=P_2(\{c\})=\frac{1}{6}$$

- 显然, 它们并不相同, 但它们在 $\mathcal{A}$ 上却完全一致

$$P_1(\{a,b\})=\frac{1}{2},P_1(\{a,c\})=\frac{1}{2},P_1(\{b,d\})=\frac{1}{2},P_1(\{c,d\})=\frac{1}{2}$$
$$P_2(\{a,b\})=\frac{1}{2},P_2(\{a,c\})=\frac{1}{2},P_2(\{b,d\})=\frac{1}{2},P_2(\{c,d\})=\frac{1}{2}$$

定理 0.5 (概率扩张的唯一性) 给定样本空间 Ω 和其上的 σ-代数 $\mathcal{F}$. 若

- P_1 和 P_2 是定义在 $\mathcal{F}$ 上的两个概率;
- 集族 $\mathcal{A}$ 关于交运算封闭, 且 $\sigma(\mathcal{A})=\mathcal{F}$;
- 对任意的 $A\in\mathcal{A}$ 均有 $P_1(A)=P_2(A)$;

则对任意的 $A\in\mathcal{F}$ 一定有 $P_1(A)=P_2(A)$.

证明: 证明思路如下:

- 令 $\mathcal{G}:=\{A\in\sigma(\mathcal{A}):P_1(A)=P_2(A)\}$;
- 注意到 $\mathcal{A}$ 为 π 类且 $\mathcal{A}\subset\mathcal{G}$, 故只需证明 $\mathcal{G}$ 为 λ 类即验证即
  - $\Omega\in\mathcal{G}$;
  - $A,B\in\mathcal{G},B\subset A\Rightarrow A-B\in\mathcal{G}$;
  - $A_n\in\mathcal{G},n\geq 1,A_n\uparrow A\Rightarrow A\in\mathcal{G}$
- 利用单调类定理可知 $\mathcal{G}=\sigma(\mathcal{A})$

定理 0.6 (Carathéodory 测度扩张定理) 设 Ω 为样本空间,$\mathcal{A}$ 为代数. 如果 P_0 为定义在 $\mathcal{A}$ 上的概率, 且满足可数可加性, 那么存在唯一的定义在 $\sigma(\mathcal{A})$ 上的概率 P, 满足

$$P(A) = P_0(A), \quad \forall A \in \mathcal{A},$$

注 0.3 更一般的测度扩张定理详见测度论教材, 如严加安院士的 “测度论讲义” 等.

如何理解概率

- 概率空间定义中的样本空间 Ω 和概率 P 都很好理解，但是 σ-代数 $\mathcal{F}$ 的出现总是显得有些突兀.
- 概率是一个以事件 (集合) 为自变量的函数.
- 给定一个事件 (集合), 概率函数即赋予它一个非负实数, 用以表示该事件 (集合) 的样本点在统计实验的结果中出现的可能性大小.
- 函数的基本概念中，" 定义域" 占据重要位置. 概率函数也不例外.
- 函数的定义域中的数对 “加减乘除” 运算封闭. 类似的, 概率函数应对事件的 “交差并补” 等运算封闭.
- σ-代数 $\mathcal{F}$ 本质上就是概率函数的 “定义域”, 对自变量 (事件) 的 “交差并补” 等运算封闭..

概率测度的性质 I

定理 0.7 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间，则其上的概率 P 具有如下性质:

1. $P(\emptyset) = 0$;
2. 有限可加性：若 $A_1, \cdots, A_n$ 为 $\mathcal{F}$ 中两两互不相容事件，则

$$P(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$$

3. 可减性：若 $A, B \in \mathcal{F}$ 且 $A \subset B$，则 $P(B - A) = P(B) - P(A)$;
4. 单调性：若 $A, B \in \mathcal{F}$ 且 $A \subset B$，则 $P(A) \leq P(B)$;
5. $P(\bar{A}) = 1 - P(A)$;

6. 加法公式 (容斥原理): 对任意两个事件 A, B, 有

$$P(A \cup B) = P(A) + P(B) - P(AB). \tag{1}$$

一般的，对任意 n 个事件 $A_1, \cdots, A_n$, 有

$$\begin{aligned} P(\cup_{i=1}^n A_i) &= \sum_{i=1}^n P(A_i) - \sum_{1 \le i < j \le n} P(A_i A_j) + \sum_{1 \le i < j < k \le n} P(A_i A_j A_k) \\ &\quad + \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n). \end{aligned} \tag{2}$$

7. 次可加性 (Boole 不等式): 对任意两个事件 A, B, 有

$$P(A \cup B) \le P(A) + P(B).$$

一般的，对任意 n 个事件 $A_1, \cdots, A_n$, 有

$$P(\cup_{i=1}^n A_i) \le \sum_{i=1}^n P(A_i).$$

8. Bonferroni 不等式:

$$P(\cup_{i=1}^n A_i) \geq \sum_{i=1}^n P(A_i) - \sum_{1\leq i<j\leq n} P(A_i \cap A_j),$$

$$P(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i) - \sum_{1\leq i<j\leq n} P(A_i \cap A_j) + \sum_{1\leq i<j<k\leq n} P(A_i \cap A_j \cap A_k), \cdots.$$

9. 下连续性: 若 $A_n \in \mathcal{F}$ 且 $A_n \subset A_{n+1}, n = 1, 2, \cdots$, 则

$$P(\cup_{n=1}^\infty A_n) = P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

10. 上连续性: 若 $A_n \in \mathcal{F}$ 且 $A_n \supset A_{n+1}, n = 1, 2, \cdots$, 则

$$P(\cap_{n=1}^\infty A_n) = P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

11. 概率 P 的可列可加性等价于 P 有限可加且下连续.

注意到

$$A \cup B = A \cup (B - AB).$$

故由概率的有限可加性及减法公式可得:

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

对于一般情形，可由归纳法证明。当 $n = 2$ 时，(2) 式即为 (1) 式。设 (2) 式对 $n-1$ 成立，则先对两个事件 $\cup_{i=1}^{n-1} A_i$ 与 A_n 应用 (1) 得

$$\begin{aligned} P(\cup_{i=1}^{n} A_i) &= P(\cup_{i=1}^{n-1} A_i) + P(A_n) - P((\cup_{i=1}^{n-1} A_i) \cap A_n) \\ &= P(\cup_{i=1}^{n-1} A_i) + P(A_n) - P((\cup_{i=1}^{n-1} (A_i A_n)) \end{aligned} \tag{3}$$

加法公式证明 II

而

$$
\begin{aligned}
P(\cup_{i=1}^{n-1} A_i) &= \sum_{i=1}^{n-1} P(A_i) - \sum_{1\le i<j\le n-1} P(A_iA_j) + \sum_{1\le i<j<k\le n-1} P(A_iA_jA_k) \\
&\quad + \cdots + (-1)^{n-2} P(A_1A_2\cdots A_{n-1}),
\end{aligned}
$$

$$
\begin{aligned}
P((\cup_{i=1}^{n-1}(A_iA_n)) &= \sum_{1\le i\le n-1} P(A_iA_n) - \sum_{1\le i<j\le n-1} P(A_iA_jA_n) \\
&\quad + \cdots + (-1)^{n-2} P(A_1A_2\cdots A_n) \\
&= \sum_{1\le i<j=n} P(A_iA_n) - \sum_{1\le i<j<k=n} P(A_iA_jA_n) \\
&\quad + \cdots + (-1)^{n-2} P(A_1A_2\cdots A_n)
\end{aligned}
$$

故将上面两式代入 (3) 式即得加法公式.

Boole 不等式的证明：归纳法

- $n=2$ 时显然有

$$P\left(A_1 \cup A_2\right)=P\left(A_1\right)+P\left(A_2\right)-P\left(A_1 \cap A_2\right) \leqslant P\left(A_1\right)+P\left(A_2\right),$$

- 假定 $n=m$ 时结论成立，那么 $n=m+1$ 时，有

$$\begin{aligned}
P(\bigcup_{i=1}^{m+1} A_i) &\leq P(\bigcup_{i=1}^{m} A_i)+P(A_{m+1}) \\
&\leqslant \sum_{i=1}^{m} P(A_i)+P(A_{m+1}) \\
&= \sum_{i=1}^{m+1} P(A_i)
\end{aligned}$$

Boole 不等式证明: 非归纳法

- 注意到:

$$\begin{aligned}\cup_{i=1}^{n} A_i &= A_1 \cup (A_2 - A_1) \cup (A_3 - (A_1 \cup A_2)) \cup \cdots \cup (A_n - \cup_{i=1}^{n-1} A_i) \\ &= A_1 \cup (A_2\overline{A}_1) \cup (A_3\overline{A}_2\overline{A}_1)) \cup \cdots \cup (A_n\overline{A}_{n-1}\cdots\overline{A}_1) \\ &= A_1 \cup \cup_{i=2}^{n} (A_i\overline{A}_{i-1}\cdots\overline{A}_1)\end{aligned}$$

- 从而

$$\begin{aligned}P(\cup_{i=1}^{n} A_i) &= P(A_1) + \sum_{i=2}^{n} P(A_i\overline{A}_{i-1}\cdots\overline{A}_1) \\ &= P(A_1) + \sum_{i=2}^{n} \left[P(A_i) - P(A_i \cap (A_{i-1} \cup \cdots \cup A_1))\right] \\ &= \sum_{i=1}^{n} P(A_i) - \sum_{i=2}^{n} P(\cup_{j=1}^{i-1}(A_iA_j)) \leq \sum_{i=1}^{n} P(A_i)\end{aligned}$$

Bonferroni 不等式证明: 非归纳法

- 注意到

$$P(\cup_{j=1}^{i-1}(A_iA_j)) \le \sum_{j=1}^{i-1} P(A_iA_j)$$

- 从而

$$\begin{aligned}
P(\cup_{i=1}^{n}A_i) &= \sum_{i=1}^{n} P(A_i) - \sum_{i=2}^{n} P(\cup_{j=1}^{i-1}(A_iA_j)) \\
&\ge \sum_{i=1}^{n} P(A_i) - \sum_{i=2}^{n}\sum_{j=1}^{i-1} P(A_iA_j) \\
&= \sum_{i=1}^{n} P(A_i) - \sum_{1\le j<i\le n} P(A_iA_j).
\end{aligned}$$

- 再由$P(\cup_{j=1}^{i-1}(A_iA_j)) \ge \sum_{j=1}^{i-1} P(A_iA_j) - \sum_{1\le j<k\le i-1} P(A_iA_j \cap A_iA_k)$ 可得

$$P(\cup_{i=1}^{n}A_i) \le \sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_i\cap A_j) + \sum_{1\le i<j<k\le n} P(A_i\cap A_j\cap A_k)$$

概率 P 的下连续性

设 F_n 是 $\mathcal{F}$ 中的一个单调不减的事件序列，即

$$\lim_{n\to\infty} F_n = \cup_{i=1}^{\infty} F_i$$

若令 $F_0 = \emptyset$, 则

$$\cup_{i=1}^{\infty} F_i = \cup_{i=1}^{\infty} (F_i - F_{i-1}).$$

由于 $F_{i-1} \subset F_i$, 显然诸 $F_i - F_{i-1}$ 互不相容，故由可列可加性得

$$\begin{aligned} P(\cup_{i=1}^{\infty} F_i) &= P(\cup_{i=1}^{\infty} (F_i - F_{i-1})) = \sum_{i=1}^{\infty} (P(F_i) - P(F_{i-1})) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} (P(F_i) - P(F_{i-1})) = \lim_{n\to\infty} P(F_n). \end{aligned}$$

设 $\{E_n\}$ 是一列单调不增的事件序列，则 $\{\bar{E}_n\}$ 为单调不减事件序列，由概率的下连续性可得

$$P(\cup_{i=1}^{\infty}\bar{E}_i) = \lim_{n\to\infty} P(\bar{E}_n) = 1 - \lim_{n\to\infty} P(E_n) \tag{4}$$

另一方面，

$$P(\cup_{i=1}^{\infty}\bar{E}_i) = P(\overline{\cap_{i=1}^{\infty}E_i}) = 1 - P(\cap_{i=1}^{\infty}E_i) = 1 - P(\lim_{n\to\infty} E_n) \tag{5}$$

综合 (4) 式与 (5) 式可得

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$

可列可加 $\Leftrightarrow$ 有限可加且下连续

$\Rightarrow$, 显然，下证 $\Leftarrow$。

设 $A_i \in \mathcal{F}, i = 1, 2, \cdots$, 是两两互不相容的事件序列，则由有限可加性知，对任意的 n 均有 $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$, 等式两边同时令 $n \to \infty$ 可得

$$\lim_{n\to\infty} P(\cup_{i=1}^n A_i) = \lim_{n\to\infty} \sum_{i=1}^n P(A_i) = \sum_{i=1}^\infty P(A_i),$$

令 $F_n := \cup_{i=1}^n A_i$, 则易有 F_n 为单调不减事件序列，故有概率的下连续性可得

$$\lim_{n\to\infty} P(\cup_{i=1}^n A_i) = \lim_{n\to\infty} P(F_n) = P(\cup_{n=1}^\infty F_n) = P(\cup_{i=1}^\infty A_i)$$

例 0.10 (Bernoulli 概率空间) 取 $\mathcal{F}=\{\emptyset, A, \bar{A}, \Omega\}$, 其中 A 为 Ω 的非空真子集。任取两个正数 $p, q(p+q=1)$, 令

$$P(\emptyset)=0, P(A)=p, P(\bar{A})=q, P(\Omega)=1.$$

易证 P 是一个概率测度。从而 $(\Omega, \mathcal{F}, P)$ 是一个概率空间.

例 0.11 (有限概率空间) 样本空间是有限集 $\Omega=\{\omega_1, \cdots, \omega_n\}$. $\mathcal{F}$ 取为 Ω 的一切子集 (共 2^n 个) 组成的集类。设 $\{p_i\}_{1\le i\le n}$ 为满足 $\sum_{i=1}^n p_i=1$ 的 n 个非负实数。定义集函数 $P(\cdot)$ 如下

$$P(A)=\sum_{\omega_i\in A} p_i.$$

易证 P 为概率测度，称此 $(\Omega, \mathcal{F}, P)$ 为有限概率空间。特别的若取 $p_i=\frac{1}{n}, 1\le i\le n$, 则为古典概率空间.

几个概率空间的例子

- 例 0.12 (离散概率空间) 样本空间 $\Omega=\{\omega_1,\omega_2,\cdots\}$ 为可列集，$\mathcal{F}$ 取为 Ω 的所有子集所构成的集类。设 $\{p_i\}_{i\geq 1}$ 为满足 $\sum_{i=1}^{\infty}p_i=1$ 的一列非负实数。对任意的 $E\subset\Omega$, 定义集函数 $P(\cdot)$ 如下

$$P(E)=\sum_{j:\omega_j\in E}p_j.$$

易证 P 是一个概率测度，于是 $(\Omega,\mathcal{F},P)$ 是一个概率空间，称为离散概率空间.

- 例 0.13 (一维几何概率空间) 设样本空间 Ω 为实直线上的具有正 Lebesgue 测度的某个 Borel 集，取 $\mathcal{F}$ 为 Ω 的所有 Borel 子集所构成的类。对每个 $E\in\mathcal{F}$, 令

$$P(E)=\frac{m(E)}{m(\Omega)}$$

其中 $m(E)$ 和 $m(\Omega)$ 分别表示集合 E 与 Ω 的 Lebesgue 测度。易证 $P(\cdot)$ 为概率测度.

例 0.14 口袋中有编号为 $1, 2, \cdots, n$ 的 n 个球，从中有放回地任取 m 次，求取出 m 个球最大号码为 k 的概率.

解: 令

$$
\begin{aligned}
A_k &:= \{\text{取出的}m\text{个球的最大号码为}k\}, \\
B_k &:= \{\text{取出的}m\text{个球的最大号码为小于等于}k\}
\end{aligned}
$$

则易有 $A_k = B_k - B_{k-1}$ 且 $P(B_k) = \frac{k^m}{n^m}$.另一方面，显然有 $B_{k-1} \subset B_k$，故

$$
P(A_k) = P(B_k - B_{k-1}) = P(B_k) - P(B_{k-1}) = \frac{k^m - (k-1)^m}{n^m}
$$

例 0.15 口袋中有 $n-1$ 个黑球，1 个白球，每次从口袋中随机地摸出一球，并换入一只黑球，求第 k 次摸到黑球的概率.

解: 记 $A :=$ {第k次取到黑球},则 A 的对立事件为

$$\overline{A} = \{\text{第}k\text{次取到白球}\}$$

这意味着，第 1 次，$\cdots$，第 $k-1$ 次只能取到黑球，而第 k 次取到白球.

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{(n-1)^{k-1}}{n^k}$$

巧用对立事件解决无空盒问题

例 0.16 将 m 个不同的小球等可能地放入 n 个不同的盒子，$m>n$，试求无空盒出现的概率.

- 易知 $|\Omega|=n^m$，记 $E=\{无空盒出现\}$，但 $|E|$ 不易求得;
- 记 $A_k=\{第k号盒子是空盒\}$，$|A_k|, |A_{j_1}\cdots A_{j_i}|$ 均易求;
- E 与 A_k 的关系: $\overline{E}=\bigcup\limits_{k=1}^{n}A_k$;
- 加法公式:

$$P(\overline{E})=P\left(\cup_{k=1}^{n}A_k\right)=\sum_{i=1}^{n}(-1)^{i-1}\sum_{1\le j_1<\cdots<j_i\le n}P(A_{j_1}\cdots A_{j_i})$$

巧用对立事件解决无空盒问题

- 计算概率 $P(A_{j_1}\cdots A_{j_i})$:

$$P(A_{j_1}\cdots A_{j_i})=\frac{(n-i)^m}{n^m}=\left(1-\frac{i}{n}\right)^m;$$

- $P(\overline{E})=\sum_{i=1}^{n}(-1)^{i-1}C_n^i\left(1-\frac{i}{n}\right)^m$;
- $P(E)=1-P(\overline{E})=\sum_{i=0}^{n}(-1)^{i}C_n^i\left(1-\frac{i}{n}\right)^m$.

例 0.17 向画满间隔为 a 的平行直线的桌面上任投一直径为 $l(l<a)$ 的半圆形纸片，求事件 $E=\{$纸片与某直线相交$\}$ 的概率.

解: 设想把半圆形纸片拼成一个圆形，并记

$$F=\{\text{新拼的半圆形与某直线相交}\},$$

则

$$\begin{aligned} E\cup F &= \{\text{直径为}l\text{的圆形与某直线相交}\} \\ E\cap F &= \{\text{长为}l\text{的线段 (即公共直径) 与某直线相交}\} \end{aligned}$$

由前面结论可知 $P(E\cap F)=\dfrac{2l}{\pi a}$. 为计算 $P(E\cup F)$, 可令 d 表示圆心到直线的最近距离，则有 $\Omega=\{d|0\le d\le \frac{a}{2}\}, E\cup F=\{d|0\le d\le \frac{l}{2}\}$. 因此

$$P(E\cup F)=\frac{l}{a}.$$

从而: $P(E)=\dfrac{P(E)+P(F)}{2}=\dfrac{P(E\cup F)+P(E\cap F)}{2}=\dfrac{\pi l/2+l}{\pi a}.$

例 0.18 在平面上画有间距为 a 与 b 的水平直线以及垂直直线，向平面随机地抛掷长度为 l 的针，$l < \min\{a, b, a+b-\sqrt{(a+b)^2-\pi ab}\}$，试求事件 E={针与某条直线相交 } 的概率.

- 记 A={针与某条水平直线相交 }, B={针与某条垂直直线相交 };
- $E = A \cup B$, 从而 $P(E) = P(A \cup B) = P(A) + P(B) - P(AB)$;
- $P(A) = \dfrac{2l}{\pi a}, \quad P(B) = \dfrac{2l}{\pi b}$;

- 计算 $P(AB)$:
  - 设 ρ 和 δ 分别表示针的中点与水平直线的最近距离和与垂直直线的最近距离，设 θ 为针与水平直线的夹角
  - 易知

$$\Omega=\{(\rho,\delta,\theta)|0\le\rho\le\frac{a}{2},0\le\delta\le\frac{b}{2},0\le\theta\le\frac{\pi}{2}\},$$

$$AB=\{(\rho,\delta,\theta)|(\rho,\delta,\theta)\in\Omega,\rho\le\frac{l}{2}\sin\theta,\delta\le\frac{l}{2}\cos\theta\}.$$

  - 从而

$$P(AB)=\frac{m(AB)}{m(\Omega)}=\frac{\int_0^{\frac{\pi}{2}}d\theta\int_0^{\frac{l}{2}\cos\theta}d\delta\int_0^{\frac{l}{2}\sin\theta}d\rho}{\frac{a}{2}\cdot\frac{b}{2}\cdot\frac{\pi}{2}}=\frac{l^2}{\pi ab}$$

- 故由加法公式可得

$$P(E)=P(A\cup B)=P(A)+P(B)-P(AB)=\frac{2l(a+b)-l^2}{\pi ab}.$$

例 0.19 设有 n 个孩子, 每个人戴着一顶帽子. 如果他们将各自的帽子放在桌子上混杂在一起, 然后每人从中随机取一顶帽子戴上, 试问至少有一人戴对自己的帽子的概率是多少?

解: 若记

$$A_i := \{第i个人戴对自己的帽子\},$$

则所求事件概率为 $P(\cup_{i=1}^n A_i)$, 易知有

$$\begin{aligned}
P(A_i) &= \frac{(n-1)!}{n!} = \frac{1}{n}, \ \sum_{i=1}^n P(A_i) = 1,\\
P(A_iA_j) &= \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}, i \neq j,\\
\sum_{1\le i<j\le n} P(A_iA_j) &= C_n^2 \frac{1}{n(n-1)} = \frac{1}{2!},
\end{aligned}$$

配对问题一（续）

同理可得

$$\begin{aligned} P(A_iA_jA_k) &= \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)}, i \neq j \neq k, \\ \sum_{1\le i<j<k\le n} P(A_iA_jA_k) &= C_n^3 \frac{1}{n(n-1)(n-2)} = \frac{1}{3!}, \\ \cdots &, \\ P(\cap_{i=1}^n A_i) &= \frac{1}{n!}. \end{aligned}$$

故由加法公式可得

$$P(\cup_{i=1}^n A_i) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + (-1)^{n-1}\frac{1}{n!} \xrightarrow{n\to\infty} 1 - e^{-1}.$$

$$e^x = 1 + \sum_{n=1}^{\infty} \frac{x^n}{n!}.$$

配对问题二

例 0.20 接上面例子，试求下述各事件的概率：

(1) 恰有 k 个孩子戴对帽子；(2) 至少有 m 个孩子戴对帽子.

- 记 E_k ={恰有 k 个孩子戴对帽子 },A_m ={至少有 m 个孩子戴对帽子 };
- $\Omega = n!$, $|E_k| =?$, $|A_m| =?$;
- 若记 D_k ={给定的某 $n-k$ 个孩子均未戴对自己的帽子 },则有 $|E_k| = C_n^k |D_k|$
- D_k 表示给定的某 $n-k$ 个孩子未戴对自己的帽子，不涉及其余 k 个人，故可套用 $n-k$ 情形下的配对问题即

$$\frac{|D_k|}{(n-k)!} = P(D_k) = 1 - P(\overline{D}_k) = \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-k}\frac{1}{(n-k)!}$$

- 因此，

$$|D_k| = (n-k)!P(D_k) = (n-k)!\left\{\frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-k}\frac{1}{(n-k)!}\right\}$$

配对问题二

- 从而

$$
\begin{aligned}
P(E_k) &= \frac{|E_k|}{n!} = \frac{C_n^k |D_k|}{n!} \\
&= C_n^k \frac{(n-k)!}{n!} \left\{ \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-k} \frac{1}{(n-k)!} \right\} \\
&= \frac{1}{k!} \sum_{j=2}^{n-k} (-1)^j \frac{1}{j!} = \frac{1}{k!} \sum_{j=0}^{n-k} (-1)^j \frac{1}{j!}.
\end{aligned}
$$

- 在上式中令 $n \to \infty$, 可得极限概率为 $\dfrac{1}{e \cdot k!}$
- 由于 $A_m = \bigcup_{k=m}^{n} E_k$, 且事件 $E_m, E_{m+1}, \cdots, E_n$ 两两不交， 故

$$
P(A_m) = \sum_{k=m}^{n} P(E_k) = \sum_{k=m}^{n} \frac{1}{k!} \sum_{j=0}^{n-k} (-1)^j \frac{1}{j!} = \sum_{k=m}^{n} \sum_{j=0}^{n-k} (-1)^j \frac{1}{k! j!}.
$$

配对问题三

- 允许一个孩子拿多只帽子的情况: 即允许有孩子没有拿到帽子. 某一个孩子没有拿到帽子的概率是

$$P_1 = \frac{(n-1)^n}{n^n} = \left(1 - \frac{1}{n}\right)^n$$

- 孩子人数 n 并不等于帽子数目 m, 即并不是每一个孩子都有自己的帽子. 某一个孩子没有拿到帽子的概率是

$$\widetilde{P}_1 = \frac{(n-1)^m}{n^m} = \left(1 - \frac{1}{n}\right)^m$$

令 $n, m \to \infty$, 且保持 $m/n = \lambda$ 为常数, 得到 $\widetilde{P}_1 \to \exp(-\lambda)$.

- 某一个孩子恰好拿到 k 顶帽子的概率为

$$\widetilde{P}_k = \binom{m}{k} \frac{(n-1)^{m-k}}{n^m} = \frac{1}{k!} \frac{m(m-1)\cdots(m-k+1)}{(n-1)^k} \left(1 - \frac{1}{n}\right)^m \to \frac{\lambda^k}{k!} \exp(-\lambda)$$

第六讲：条件概率、全概率公式与贝叶斯公式

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

条件思考的重要性

- 条件概率说明了如何以合乎逻辑且相一致的方式将证据纳入人们对世界的理解当中.
- 添加条件是一种非常有用的解决问题的策略.
- 条件既可以作为更新判断依据的方法，也是解决问题的策略.
- 条件是统计学的灵魂.

例 0.21 抛掷一枚均匀的骰子一次，已知掷出的点数为奇数，试求点数大于 1 的概率。

- $\Omega=\{1,2,\cdots,6\}, A=\{1,3,5\}, B=\{2,3,4,5,6\}$;
- 令 $P(B|A)$ 表示“已知 A 发生的情况下 B 发生的概率”;
- $A=\{1,3,5\}$, B 发生就是 $AB=\{3,5\}$ 发生，于是

$$P(B|A)=\frac{|AB|}{|A|}=\frac{2}{3}.$$

- 在原样本空间考虑:

$$P(AB)=\frac{|AB|}{|\Omega|}=\frac{2}{6}, \quad P(A)=\frac{|A|}{|\Omega|}=\frac{3}{6}.$$

- 易验证:

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{2}{3}.$$

例 0.22 从分别写有号码 $1,2,\cdots,10$ 的 10 张卡片中随机抽取一张，已知抽出的卡片的号码不小于 3，试求其号码为偶数的概率。

- $\Omega=\{1,2,\cdots,10\}, A=\{3,4,\cdots,10\}, B=\{2,4,6,8,10\}$;
- 类似上一例:

$$P(B|A)=\frac{|AB|}{|A|}=\frac{4}{8}=\frac{1}{2}.$$

- 在原样本空间考虑:

$$P(AB)=\frac{|AB|}{|\Omega|}=\frac{4}{10}=\frac{2}{5},\quad P(A)=\frac{|A|}{|\Omega|}=\frac{8}{10}=\frac{4}{5}.$$

- 易验证:

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{1}{2}.$$

例 0.23 考察有两个小孩的家庭，其样本空间为 $\Omega := \{bb, bg, gb, gg\}$，其中 b, g 分别表男孩与女孩，bg 表示大的是男孩，小的是女孩，其他样本点类似说明. 在 Ω 中的四个样本点可能情况下，我们讨论如下事件的概率：

- 事件 $A := \{$家中至少有一个女孩$\}$ 发生的概率为 $P(A) = \dfrac{3}{4}$
- 若已知 $B := \{$家中至少有一个男孩$\}$ 发生，再求 A 发生的概率为

$$P(A|B) = \frac{2}{3}$$

- 若对上述条件概率的分子分母各除以 4，则有

$$P(A|B) = \frac{2/4}{3/4} = \frac{P(AB)}{P(B)}$$

条件概率的定义

定义 0.6 设 $(\Omega, \mathcal{F}, P)$ 为概率空间. $A \in \mathcal{F}$, $B \in \mathcal{F}$, 若 $P(B) > 0$, 则称

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

为 “在已知 B 发生的情况下，A 发生的 条件概率”, 简称条件概率.

注 0.4 条件概率的两种计算方法:

- 原则性方法: $P(A|B) = \dfrac{P(AB)}{P(B)}$
- 把 B 作为样本空间看待 (经常显得非常方便): $P(A|B) = \dfrac{|AB|}{|B|}$

条件概率的例子

例 0.24 (两张牌) 洗好一副标准扑克后。从中随机抽取两张牌，无放回地一次抽一张。设 A 事件表示第一张牌为红桃，事件 B 表示第二张牌为红色。求 $P(A|B)$ 和 $P(B|A)$.

解: 由题意易知

$$P(A\cap B) = \frac{13\cdot 25}{52\cdot 51} = \frac{25}{204}$$
$$P(B) = \frac{26\cdot 51}{52\cdot 51} = \frac{1}{2}$$
$$P(A) = \frac{1}{4}$$

从而，

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{25/204}{1/2} = \frac{25}{102}$$
$$P(B|A) = \frac{P(B\cap A)}{P(B)} = \frac{25/204}{1/4} = \frac{25}{51}$$

条件概率的注

- 条件概率大小与原事件无条件概率并没有明确的大小关系，见教材例 6.5.
- 注意哪些事件放在竖线的哪一边是非常重要的，具体来说就是 $P(A|B) \neq P(B|A)$.
- 无论 $P(A|B)$ 还是 $P(B|A)$ 都是有意义的 (直观上或数学上):
  - 牌抽取的时间顺序并不能决定出现何种条件概率.
  - 在计算条件概率时，我们考虑的是一个事件给另一个事件带来的信息，而不是一个事件是否导致了另一个事件.
- 此外，也可以通过条件概率的直接解释得出 $P(B|A) = 25/51$:
  - 如果第一张抽的牌为红桃，那么剩下的牌就由 25 张红色牌和 26 张黑色牌组成 (所有牌被下一次抽中的可能性是相同的)
  - 所以抽取一张红牌的条件概率是 $25/(25+26) = 25/51$.

定理 0.8 条件概率是概率，即若设 $P(B)>0$，则

- 非负性: $P(A|B)\geq 0, \forall A\in\mathcal{F}$;
- 规范性: $P(\Omega|B)=1$;
- 可列可加性：若 $A_n\in\mathcal{F}, n=1,2,\cdots$，互不相容，则

$$P(\cup_{n=1}^{\infty}A_n|B)=\sum_{n=1}^{\infty}P(A_n|B)$$

证明: 由条件概率的定义易证非负性与规范性。下面说明可列可加性。假设 $A_n, n=1,\cdots$，互不相容，则 $A_nB, n=1,\cdots$，也互不相容。从而

$$\begin{aligned}P(\cup_{n=1}^{\infty}A_n|B) &= \frac{P((\cup_{n=1}^{\infty}A_n)B)}{P(B)}=\frac{P(\cup_{n=1}^{\infty}(A_nB))}{P(B)}\\ &= \sum_{n=1}^{\infty}\frac{P(A_nB)}{P(B)}=\sum_{n=1}^{\infty}P(A_n|B).\end{aligned}$$

定理 0.9 (乘法公式)

- 若 $P(A)>0$, 则

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

- 若 $P(A_1A_2\cdots A_{n-1})>0$, 则

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \tag{7}$$

证明: 由条件概率的定义，移项即得 (6) 式，下证 (7). 因为

$$P(A_1)\geq P(A_2)\geq\cdots\geq P(A_1A_2\cdots A_{n-1})>0$$

故 (7) 式中条件概率均有意义。从而由两个事件的乘法公式可得

$$\begin{aligned}P(A_1A_2\cdots A_n)&=P(A_1A_2\cdots A_{n-1})P(A_n|A_1A_2\cdots A_{n-1})\\&=P(A_1\cdots A_{n-2})P(A_{n-1}|A_1\cdots A_{n-2})P(A_n|A_1A_2\cdots A_{n-1})\\&=\cdots=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})\end{aligned}$$

例 0.25 有两个罐子

- 在第一个罐中放有 7 个红球、2 个白球和 3 个黑球
- 在第二个罐中放有 5 个红球、4 个白球和 3 个黑球
- 从第一个罐中随机取出 1 个球放入第二个罐中
- 再从第二个罐中随机取出 1 个球来
- 试求 B={从第二个罐中随机取出的球为红球} 的概率

问题简要分析:

- 从第二个罐中随机取出的球这一随机试验结果依赖于第一次从第一个罐中所取球的结果;
- 因此, 无论是计算 $|B|$ 还是直接计算 $P(B)$ 都不很容易
- 若已知第一次从第一个罐中所取球的结果后, 则从第二个罐中随机取出的球为红球的概率很容易计算.

解:

- 以 A_1, A_2, A_3 分别表示由第一个罐子取出的是红球, 白球和黑球;
- 显然 $A_k, k=1,2,3$ 两两互不相容且 $\bigcup_{k=1}^{3} A_k = \Omega$;
- $B = \bigcup_{k=1}^{3} A_k B$;
- $P(B) = \sum_{k=1}^{3} P(A_k B) = \sum_{k=1}^{3} P(A_k)P(B|A_k)$;
  - $P(A_1) = \dfrac{7}{12}, P(A_2) = \dfrac{1}{6}, P(A_3) = \dfrac{1}{4}$;
  - $P(B|A_1) = \dfrac{6}{13}, P(B|A_2) = \dfrac{5}{13}, P(B|A_3) = \dfrac{5}{13}$;
- 故 $P(B) = \dfrac{7}{12} \cdot \dfrac{6}{13} + \dfrac{1}{6} \cdot \dfrac{5}{13} + \dfrac{1}{4} \cdot \dfrac{5}{13} = \dfrac{67}{156}$;

定理 0.10 (全概率公式) 设 $A_1, A_2, \cdots, A_n$ 为样本空间 Ω 的一个分割，即 $A_1, \cdots, A_n$ 互不相容，且 $\cup_{k=1}^n A_k = \Omega$，若 $P(A_k) > 0, k = 1, 2, \cdots, n$，则对任一事件 B 有

$$P(B) = \sum_{k=1}^{n} P(A_k)P(B|A_k)$$

证明：因为

$$B = B\Omega = B(\cup_{k=1}^n A_k) = \cup_{k=1}^n (BA_k)$$

且 $BA_1, BA_2, \cdots, BA_n$ 互不相容，所以由有限可加性得

$$\begin{aligned} P(B) &= P(\cup_{k=1}^n (BA_k)) = \sum_{k=1}^{n} P(BA_k) \\ &= \sum_{k=1}^{n} P(A_k)P(B|A_k) \end{aligned}$$

全概率公式图示

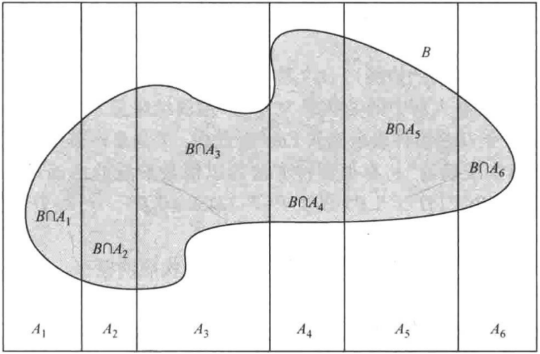


图 2.3　A_i 构成了样本空间的划分；$P(B)$ 等于 $\sum_i P(B\cap A_i)$。

全概率公式（续）

对于全概率公式，我们要注意以下几点

- 全概率公式的最简单形式：如果 $0 < P(A) < 1$，则

$$P(B) = P(A)P(B|A) + P(\overline{A})P(B|\overline{A})$$

- 条件 $A_1, A_2, \cdots, A_n$ 为样本空间 Ω 的一个分割，可改为 $A_1, A_2, \cdots, A_n$ 互不相容，且 $B \subset \cup_{k=1}^{n} A_k$
- 在必要时，可把分割的概念推广到可列个事件 $(n = \infty)$ 的情形，相应地有

$$P(B) = \sum_{k=1}^{\infty} P(A_k B) = \sum_{k=1}^{\infty} P(A_k)P(B|A_k).$$

定理 0.11 设 $A_1, A_2, \cdots, A_n$ 为样本空间 Ω 的一个分割，即 $A_1, A_2, \cdots, A_n$ 互不相容，且 $\cup_{k=1}^n A_k = \Omega$，若

$$P(B) > 0, P(A_k) > 0, k = 1, 2, \cdots, n$$

则

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{j=1}^n P(A_j)P(B|A_j)}.$$

特别的，我们有

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

贝叶斯几率

定义 0.7 (几率) 一个事件的几率 (odds) 为

$$\text{odds}(A) = P(A)/P(\overline{A}) = \frac{P(A)}{1-P(A)}$$

概率的几率表示: $P(A) = \text{odds}(A)/(1+\text{odds(A)})$

定理 0.12 (贝叶斯公式的几率形式) 对于任意两个正概率事件 A 和 B, 给定以 B 为条件的情况下, A 的几率如下:

$$\frac{P(A|B)}{P(\overline{A}|B)} = \frac{P(B|A)P(A)}{P(B|\overline{A})P(\overline{A})}$$

- 后验几率 $\frac{P(A|B)}{P(\overline{A}|B)}$ 等于先验几率 $\frac{P(A)}{P(\overline{A})}$ 乘以似然比因子 $\frac{P(B|A)}{P(B|\overline{A})}$.
- 有时候用上述形式的贝叶斯准则可以更方便地求出后验几率, 如果需要的话还可以将几率形式转换成概率形式形式.

全概率与贝叶斯内涵：由因到果，由果推因

- 在现实中把事件 B 看作结果，把事件 $A_1, A_2, \cdots, A_n$ 看作导致结果 B 的各种原因
- 全概率公式是由各种原因推理出结果事件发生的概率，是由因到果
- 在日常生活中常常是观察到某种现象，反推造成这种现象的各种原因的概率，即由果推因
- 条件概率 $P(A_k|B)$，就是在观察到结果事件 B 已经发生的情况下，推断结果事件 B 是由原因 A_k 造成的概率的大小
- $P(A_k)$：先验概率，在没有别的前提信息情况下的概率值，一般借助经验来估计，或赋予所有原因以相同的先验概率
- $P(A_k|B)$：后验概率，在获得"结果事件 B 发生"这个信息之后原因 A_k 出现的概率
- 后验概率可看作先验概率在获取了新信息之后的一种修正，贝叶斯公式恰恰提供了一种计算后验概率的工具
- 贝叶斯理论对于人工智能、深度学习等理论具有重要的指导意义，贝叶斯统计受到了从未有过的青睐，迎来了前所未有的发展机遇

条件全概率公式和条件贝叶斯公式

定理 0.13 (条件全概率公式) 令 $A_1, \cdots, A_n$ 为样本空间 Ω 的一个划分. 设对于所有的 i 满足 $P(A_iE) > 0$, 则

$$P(B|E) = \sum_{i=1}^{n} P(B|A_i, E)P(A_i|E)$$

定理 0.14 (条件贝叶斯公式) 设 $P(AE) > 0$ 且 $P(BE) > 0$, 则有

$$P(A|B, E) = \frac{P(B|A, E)P(A|E)}{P(B|E)}$$

例 0.26 (控方证人的错误) 1988 年，Sally Clark 由于她的两个孩子在出生不久便死亡，因而被指控谋杀幼童。在审讯期间，控方的一个专家证人证实

- 新生儿因婴儿猝死综合症 (SIDS) 而死亡的概率为 $1/8500$
- 所以两个新生儿由于婴儿猝死综合症死亡的概率为 $(1/8500)^2$，大约为 7300 万分之一
- 因此，他认为 Clark 清白的概率仅为 7300 万分之一.

解: 这个推理过程至少有两个问题

- 一个家庭内部成员之间死于 SIDS 是否相互独立？
  - 家庭内部成员之间死于 SIDS 相互独立时，“第一个孩子死于 SIDS” 且 “第二个孩子也死于 SIDS” 的概率是相应的两个事件概率相乘
  - 如果遗传因素或其他家庭特有的风险因素导致某些家庭内的所有新生儿面临 SIDS 的风险增加，这种独立性就不再成立了.

条件性思维谬误：控方证人的错误

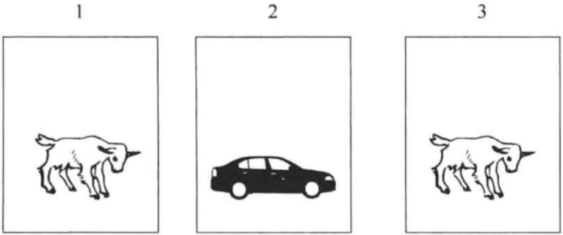
- 这个所谓的专家将两个不同的条件概率混淆了: $P(\text{清白} \mid \text{证据})$ 和 $P(\text{证据} \mid \text{清白})$ 是不一样的.
  - 专家 1 声称：如果在被告人是清白的情况下，两个孩子死亡的概率很低；那就是说 $P(\text{证据} \mid \text{清白})$ 非常小.
  - 但人们感兴趣的是，给定现在所有的证据 (孩子均死) 条件下，报告人仍清白的概率，即 $P(\text{清白} \mid \text{证据})$.
  - 由贝叶斯准则可知，

$$P(\text{清白} \mid \text{证据}) = \frac{P(\text{证据} \mid \text{清白})P(\text{清白})}{P(\text{证据})},$$

  - 所以为了计算 $P(\text{清白} \mid \text{证据})$, 这里需要考虑 $P(\text{清白})$, 也就是被告清白的先验概率。这个概率是很高的;
  - 虽然 SIDS 造成两个婴儿死亡是很罕见的，但是蓄意杀害两个婴儿的情况也很少见!
  - 基于现有证据的后验概率是对很低的 $P(\text{证据} \mid \text{清白})$ 和很高的 $P(\text{清白})$ 的一个平衡。专家的结果 $(1/8500)^2$ 是有问题的，它只是整个计算式中的一部分.

例 0.27 在 Monty Hall 主持的 “Let's Make a Deal” 节目中，有三扇门，其中有两扇门后面是一只山羊，一扇门后面是一辆车．选手将获得其所选中的那扇门后面的物品．

- 一位选手从三扇最近的门中选一扇
- Monty 知道车在哪扇门后面，并以不暴露车的位置打开剩下两扇门中的一扇，即他打开的门后面永远是山羊．
- 若剩下的两扇门后都是山羊的话，Monty 会等可能地随机选一扇门．
- 然后 Monty 会让对手选择，是换另一扇没打开的门还是不换．
- 如果对手的目标是得到车，她应该换吗？



Monty Hall 问题 (续)

解:

- 先将三扇门从 1 到 3 编号.
- 不失一般性, 可以假设选手选择的是 1 号门
- 令

$$
\begin{aligned}
A &= \{\text{换门得到车}\} \\
B &= \{\text{不换门得到车}\} \\
C_i &:= \{\text{车在第}i\text{扇门后面}\}, i = 1, 2, 3.
\end{aligned}
$$

- 显然, $P(B) = P(\text{不换门得到车}) = 1/3$
- 则由全概率公式

$$
\begin{aligned}
P(A) &= P(A|C_1) \cdot \frac{1}{3} + P(A|C_2) \cdot \frac{1}{3} + P(A|C_3) \cdot \frac{1}{3} \\
&= 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}
\end{aligned}
$$

Monty Hall 三门问题图示

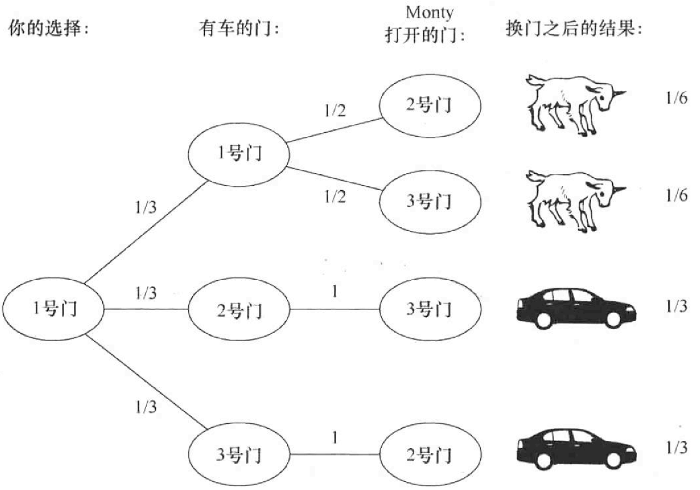


图 2.5　蒙提·霍尔（Monty Hall）问题的树状图，换门赢车的概率为 2/3。

随机徘徊的吸收概率

例 0.28 质点在数轴上整数点运动, 若质点处在整数点 i 上, 则下一时刻

- 以概率 p 向右移动到 $i+1$
- 以概率 q 向左移动到 $i-1$, 其中 $p,q>0, p+q=1$
- 我们把质点的上述运动称之为随机徘徊或随机游动
- 考虑数轴上的两个吸收壁 0 与 $a(a>1)$: 质点运动到 0 或 a 之后就永远不再移动
- 现求自 $i(0<i<a)$ 出发的质点将被 0 或 a 吸收的概率

解: 令

$E_i=\{$质点自i出发$\}, F=\{$质点在0被吸收$\}, G=\{$质点在a被吸收$\}$.
下面对 $i=0,1,\cdots,a$ 计算

$$P_i=P(F|E_i),\quad Q_i=P(G|E_i)$$

记 $B=\{$质点第一次向左移动 1 单位$\}$ 并以第 1 次可能的运动情况为条件进行全概率展开可得

$$\begin{aligned}Q_i &= P(GB|E_i)+P(G\bar{B}|E_i)\\ &= P(B|E_i)P(G|BE_i)+P(\bar{B}|E_i)P(G|\bar{B}E_i)\\ &= pQ_{i+1}+qQ_{i-1},\quad i=1,2,\cdots a-1.\end{aligned}$$

将上述递推式重写为

$$Q_{i+1}-Q_i=\frac{q}{p}(Q_i-Q_{i-1}),\quad i=1,\cdots,a-1. \tag{8}$$

1. 当 $p=q=1/2$ 时, 由 $Q_0=0$ 及 $Q_a=1$ 可得

$$Q_i=\frac{i}{a},\quad i=0,\cdots,a.$$

2. 当 $p\neq q$ 时, 反复利用递推式(8)并注意到 $Q_0=0$ 可导出

$$Q_{i+1}-Q_i=\left(\frac{q}{p}\right)^i Q_1,\quad i=1,\cdots,a-1. \tag{9}$$

上式对 $i=1,\cdots,a-1$ 求和并利用 $Q_a=1$ 可得

$$Q_1=\left[\sum_{i=0}^{a-1}\left(\frac{q}{p}\right)^i\right]^{-1}=\frac{1-q/p}{1-(q/p)^a}$$

再将 (9) 式对 $i=1,\cdots,i-1$ 求和并将上述 Q_1 代入可得

$$Q_i=\frac{1-(q/p)^i}{1-(q/p)^a},\quad i=0,\cdots,a.$$

综上可得

$$Q_i = \begin{cases} \dfrac{i}{a}, & p = q, \\ \dfrac{1-(q/p)^i}{1-(q/p)^a}, & p \neq q. \end{cases}$$

类似可得

$$P_i = \begin{cases} \dfrac{a-i}{a}, & p = q, \\ \dfrac{(q/p)^i-(q/p)^a}{1-(q/p)^a}, & p \neq q. \end{cases}$$

例 0.29 有两种治疗肾结石的方案，其治疗效果如下：

- 方案 1：小结石患者占 25%，大结石患者占 75%，小结石患者的治愈率是 93%，大结石患者的治愈率是 73%
- 方案 2：小结石患者占 77%，大结石患者占 23%，小结石患者的治愈率是 87%，大结石患者的治愈率是 69%
- 不管是对小结石患者，还是大结石患者，方案 1 的治愈率都要高于方案 2
- 方案 1 优于方案 2 吗?

解： 计算两种方案的治愈率

- 记 A={患者是小结石患者}, B={患者被治愈}
- 根据全概率公式

$$P_1(B) = P_1(A)P_1(B|A) + P_1(\overline{A})P_1(B|\overline{A}) = 0.25{\cdot}0.93 + 0.75{\cdot}0.73 = 0.78$$

$$P_2(B) = P_2(A)P_2(B|A) + P_2(\overline{A})P_2(B|\overline{A}) = 0.77{\cdot}0.87 + 0.23{\cdot}0.69 = 0.8286$$

- 方案 2 的治愈率高于方案 1，可见方案 1 并不优于方案 2

敏感性问题调查方案的关键在于要使被调查者愿意作出真实的回答又能保守个人秘密，如果调查方案有误，被调查者就会拒绝配合，所得调查数据将失去真实性.

调查方案: 被调查者只需要加答以下两个问题中的一个问题，且只需要回答“是”或“否”.

问题 A: 你的生日是否在 7 月 1 日之前?

问题 B: 所调查的敏感性问题.

调查方案的操作:

(1) 被调查者在没有旁人的情况下独自一人在房间内操作回答问题

(2) 被调查者从一个罐子中随机抽一球，看过颜色后放回，若抽到白球，回答问题 A，抽到红球，回答问题 B.

(3) 被调查者无论回答问题 A 还是问题 B，只需在仅有“是”与“否”选项的答卷上作答然后将答卷放入密封的投票箱内.

敏感性问题调查

问题: 假若我们有 n 张问卷, 其中 k 张回答“是”, 我们如何确定选定红球回答问题 B 为“是”的概率?

已知:

- 红白球的比例, 即 $P(\text{红球}) := \pi, P(\text{白球}) = 1 - \pi$;
- $P(\text{是}|\text{白球}) = 0.5$;
- $P(\text{是}) \approx \dfrac{k}{n}$.

待求: $p := P(\text{是}|\text{红球})$.

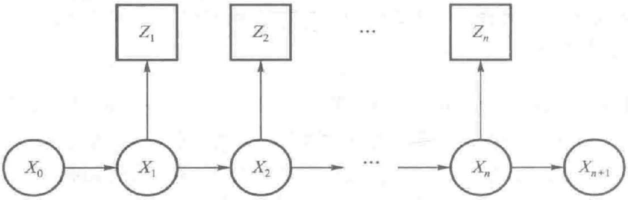
解: 由全概率公式可得

$$P(\text{是}) = P(\text{白球})P(\text{是}|\text{白球}) + P(\text{红球})P(\text{是}|\text{红球})$$

即

$$\frac{k}{n} \approx (1-\pi)\cdot 0.5 + \pi \cdot p \Rightarrow p = \frac{k/n - 0.5(1-\pi)}{\pi}$$

- 许多工程科学类问题都可以归结为如下图所示形式



- 系统自身的状态 X 是我们关注的随机事件，但是由于技术手段等方面的限制，无法直接对 X 进行观测，只能获取另一随机事件，即间接反映系统状态 X 的观测量 Z.
- 问题转化为已知观测量 Z 的条件下，如何对系统状态 X 进行推断. 用条件概率的语言讲，就是计算 $P(X \mid Z)$.
- 通常情况下，X 和 Z 随时间变化，分别记作 X_n 和 Z_n ，其中 n 表示时间. 那么在实际应用中，到 n 时刻为止，我们掌握的实际观测数据为 $\{Z_1, \cdots, Z_n\}$,

- 如何实现 $P(X_n|Z_1,\cdots,Z_n) \longrightarrow P(X_{n+1}|Z_1,\cdots,Z_{n+1})$ 的递推计算?
- 注意到, 若已知 X_{n+1} 时,Z_{n+1} 与 $\{Z_1,\cdots,Z_n\}$ 独立, 则

$$
\begin{aligned}
P(X_{n+1}|Z_1,\cdots,Z_{n+1}) &= \frac{P(X_{n+1},Z_1,\cdots,Z_{n+1})}{P(Z_1,\cdots,Z_{n+1})} \\
&= \frac{P(Z_{n+1}|X_{n+1},Z_1,\cdots,Z_n)\,P(X_{n+1},Z_1,\cdots,Z_n)}{P(Z_1,\cdots,Z_{n+1})} \\
&= \frac{P(Z_{n+1}|X_{n+1})\,P(X_{n+1},Z_1,\cdots,Z_n)}{P(Z_{n+1}|Z_1,\cdots,Z_n)P(Z_1,\cdots,Z_{n+1})}
\end{aligned}
$$

Bayesian 滤波

- 再由

$$\begin{aligned} & P\left(X_{n+1}, Z_1, \cdots, Z_n\right) \\ = & \sum_{X_n} P\left(X_{n+1}, X_n, Z_1, \cdots, Z_n\right) \\ = & \sum_{X_n} P\left(X_{n+1} \mid X_n, Z_1, \cdots, Z_n\right) P\left(X_n \mid Z_1, \cdots, Z_n\right) P\left(Z_1, \cdots, Z_n\right) \\ = & \sum_{X_n} P\left(X_{n+1} \mid X_n\right) P\left(X_n \mid Z_1, \cdots, Z_n\right) P\left(Z_1, \cdots, Z_n\right) \end{aligned}$$

- 将上式代入前面一页表达式中可得

$$P\left(X_{n+1} \mid Z_1, \cdots, Z_{n+1}\right)=\frac{P\left(Z_{n+1} \mid X_{n+1}\right)}{P\left(Z_{n+1} \mid Z_n, \cdots, Z_1\right)} \sum_{X_n} P\left(X_{n+1} \mid X_n\right) P\left(X_n \mid Z_1, \cdots, Z_n\right)$$

- 上式便是对系统状态进行递推估计的 Bayesian 滤波公式，在雷达、声呐、通信、导航、机器学习等许多领域有广泛应用.

第七讲：条件概率、全概率与贝叶斯公式的应用

Email: xzhangseu@seu.edu.cn

东南大学　　数学学院

条件概率的例子：男孩女孩概率问题

例 0.30 (年长的是女孩 vs 至少一个女孩) 某家庭有两个孩子，已知至少有一个是女孩。两个孩子都是女孩的概率是多少？如果条件改为年长的孩子是女孩，那么两个都是女孩的概率又是多少？

解: 假设每个孩子都是女孩和男孩的可能性相同且不相关，那么

$$P(\text{都是女孩} \mid \text{至少有一个是女孩}) \tag{10}$$

$$= \frac{P(\text{都是女孩，至少有一个是女孩})}{P(\text{至少有一个是女孩})} = \frac{1/4}{3/4} = 1/3 \tag{11}$$

$$P(\text{都是女孩} \mid \text{年长的是女孩}) \tag{12}$$

$$= \frac{P(\text{都是女孩，年长的是女孩})}{P(\text{年长的是女孩})} = \frac{1/4}{1/2} = 1/2 \tag{13}$$

男孩女孩概率问题续：随机的一个孩子是女孩

例 0.31 某家庭有两个孩子。随机遇到其中的一个，发现是女孩。给定这个信息后，两个孩子都是女孩的概率是多少？假设随机遇到两个孩子的可能性相同，且与性别无关.

解:

- 直观来看，结果应为 1/2.
- 令 G_1, G_2, G_3 分别表示年长、年幼、随机的孩子是女孩这三个事件。由对称性可得: $P(G_1) = P(G_2) = P(G_3) = 1/2$
- 根据朴素概率的定义，或者独立性，可得: $P(G_1 \cap G_2) = 1/4$
- 因此，$P(G_1 \cap G_2 | G_3) = P(G_1 \cap G_2 \cap G_3)/P(G_3) = 1/2$
- 又因为 $G_1 \cap G_2 \cap G_3 = G_1 \cap G_2$, 所以概率为 1/2.

注 0.5 假设一个强制性法律规定：如果一个男孩有姐妹则禁止他走出家门。那么这时“随机遇到的孩子是女孩”就等价于“至少有一个孩子是女孩”

男孩女孩概率问题续：冬天出生的女孩

例 0.32 某家庭有两个孩子，给定条件至少一个是女孩且在冬天出生，求两个孩子都是女孩的概率。假设四个季节出生的可能性相同且性别和季节是相互独立的.

解: 由条件概率的定义，可得:

$$P(\text{两个都是女孩} \mid \text{至少有一个是冬天出生的女孩}) = \frac{P(\text{两个都是女孩，至少有一个是冬天出生的女孩})}{P(\text{至少有一个是冬天出生的女孩})}$$

由于指定的孩子是在冬天出生的女孩的概率为 $1/8$, 所以，

$$P(\text{至少有一个是在冬天出生的女孩}) = 1 - (7/8)^2.$$

男孩女孩概率问题续：冬天出生的女孩

利用性别和季节是相互独立的假设，得到:

$$
\begin{aligned}
&P(\text{两个都是女孩，至少有一个是冬天出生的女孩})\\
&= P(\text{两个都是女孩，至少有一个是冬天出生的})\\
&= (1/4)P(\text{至少有一个是冬天出生的女孩})\\
&= (1/4)(1 - P(\text{所有孩子都不是在冬天出生的})
\end{aligned}
$$

合在一起得到，

$$P(\text{两个都是女孩} \mid \text{至少有一个是冬天出生的女孩}) = 7/15$$

例 0.33 n 根绳 $2n$ 个头两两相接，求 $A=\{恰好结成n个圈\}$ 的概率.

- 设想 $2n$ 个头排成一行，规定将第 $2k-1$ 个头与第 $2k$ 个端头相接;
- 令 B_i 表示第 i 根绳的头与尾恰好相接，则 $A=B_1B_2\cdots B_n$;
- 若以 $n(A)$ 表示事件 A 所包含的样本点个数，则易知

$$n(\Omega)=(2n)!, n(B_1)=2n(2n-2)! \Rightarrow P(B_1)=\frac{n(B_1)}{n(\Omega)}=\frac{1}{2n-1};$$

- $P(B_2|B_1)$ 可看作 $n-1$ 根绳某根绳头尾相接的概率，类比 n 根绳情形可得 $P(B_2|B_1)=\dfrac{1}{2(n-1)-1}=\dfrac{1}{2n-3}$;
- 同理可得$P(B_k|B_1B_2\cdots B_{k-1})=\dfrac{1}{2[n-(k-1)]-1}=\dfrac{1}{2n-2k+1}, 3\leq k\leq n$;
- 利用乘法公式可得

$$\begin{aligned}P(A)&=P(B_1B_2\cdots B_n)=P(B_1)P(B_2|B_1)\cdots P(B_n|B_1\cdots B_{n-1})\\&=\frac{1}{(2n-1)!!}\end{aligned}$$

取球问题

例 0.34 在计算机中输入程序，让它自动完成如下操作:

- 在 $1-\frac{1}{2^n}$ 时刻，往盒中放入标号 $10(n-1)+1 \sim 10n$ 的 10 个球，同时取出标号为 $10(n-1)+1$ 的球，$n \geq 1$;
- 在 $1-\frac{1}{2^n}$ 时刻，往盒中放入标号 $10(n-1)+1 \sim 10n$ 的 10 个球，同时取出标号为 n 的球，$n \geq 1$;
- 在 $1-\frac{1}{2^n}$ 时刻，往盒中放入标号 $10(n-1)+1 \sim 10n$ 的 10 个球，同时随机地从盒中取出一个球，$n \geq 1$.

则在时刻 1, 盒中的球数结果如下

- 盒子中有无穷多个球;
- 盒子变为空的;
- 盒子变为空的概率等于 1?

解:

- 记 E={在时刻 1 时盒子变空 };
- $\overline{E}$={在时刻 1 时盒中有球未被取出 };
- 记 A_k={在时刻 1 时 k 号球仍在盒中未被取出 }，则 $\overline{E}=\bigcup\limits_{k=1}^{\infty} A_k$;
- 由概率的次可加性知

$$P(\overline{E})=P\Big(\bigcup_{k=1}^{\infty} A_k\Big)\leq\sum_{k=1}^{\infty} P(A_k);$$

- 为证 $P(\overline{E})=0$，只需证明 $P(A_k)=0,\ k\geq 1$;
- 由于证法类似，仅以证明 $P(A_1)=0$ 为例;

取球问题

- 记 B_n={在 $1-\dfrac{1}{2^n}$ 时刻 1 号球未被取出 }，易知 $A_1=\bigcap\limits_{n=1}^{\infty}B_n$;
- 令 $C_m=\bigcap\limits_{n=1}^{m}B_n$, 则有

$$C_m=\bigcap_{n=1}^{m}B_n\supset\bigcap_{n=1}^{m+1}B_n=C_{m+1}，\text{且}\lim_{m\to\infty}C_m=A_1;$$

- 由概率的上连续性知

$$P(A_1)=P\Big(\lim_{m\to\infty}C_m\Big)=\lim_{m\to\infty}P\big(C_m\big);$$

- 下面求 $P(C_m)$, 即 $P\big(\bigcap_{n=1}^{m}B_n\big)=\prod_{n=1}^{m}\frac{9n}{9n+1}=\prod_{n=1}^{m}\left(1-\frac{1}{9n+1}\right)$;
  - B_1: 在 1/2 时刻，盒中 10 个球，1 号球未被取出，故 $P(B_1)=\frac{9}{10}$;
  - B_2: 在 3/4 时刻，盒中 19 个球，1 号球未被取出，故 $P(B_2|B_1)=\frac{18}{19}$;
  - B_n: $P(B_n|B_1B_2\cdots B_{n-1})=\frac{9n}{9n+1}$.
- $P(A_1)=\prod_{n=1}^{\infty}\big(1-\frac{1}{9n+1}\big)=0$. 同理可证 $P(A_k)=0,\ k=2,3,\cdots$.

罐子模型（波利亚模型）

例 0.35 设罐中有 b 个黑球，r 个红球，每次随机的取出一球，取出后将原球放回，还加进 c 个同色球和 d 个异色球。记 $B_i := \{第i次取出的是黑球\}, R_j := \{第j次取出的是红球\}$. 若连续从罐子中取出三个球，其中有两个红球，一个黑球，则由乘法公式可得

$$
\begin{aligned}
P(B_1R_2R_3) &= P(B_1)P(R_2|B_1)P(R_3|B_1R_2) \\
&= \frac{b}{b+r}\cdot\frac{r+d}{b+r+c+d}\cdot\frac{r+d+c}{b+r+2c+2d} \\
P(R_1B_2R_3) &= P(R_1)P(B_2|R_1)P(R_3|R_1B_2) \\
&= \frac{r}{b+r}\cdot\frac{b+d}{b+r+c+d}\cdot\frac{r+d+c}{b+r+2c+2d} \\
P(R_1R_2B_3) &= P(R_1)P(R_2|R_1)P(B_3|R_1R_2) \\
&= \frac{r}{b+r}\cdot\frac{r+c}{b+r+c+d}\cdot\frac{b+2d}{b+r+2c+2d}
\end{aligned}
$$

显然以上概率与黑球在第几次抽取有关.

罐子模型（波利亚模型）

- 当 $c=-1, d=0$ 时，即为不返回抽样。此时前次抽取结果会影响后次抽取结果，但只要抽取的黑球与红球个数确定，则概率不依赖其抽出球的次序，都是一样的.

$$P(B_1R_2R_3)=P(R_1B_2R_3)=P(R_1R_2B_3)=\frac{br(r-1)}{(b+r)(b+r-1)(b+r-2)}.$$

- 当 $c=0, d=0$ 时，即为返回抽样。此时前次抽取结果不会影响后次抽取结果，故上述三个概率相等且都等于

$$P(B_1R_2R_3)=P(R_1B_2R_3)=P(R_1R_2B_3)=\frac{br^2}{(b+r)^3}.$$

- 当 $c>0, d=0$ 时，称为传染病模型。此时每次取出球后会增加下一次取出同色球的概率，或换言之，每发现一个传染病患者，以后都会增加再传染的概率。与前两种情况一样，三个概率都等于

$$P(B_1R_2R_3)=P(R_1B_2R_3)=P(R_1R_2B_3)=\frac{br(r+c)}{(b+r)(b+r+c)(b+r+2c)}.$$

- 从上面的结果可以看出，只要 $d=0$，以上三个概率都相等，即只要抽取的黑球与红球的个数确定，则概率不依赖于抽出黑红球的次序.
- 当 $c=0, d>0$ 时，称为安全模型。此模型可解释为：每当事故发生了 (当红球被取出)，安全工作就抓紧一些，下次再发生事故的概率就会减少，而当事故没有发生时 (黑球被取出)，安全工作就放松一些，下次再发生事故的概率就会增大，此时，上述三个概率分别为

$$\begin{aligned}
P(B_1R_2R_3) &= \frac{b}{b+r}\cdot\frac{r+d}{b+r+d}\cdot\frac{r+d}{b+r+2d}\\
P(R_1B_2R_3) &= \frac{r}{b+r}\cdot\frac{b+d}{b+r+d}\cdot\frac{r+d}{b+r+2d}\\
P(R_1R_2B_3) &= \frac{r}{b+r}\cdot\frac{r}{b+r+d}\cdot\frac{b+2d}{b+r+2d}
\end{aligned}$$

例 0.36 设罐中有 b 个黑球，r 个红球，每次随机取出一球后将原球放回并加进 c 个同色球，如此反复进行。试证明：在前 $n=n_1+n_2$ 次取球中，取出了 n_1 个红球和 n_2 个黑球的概率为

$$C_n^{n_1}\frac{a(a+c)(a+2c)\cdots(a+n_1c-c)b(b+c)(b+2c)\cdots(b+n_2c-c)}{(a+b)(a+b+c)(a+b+2c)\cdots(a+b+nc-c)}.$$

摸球问题

例 0.37 有三个罐子, 各装有两个球, 分别为两个白球、一白一黑和两个黑球. 任意取出一个罐子, 摸出一球, 发现是白球. (1) 求该罐中另一个球也是白球的概率; (2) 把摸出的球放回罐中, 再从该罐中随机摸出一球, 求该球也是白球的概率.

解:

- $A_k, k = 1, 2$: 表示第 k 次取球取出的是白球的事件;
- $B_k, k = 1, 2, 3$: 表示取出的是装有两白、一白一黑和两黑球的罐子;
- 问题 (1) 要求的是该白球取自两白的罐子的概率, 即 $P(B_1|A_1)$;
- 由条件概率公式, 得 $P(B_1|A_1) = \frac{P(A_1B_1)}{P(A_1)} = \frac{P(B_1)P(A_1|B_1)}{P(A_1)}$;
  - $P(B_1) = 1/3$, $P(A_1|B_1) = 1$;
  - $P(A_1) = \sum_{k=1}^{3} P(B_k)P(A_1|B_k) = \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 0 = \frac{1}{2}$.
- $P(B_1|A_1) = \dfrac{1/3 \cdot 1}{1/2} = 2/3$.

摸球问题

- 问题 (2) 是在同一个罐子两次有放回的取球, 要求的是在第一次取出白球的条件下, 第二次取出的还是白球的条件概率, 即 $P(A_2|A_1)$
- 由条件概率的定义, 知$P(A_2|A_1) = \frac{P(A_1A_2)}{P(A_1)}$;
  - $P(A_1) = \sum_{k=1}^{3} P(B_k)P(A_1|B_k) = \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 0 = \frac{1}{2}.$
  - $P(A_1A_2) = \sum_{k=1}^{3} P(B_k)P(A_1A_2|B_k) = \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot\frac{1}{4} + \frac{1}{3}\cdot 0 = \frac{5}{12}.$
- $P(A_2|A_1) = \frac{P(A_1A_2)}{P(A_1)} = \dfrac{5/12}{1/2} = \dfrac{5}{6}.$

例 0.38 甲盒中有球 5 红 1 黑, 乙盒中有球 5 红 3 黑. 随机取出一个盒子, 从中无放回地相继取出两个球, 试求在第一个球是红球的条件下, 第二个球也是红球的概率.

解:

- B :={第一个球是红球}, C :={第二个球是红球}
- $P(C|B) = \dfrac{P(BC)}{P(B)}$
  - A :={取出的是甲盒}
  - $P(BC) = P(A)P(BC|A) + P(\overline{A})P(BC|\overline{A}) = \frac{1}{2}\cdot\frac{5}{6}\cdot\frac{4}{5} + \frac{1}{2}\cdot\frac{5}{8}\cdot\frac{4}{7} = \frac{43}{84}$
  - $P(B) = P(A)P(B|A) + P(\overline{A})P(B|\overline{A}) = \frac{1}{2}\cdot\frac{5}{6} + \frac{1}{2}\cdot\frac{5}{8} = \frac{35}{48}$
- $P(C|B) = \dfrac{43}{84}\Big/\dfrac{35}{48} = \dfrac{172}{245}$

例 0.39 袋中有 r 个红球与 b 个黑球. 每次从袋中任摸出 1 球并连同 s 个同色球一起放回. 以 R_n 表示第 n 次摸出红球, 试证 $P(R_n)=\dfrac{r}{r+b}$.

证明: 利用归纳法来证明:$n=1$ 时, $P(R_1)=\dfrac{r}{r+b}$ 显然成立.

假设 $n-1$ 时命题成立. 为求 $P(R_n)$, 我们以第 1 次取球的可能结果 R_1 与 $\bar{R}_1$ 作为 Ω 的分割, 用全概率公式可得:

$$
\begin{aligned}
P(R_n) &= P(R_1)P(R_n|R_1)+P(\bar{R}_1)P(R_n|\bar{R}_1) \\
&= \frac{r}{r+b}\cdot\frac{r+s}{r+s+b}+\frac{b}{r+b}\cdot\frac{r}{r+b+s} \\
&= \frac{r}{r+b}
\end{aligned}
$$

注 0.6 当 $s=0$ 时相当于放回摸球, 而 $s=-1$ 相当于不放回摸球.

掷硬币问题

例 0.40 甲、乙二人抛掷一枚均匀的硬币，甲抛了 100 次，乙抛了 101 次．求事件 $E:=\{$乙抛出的正面次数比甲多$\}$ 的概率．

解:

- 如果甲和乙都抛掷这枚均匀的硬币 100 次，那么当然会有三种不同的可能结果
  - A_0:甲乙抛出的正面次数一样多
  - A_1:甲抛出的正面次数比乙多
  - A_2:乙抛出的正面次数比甲多
  - $P(A_0)+P(A_1)+P(A_2)=1$且 $P(A_1)=P(A_2)$
- 以乙第一次抛出的硬币结果对样本空间进行分割:
  - 若 $B=\{$乙第一次时抛出的是正面$\}$ 发生，则乙只要在接下来的 100 次抛掷中，抛出的正面次数不比甲少即可，即

$$P(E|B)=P(A_0)+P(A_1)$$

  - 若 $\overline{B}:=\{$乙第一次时抛出的是反面$\}$ 发生，则乙只要在接下来的 100 次抛掷中，抛出的正面次数比甲多即可，即 $P(E|\overline{B})=P(A_1)=P(A_2)$
- $P(E)=P(B)P(E|B)+P(\overline{B})P(E|\overline{B})=\frac{1}{2}\left(P(A_0)+P(A_1)+P(A_2)\right)=\frac{1}{2}$.

例 0.41 甲、乙进行某项对战比赛, 每回合胜者得 1 分, 败者不得分. 比赛进行到有 1 人比另外 1 人多 2 分终止, 多 2 分者获胜. 现知每回合甲胜的概率为 $p \in (0,1)$. 试求 $A :=\{$甲最终获胜$\}$ 的概率.

解: 法一 (经典解法)

- 显然甲只能在偶数个回合后获胜
- 记 $A_{2n}=\{$甲在 $2n$ 个回合后获胜$\}$, 则 $A = \cup_{n=1}^{\infty} A_{2n}$
- $A_{2n}, n \geq 1$ 两两互不相容且$P(A_{2n}) = (2p(1-p))^{n-1}p^2$
- 甲最终获胜的概率

$$
\begin{aligned}
P(A) &= \sum_{n=1}^{\infty} P(A_{2n}) = p^2 \sum_{n=1}^{\infty} (2p(1-p))^{n-1} \\
&= p^2 \sum_{n=0}^{\infty} (2p(1-p))^{n} = \frac{p^2}{1-2p(1-p)}.
\end{aligned}
$$

对战获胜概率问题

解: 法二 (全概率公式)

- 以前两个回合的战绩对样本空间进行分割
- 分别以 B_1, B_2, B_3 表示甲二胜、一胜一败、二败事件
- $P(B_1) = p^2$, $P(B_2) = 2p(1-p)$, $P(B_3) = (1-p)^2$
- $P(A|B_1) = 1$, $P(A|B_2) = P(A)$, $P(A|B_3) = 0$
- 由全概率公式得 $P(A) = \sum_{i=1}^{3} P(B_i)P(A|B_i) = p^2 + 2p(1-p)P(A)$
- $P(A) = \dfrac{p^2}{1-2p(1-p)}$

对战获胜概率问题

解: 法三 (随机游动)

- 考察质点在数轴整数点上的随机游动: 在整数点 $x=n$
  - 以概率 p 向右移动到整数点 $x=n+1$,
  - 以概率 $1-p$ 向左移动到整数点 $x=n-1$
- 以 p_n 表示 "质点由 $x=n$ 出发, 未达 -2 前先到达 2 的概率"
- 显然, $p_{-2}=0, p_2=1$
- 甲获胜的概率即为: p_0
- 由全概率公式易得

$$\begin{cases} p_0 = pp_1 + (1-p)p_{-1}, \\ p_1 = pp_2 + (1-p)p_0 = p + (1-p)p_0, \\ p_{-1} = pp_0 + (1-p)p_{-2} = pp_0. \end{cases}$$

- $p_0 = p(p+(1-p)p_0) + (1-p)(pp_0) = p^2 + 2p(1-p)p_0$
- 故

$$p_0 = \frac{p^2}{1-2p(1-p)}.$$

例 0.42 (随机抛硬币) 假设有一枚均匀的硬币和一枚以概率 $3/4$ 正面朝上的不均匀硬币. 随机选取一枚硬币掷 3 次, 3 次都是正面朝上. 试问:

1. 给定上述信息后, 选取的硬币是均匀硬币的概率有多大?
2. 接下去掷第四次时, 仍是正面朝上的概率是多少?

解: 令

$$A := \{\text{选取的硬币掷 3 次均正面朝上}\}$$
$$F := \{\text{选取的硬币是均匀的}\}, \quad H := \{\text{第 4 次正面朝上}\}$$

则

$$\begin{aligned}
P(F|A) &= \frac{P(A|F)P(F)}{P(A)} = \frac{P(A|F)P(F)}{P(A|F)P(F) + P(A|F^c)P(F^c)} \\
&= \frac{(1/2)^3 \cdot 1/2}{(1/2)^3 \cdot 1/2 + (3/4)^3 \cdot 1/2} \approx 0.23 \\
P(H \mid A) &= P(H \mid F, A)\, P(F \mid A) + P(H \mid F^c, A)\, P(F^c \mid A) \\
&= \frac{1}{2} \cdot 0.23 + \frac{3}{4} \cdot (1 - 0.23) \approx 0.69.
\end{aligned}$$

例 0.43 考虑以验血结果诊断某种罕见病的患病概率:

- 某罕见病的发病率为 1%
- 通过验血诊断该病的误诊率为 5%, 即非患者中有 5% 的人验血结果为阳性, 患者中有 5% 的人验血结果为阴性
- 现已知某人验血结果为阳性, 试求他患有此病的概率

解:

- 记 D :={患有此病}, T_1 :={第一次验血结果为阳性}.
- 要求的概率是: $P(D|T_1) = \dfrac{P(DT_1)}{P(T_1)} = \dfrac{P(D)P(T_1|D)}{P(T_1)}$
- 由题意可知

$$P(D) = 1\%,\ P(\overline{D}) = 99\%,\ P(T_1|D) = 95\%,\ P(T_1|\overline{D}) = 5\%$$

$$\begin{aligned} P(T_1) &= P(D)P(T_1|D) + P(\overline{D})P(T_1|\overline{D}) = 1\% \cdot 95\% + 99\% \cdot 5\% \\ &= 0.0685 \end{aligned}$$

罕见病检测诊断问题

- $P(D|T_1) = \dfrac{P(D)P(T_1|D)}{P(T_1)} = \dfrac{1\% \cdot 95\%}{0.0685} \approx 0.16.$
- 几率方法:

$$\frac{P(D|T_1)}{P(D^c|T_1)} = \frac{P(D)}{P(D^c)} \frac{P(T_1|D)}{P(T_1|D^c)} = \frac{1}{99} \cdot \frac{0.95}{0.05} \approx 0.19$$

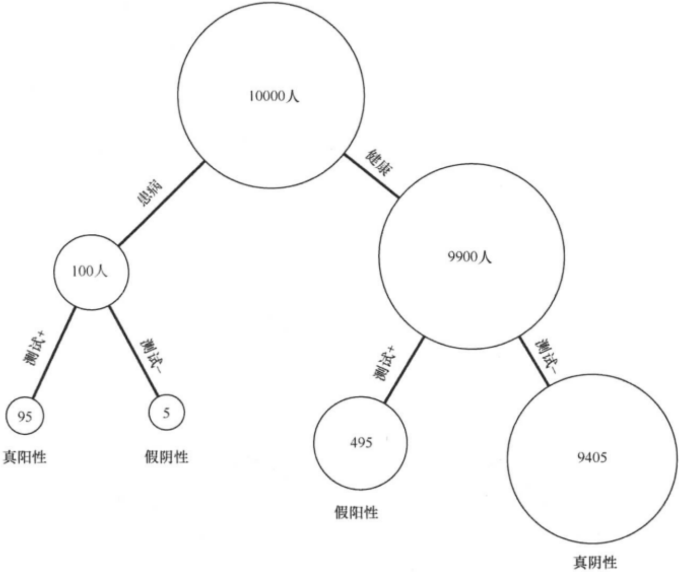
由几率与概率之间的关系可知

$$P(D|T_1) = 0.19/(1+0.19) \approx 0.16,$$

与上述结果一致.

- 用几率形式的贝叶斯准则计算更迅速的原因是此时不需要计算普通贝叶斯准则的分母.

罕见病检测诊断问题图示



罕见病检测诊断问题

例 0.44 接上例，检测结果为阳性的某人，决定进行第二次检测．假设新的检测结果与之前的结果相互独立，且有相同的敏感性和特异性．若第二次检测结果也为阳性，试求此人患有此病的概率．

解:

- 记 T_2 :={第二次验血结果为阳性}
- 要求的概率是: $P(D|T_1T_2)$
- 一步法: 将两个检测结果一次性都考虑在内以进行概率更新,
  - 计算几率

$$\frac{P(D|T_1\cap T_2)}{P(D^c|T_1\cap T_2)}=\frac{P(D)}{P(D^c)}\frac{P(T_1\cap T_2|D)}{P(T_1\cap T_2|D^c)}=\frac{1}{99}\cdot\frac{0.95^2}{0.05^2}=\frac{361}{99}\approx 3.646$$

  - $P(D|T_1T_2)=\frac{3.646}{1+3.646}\approx 0.78.$

罕见病检测诊断问题

- 两步法:
  - 在完成第一次检测后，某人患有此病的后验几率为

$$\frac{P(D|T_1)}{P(D^c|T_1)} = \frac{1}{99} \cdot \frac{0.95}{0.05} \approx 0.19$$

  - 将后验几率作为新的先验几率，然后基于第二次检测结果更新后验几率

$$\begin{aligned}\frac{P(D|T_1 \cap T_2)}{P(D^c|T_1 \cap T_2)} &= \frac{P(D|T_1)}{P(D^c|T_1)} \frac{P(T_2|D, T_1)}{P(T_2|D^c, T_1)} \\ &= (\frac{1}{99} \cdot \frac{0.95}{0.05}) \frac{0.95}{0.05} \\ &= \frac{361}{99} \approx 3.646\end{aligned}$$

  - $P(D|T_1 T_2) = \dfrac{3.646}{1 + 3.646} \approx 0.78$

电邮问题

例 0.45 甲、乙二人之间经常用 E-mail 相互联系，他们约定在收到对方信件的当天即给 E-mail 回复. 由于线路问题，每 n 份 E-mail 中会有 1 份不能在当天送达收件人. 甲在某日发了 1 份 E-mail 给乙，但未在当天收到乙的回音. 试求乙在当天收到了甲发给他的 E-mail 的概率.

解:

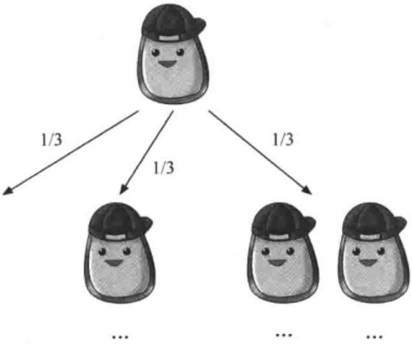
- 在这个问题中，包含了两个不确定的环节:
  - 甲发给乙的 E-mail 不一定在当天到达乙处
  - 乙回给甲的 E-mail 不一定在当天到达甲处
- A={乙在当天收到甲的 E-mail}, B={甲在当天收到乙回的 E-mail}
- $P(A|\overline{B}) = \dfrac{P(A)P(\overline{B}|A)}{P(A)P(\overline{B}|A) + P(\overline{A})P(\overline{B}|\overline{A})}$
- 由题中条件知

$$P(A) = \frac{n-1}{n},\ P(\overline{A}) = \frac{1}{n},\ P(\overline{B}|A) = \frac{1}{n},\ P(\overline{B}|\overline{A}) = 1.$$

- $P(A|\overline{B}) = \dfrac{\frac{n-1}{n}\cdot\frac{1}{n}}{\frac{n-1}{n}\cdot\frac{1}{n} + \frac{1}{n}\cdot 1} = \dfrac{n-1}{2n-1} < 1/2$

例 0.46 (分支过程) 池塘里只有一只变形虫叫作 Bobo.

- 每过 1 分钟, Bobo 有三种结果：死去、分裂成两个或保持原状
- 三种结果出现的概率相同
- 此后所有活着的 Bobo 都将继续以这种方式相互独立地进行下去
- 那么这个变形虫种族最终灭亡的概率是多少?



分支过程 (续)

解:

- 令 $D := \{$最终种族灭绝$\}$, 本题希望求出 $P(D)$.
- 我们在第一步结果的基础上即以 1 分钟后的结果进行分析:
  - 令 $B_i := \{$1分钟后Bobo变成的变形虫个数$\}(i = 0, 1, 2)$
  - 易知 $P(D|B_0) = 1$ 和 $P(D|B_1) = P(D)$, $P(D|B_2) = P(D)^2$.
- 利用全概率公式有

$$
\begin{aligned}
P(D) &= P(D|B_0) \cdot \frac{1}{3} + P(D|B_1) \cdot \frac{1}{3} + P(D|B_2) \cdot \frac{1}{3} \\
&= 1 \cdot \frac{1}{3} + P(D) \cdot \frac{1}{3} + P(D)^2 \cdot \frac{1}{3}.
\end{aligned}
$$

- 由上式可解得 $P(D) = 1$, 即变形虫种族最终会以概率 1 灭绝.

对战相遇概率问题

例 0.47 包括甲、乙二人在内的 2^n 名乒乓球运动员参加一场淘汰赛.

- 第一轮任意两两配对比赛, 然后 2^{n-1} 名胜者再任意两两配对进行第二轮比赛, 如此下去, 直至第 n 轮决出一名冠军为止
- 假定每一名运动员在各轮比赛中胜负都是等可能的
- 求 $B :=${甲、乙二人在淘汰赛中相遇} 的概率

对战相遇概率问题

解: 记 $p_n := P(B)$, 即甲、乙二人在 2^n 人参赛的比赛中相遇的概率

- 以甲、乙二人是否在第一轮比赛相遇对样本空间进行分割
- $A :=$ {甲、乙二人在第一轮比赛中相遇}
- 由全概率公式

$$p_{n+1} = P(A)P(B|A) + P(\overline{A})P(B|\overline{A}) = P(A) + (1 - P(A)) \cdot \frac{1}{4}p_n$$

- $P(A)$: 甲、乙二人在第一轮比赛中相遇的概率
  - 采用无编号分组模式考虑
  - 2^{n+1} 个人两两配对的方式一共有 $\dfrac{(2^{n+1})!}{2^{2^n}(2^n)!}$ 种
  - 甲、乙二人配为一对的配对方式有 $\dfrac{(2^{n+1}-2)!}{2^{2^n-1}(2^n-1)!}$ 种
  - 将上述两式相除, 即得 $P(A) = \dfrac{1}{2^{n+1}-1}$
- $p_{n+1} = \frac{1}{2^{n+1}-1} + \frac{1}{4}(1 - \frac{1}{2^{n+1}-1})p_n$
- $p_{n+1} = \frac{1}{2^n}$

例 0.48 甲乙轮流掷一均匀骰子. 甲先掷, 以后每当某人掷出 1 点后则交给对方掷, 否则此人继续掷. 试求事件 $A_n = \{第n次由甲掷\}$ 的概率.

解: 记 $p_n = P(A_n)$, 则以 A_{n-} 与 $\bar{A}_{n-1}$ 为分割用全概率公式可得:

$$\begin{aligned} p_n = P(A_n) &= P(A_{n-1})P(A_n|A_{n-1}) + P(\bar{A}_{n-1})P(A_n|\bar{A}_{n-1}) \\ &= p_{n-1}\frac{5}{6} + (1-p_{n-1})\frac{1}{6} = \frac{2}{3}p_{n-1} + \frac{1}{6} \end{aligned}$$

经过整理, 可将上式化为以下递推的形式

$$p_n - \frac{1}{2} = \frac{2}{3}\left(p_{n-1} - \frac{1}{2}\right), n = 2, 3, \cdots .$$

由 $p_1 = 1$ 可得 $p_n - \frac{1}{2} = \left(\frac{2}{3}\right)^{n-1}\left(p_1 - \frac{1}{2}\right) = \frac{1}{2}\left(\frac{2}{3}\right)^{n-1}$. 因此, 我们有

$$p_n = P(A_n) = \frac{1}{2}\left[1 + \left(\frac{2}{3}\right)^{n-1}\right], n = 1, 2, \cdots ,$$

结绳问题续

例 0.49 n 根绳 $2n$ 个头两两相接, 求 $A_n = \{恰好结成n个圈\}$ 的概率.

解: 此前曾用条件概率解过本题, 现利用全概率公式给出一个解答.

- 将 n 根短绳作编号并记 $p_n = P(A_n)$
- 记 B=\{1 号绳连成 1 个圈\} 并用 B 和 $\overline{B}$ 作为对 Ω 的分划
- 由全概率公式可知

$$p_n = P(A_n) = P(B)P(A_n|B) + P(\overline{B})P(A_n|\overline{B})$$

- $P(B) = \dfrac{1}{2n-1}, P(A_n|\overline{B}) = 0, P(A_n|B) = P(A_{n-1}) = p_{n-1}$
- $p_n = P(A_n) = \frac{1}{2n-1}p_{n-1},\ n = 2, 3, \cdots.$
- 反复利用上式, 并由 $p_1 = 1$ 可得

$$p_n = \frac{1}{(2n-1)!!},\ n = 1, 2, \cdots.$$

秘书问题

例 0.50 某公司需招收秘书一名，共有 n 个人报名应聘，公司面试规则与录取策略如下：

- 面试规则：逐个面试，并在面试当时对应聘者表态是否录用，一旦对应聘者表态不录用，不可改变决定
- 录用策略：
  - 不录用前 $k(1 \leq k < n)$ 个面试者
  - 自第 $k+1$ 个开始，只要发现某人比他前面的所有面试者都好，就录用他，否则就录用最后一个
- 试对该公司的策略作概率分析

秘书问题

- A :={最佳人选被录用}
- B_j :={最佳人选在面试顺序中排在第 j 位}
- 全概率公式: $P(A)=\sum_{j=1}^{n}P(A|B_j)P(B_j)$
  - $P(B_j)=\frac{(n-1)!}{n!}=\frac{1}{n}, j=1,\cdots,n$
  - 当 $1\le j\le k$ 时, $P(A|B_j)=0$ (最佳人选位于前 k 个面试者, 不会被录用)
  - 当 $k+1\le j\le n$ 时, $P(A|B_j)=\dfrac{k}{j-1}$ (最佳人选被录用当且仅当前 $j-1$ 个面试者中的最佳者在前 k 个人中)
- $P(A)=\dfrac{1}{n}\sum_{j=k+1}^{n}\dfrac{k}{j-1}=\dfrac{k}{n}\sum_{j=k}^{n-1}\dfrac{1}{j}\sim\dfrac{k}{n}\ln\dfrac{n}{k}$
- 令 $g(x)=\dfrac{x}{n}\ln\dfrac{n}{x}$, $x>0$, 则 $g'(x)=\dfrac{1}{n}\ln\dfrac{n}{x}-\dfrac{1}{n}=0\Longleftrightarrow x=\dfrac{n}{e}$
- 若要 $P(A)$ 达到最大, 只需 k 取最靠近 $\frac{n}{e}$ 的正整数
- 最大概率值为 $g\left(\frac{n}{e}\right)=\frac{1}{e}\approx 0.36788$ 与 n 无关

第八讲：事件的独立性

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

两个事件的独立性

- 直观上来说，两个事件的独立性是指：一个事件的发生不影响另一个事件的发生。比如在掷两颗骰子的实验中，第一颗骰子的点数和第二颗骰子的点数是互不影响的.
- 从概率的角度看，$P(A|B)$ 与 $P(A)$ 的差别在于：事件 B 的发生改变了事件 A 发生的概率，也即事件 B 对事件 A 有某种影响。故如果 A 与 B 的发生是相互不影响的，则有

$$P(A|B) = P(A), \quad P(B|A) = P(B).$$

上面两式均等价于

$$P(AB) = P(A)P(B) \tag{14}$$

- 注意到 (14) 式对 $P(B) = 0$ 或 $P(A) = 0$ 仍然成立，为此，我们用 (14) 作为两个事件相互独立的定义.

两个事件独立性的定义

定义 0.8 如果对事件 A 与 B 有

$$P(AB) = P(A)P(B)$$

成立，则称事件 A 与 B 相互独立，简称 A 与 B 独立. 否则称 A 与 B 不独立或相依.

注 0.7

- 零概率事件 E 与任何事件相互独立，特别的，不可能事件与任何事件相互独立
- 若 A, B 互不相容且独立，则 A, B 至少有一个零概率事件
- 非零概率不相容事件，一定不独立；非零概率独立事件，一定相容
- 若事件 A 与其自身相互独立，则 $P(A) = 0$，或 $P(A) = 1$

如何确定事件的独立性:

- 实际问题中，两个事件的独立大多根据经验及相互有无影响的直观性来判断.
- 但对于较复杂事件，有无相互影响并不是很直观，则需要验证 (14) 式是否成立来说明独立性.

对立事件的独立性

定理 0.15 若 A 与 B 独立，则 A 与 $\overline{B}$ 独立，$\overline{A}$ 与 B 独立，$\overline{A}$ 与 $\overline{B}$ 独立.

证明: 我们仅证 $P(A\overline{B}) = P(A)P(\overline{B})$, 其余类似可证.

$$\begin{aligned} P(A\overline{B}) &= P(A-B) = P(A) - P(AB) = P(A) - P(A)P(B) \\ &= P(A)(1-P(B)) = P(A)P(\overline{B}) \end{aligned}$$

对于上面的定理直观上来理解也是很容易的：因 A, B 独立，故 A 的发生不影响 B 的发生，从而也不会影响 B 的不发生，$\cdots$

例 0.51 考虑掷硬币问题, 记正面向上对应的样本点为"H", 反面向上为"T", 那么连续掷三次的结果构成样本空间

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},$$

令事件 A 为" 最后一次是反面", B 为" 三次结果相同", 则有

$$A = \{HHT, HTT, THT, TTT\}, B = \{HHH, TTT\}$$

试讨论 A, B 的独立性.

解: ：设每次抛掷反面向上的概率是 p, 那么

$$P(A) = p^3 + 2p^2(1-p) + p(1-p)^2, P(B) = p^3 + (1-p)^3, P(AB) = p^3,$$

不难验证, $p = 0$、$p = 1$ 和 $p = \frac{1}{2}$ 时,A 和 B 是独立的, 否则两者不独立.

定义 0.9 设 A, B, C 三个事件，如果有

$$\left.\begin{aligned} P(AB) &= P(A)P(B) \\ P(AC) &= P(A)P(C) \\ P(BC) &= P(B)P(C) \end{aligned}\right\} \tag{15}$$

$$P(ABC) = P(A)P(B)P(C) \tag{16}$$

则称 A, B, C 相互独立。如果仅有 (15) 式成立，则称 A, B, C 两两独立.

两两独立与相互独立的关系

- 由定义可知，三个事件相互独立必能推出两两独立.
- 但两两独立未必能推出相互独立，即 (15) 式成立，不一定能推出 (16) 成立
  - 考虑独立投掷两枚均匀硬币的随机试验，设事件 A 代表第一枚硬币正面朝上，事件 B 代表第二枚硬币正面朝上，事件 C 表示两枚硬币结果相同。易知: A B 和 C 是两两独立，但

$$P(A \cap B \cap C) = 1/4 \neq 1/8 = P(A)P(B)P(C).$$

  - 考虑一个均匀的正四面体，第一二三面分别染上红 / 白 / 黑色，第四面同时染上红白黑色。现在以 A, B, C 分别记投一次四面体出现红，白，黑色朝下的事件。则易有

$$\begin{aligned} P(A) = P(B) = P(C) &= 1/2 \\ P(AB) = P(BC) = P(AC) &= 1/4 \\ P(ABC) &= 1/4 \end{aligned}$$

- 反之，如果 (16) 成立，是否能推出 (15) 成立？
  - 考虑一个均匀的正八面体，第 1, 2, 3, 4 面染上红色，第 1, 2, 3, 5 面染上白色，第 1, 6, 7, 8 面染上黑色。现在以 A, B, C 分别记投一次八面体出现红，白，黑色朝下的事件，则

$$
\begin{aligned}
P(A) = P(B) = P(C) = 4/8 = 1/2 \\
P(ABC) = 1/8 \\
P(AB) = 3/8 \neq 1/4 = P(A)P(B)
\end{aligned}
$$

三个以上事件的独立性

定义 0.10 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间，$A_1, A_2, \cdots, A_n \in \mathcal{F}$，对任意的 $1 \le k \le n$ 及任意的 $1 < j_1 < j_2 < \cdots < j_k \le n$ 均有：

$$P(A_{j_1}A_{j_2}\cdots A_{j_k}) = P(A_{j_1})P(A_{j_2})\cdots P(A_{j_k}) \tag{17}$$

成立，则称事件 $A_1, \cdots, A_n$ 相互独立.

- (17) 式共有多少个等式？

$$\left.\begin{array}{l} P(A_{j_1}A_{j_2}) = P(A_{j_1})P(A_{j_2}) \\ P(A_{j_1}A_{j_2}A_{j_3}) = P(A_{j_1})P(A_{j_2})P(A_{j_3}) \\ \qquad \vdots \\ P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n) \end{array}\right\} C_n^2 + \cdots + C_n^n = 2^n - n - 1$$

- 从定义可以看出，n 个相互独立事件中的任取 $m(2 \le m \le n)$ 个事件仍是相互独立的，而且任意一部分与另一部分也是独立的.
- 类似于前面的证明，将相互独立事件中的任一部分换为对立事件，所得诸事件仍是相互独立的.

定义 0.11 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间，每个 $t \in T$ 有 $A_t \in \mathcal{F}$. 称 $\{A_t, t \in T\}$ 为独立事件族，如果对 T 的任意有限子集 $\{t_1, t_2, \cdots, t_s\}$, 事件 $A_{t_1}, A_{t_2}, \cdots, A_{t_s}$ 相互独立.

例 0.52 $\mathcal{F}$ 中事件序列 $\{A_n\}$ 为相互独立的充分必要条件是，任意 $n \geq 1$, 事件 $A_1, A_2, \cdots, A_n$ 独立；等价的，任意有限个自然数 $k_1, \cdots, k_s$ 有

$$P(A_{k_1} A_{k_2} \cdots A_{k_s}) = P(A_{k_1}) P(A_{k_2}) \cdots P(A_{k_s})$$

条件独立

定义 0.12 称事件 A 和 B 是关于 E 条件独立的，如果

$$P(A \cap B|E) = P(A|E)P(B|E)$$

- 两个事件可以在给定事件 E 的条件下是条件独立的，但它们不是独立的.
- 两个事件可以是独立但却不是关于 E 条件独立的.
- 两个事件可以关于 E 条件独立但关于 $\overline{E}$ 不存在条件独立.

条件独立不意味着独立

例 0.53 假设有两枚硬币，一枚是均匀的，一枚是不均匀的。从两枚硬币中随机的选一枚硬币并进行抛掷 2 次，若令

$$
\begin{aligned}
F &:= \{\text{选取的硬币是均匀的}\} \\
A_1 &:= \{\text{第一次投掷硬币正面朝上}\} \\
A_2 &:= \{\text{第二次投掷硬币正面朝上}\}
\end{aligned}
$$

则给定 F 为条件，A_1 和 A_2, 是相互独立的，A_1 和 A_2 并不是无条件独立的，因为 A_1 会提供关于 A_2 的信息.

独立不意味着条件独立

例 0.54 假设只有我的朋友 Alice 和 Bob 给我打过电话。每天他俩都会相互独立地决定是否给我打电话。若令

$$A := \{\text{Alice 给我打电话}\}$$
$$B := \{\text{Bob 给我打电话}\}$$
$$R := \{\text{听到电话铃响}\}$$

- 显然，A 和 B 是无条件独立的.
- 但现在我听到一声电话铃响，那 A 和 B 就不再独立了：如果这个电话不是 Alice 打的，那就肯定是 Bob 打的。从而

$$P(B|R) < 1 = P(B|\overline{A}R) = \frac{P(B\overline{A}R)}{P(\overline{A}R)} = \frac{P(B\overline{A}|R)}{P(\overline{A}|R)}.$$

  显然: $P(B\overline{A}|R) > P(B|R)P(\overline{A}|R)$
- B 与 $\overline{A}$ 关于 R 不条件独立，A, B 亦是如此.

例 0.55 假设有两种课程：好的课程和坏的课程。在好的课上，如果你努力，就很有可能得到 A. 在坏的课上，教授随机分配给学生分数，而不管他们是否努力。若令

$$
\begin{aligned}
G &:= \{\text{这个课程是好的}\} \\
W &:= \{\text{你学习努力}\} \\
A &:= \{\text{你的得分为}A\}
\end{aligned}
$$

这时，给定 $\overline{G}$, A 和 W 是条件独立的，但给定 G, A 和 W 却不是独立的！

定义 0.13 (集类 (族) 的独立性) 考虑样本空间 $\Omega, \mathcal{A}_k \subset \Omega$, $k = 1, \cdots, n$. 称集类 (族)$\{\mathcal{A}_k\}_{k=1}^n$ 是相互独立的, 如果 $\{\mathcal{A}_k\}_{k=1}^n$ 满足

$$P\left(\bigcap_{k\in I} A_k\right) = \prod_{k\in I} P(A_k), \forall A_k \in \mathcal{A}_k, \forall I \subset \{1, 2, \cdots, n\}.$$

定理 0.16 (σ-代数的独立性) 设 $\mathcal{F}_1, \cdots, \mathcal{F}_n$ 为 Ω 上的 σ-代数, 若

$$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2) \cdots P(A_n), \quad \forall A_k \in \mathcal{F}_k, k = 1, 2, \cdots, n$$

则 $\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_n$ 是独立的.

例 0.56 (硬币实验的独立性) 抛掷不均匀硬币的实验, 正面 (用 1 表示) 向上的概率是 p, 反面 (用 0 表示) 向上的概率是 q. 假设连抛 n 次, 则样本空间 Ω 为 $\Omega = \{a_1 a_2 \cdots a_n : a_k = 0, 1\}$. 考虑事件 $A_k = \{a_k = 1\}$, $k = 1, \cdots, n$, 构造 σ-代数 $\mathcal{F}_k$: $\mathcal{F}_k = \left\{\Omega, \varnothing, A_k, A_k^C\right\}$.可以验证, 这些 σ-代数是独立的.

独立的集类 (族) 生成的 σ-代数未必独立

- 考虑 $\Omega=\{1,2,3,4\}$, 集类 (族)$\mathcal{A}=\{\{1,2\},\{2,3\}\}$,$\mathcal{B}=\{\{2,4\}\}$
- 令 $P(\{1\})=P(\{2\})=P(\{3\})=P(\{4\})=\dfrac{1}{4}$, 则

$$P(\{1,2\}\cap\{2,4\})=P(\{2\})=\frac{1}{4}=\frac{1}{2}\times\frac{1}{2}=P(\{1,2\})P(\{2,4\}),$$

$$P(\{2,3\}\cap\{2,4\})=P(\{2\})=\frac{1}{4}=\frac{1}{2}\times\frac{1}{2}=P(\{2,3\})P(\{2,4\})$$

- 集类 (族)$\mathcal{A}$ 和 $\mathcal{B}$ 独立
- 但由于

$$\sigma(\mathcal{A})=2^{\Omega},\sigma(\mathcal{B})=\{\{2,4\},\{1,3\},\varnothing,\Omega\}$$

  且明显有

$$P(\{2,4\}\cap\{3\})=0\neq\frac{1}{8}=\frac{1}{2}\times\frac{1}{4}=P(\{2,4\})P(\{3\})$$

- 故, $\sigma(\mathcal{A})$ 和 $\sigma(\mathcal{B})$ 不独立.

定理 0.17 如果 $\mathcal{A}$ 和 $\mathcal{B}$ 是独立的集类 (族), $\mathcal{B}$ 是 π-类, 那么 $\mathcal{A}$ 和 $\sigma(\mathcal{B})$ 也独立.

证明: 应用单调类定理.

- 任意固定 $A \in \mathcal{A}$, 令

$$\mathscr{A} = \{B \in \sigma(\mathcal{B}) : P(AB) = P(A)P(B)\}$$

则 $\mathscr{A}$ 是 λ-类 (系统)

- 由 $\mathcal{B} \subset \mathscr{A}$ 及单调类定理可得: $\sigma(\mathcal{B}) \subset \mathscr{A}$.
- $\mathcal{A}$ 和 $\sigma(\mathcal{B})$ 独立.

定理 0.18 对于事件列 $\{A_j\}$，有

- 如果 $\sum_{j=1}^{\infty} P(A_j) < \infty$，则 $P(A_n\ i.o.) = 0$
- 如果 $\{A_j\}$ 相互独立，$\sum_{j=1}^{\infty} P(A_j) = \infty$，则 $P(A_n\ i.o.) = 1$

证明: 注意到

$$P(A_n\ i.o.) = P(\lim \cup_{j=n}^{\infty} A_j) = \lim P(\cup_{j=n}^{\infty} A_j) \le \lim \sum_{j=n}^{\infty} P(A_j) = 0$$

由于

$$\begin{aligned} P(\cup_{j=n}^{\infty} A_j) &= \lim_{m\to\infty} P(\cup_{j=n}^{m} A_j) = \lim_{m\to\infty} (1 - P(\cap_{j=n}^{m} \overline{A}_j)) \\ P(\cap_{j=n}^{m} \overline{A}_j) &= \Pi_{j=n}^{m} P(\overline{A}_j) = \Pi_{j=n}^{m} (1 - P(A_j)) \\ &\le \Pi_{j=n}^{m} \exp(-P(A_j)) = \exp(-\sum_{j=n}^{m} P(A_j)) \xrightarrow{m\to\infty} 0 \end{aligned}$$

随机试验的独立性

- 先考虑两个随机试验，假定 $(\Omega_i, \mathcal{F}_i, P_i), i = 1, 2$ 为第 i 个随机试验对应的概率空间。按照之前独立性的理解，两个试验的独立性应当叙述为:

  对任何的$A_i \in \mathcal{F}_i, i = 1, 2, A_1$与$A_2$同时发生的概率等于它们各自概率之乘积

- 两个不妥:
  - “A_1 与 A_2 同时发生” 应当是这两个事件的交，但它们分别是两个样本空间 Ω_1, Ω_2 的子集，无法进行运算;
  - 两个概率空间有各自的概率 P_1, P_2, 但涉及两个度验，命题中 “同时发生的概率” 既不能用 P_1 也不能用 P_2 来度量.
- 解决方法：构造可以同时描述两个试验的新概率空间 $(\Omega, \mathcal{F}, P)$.

乘积空间的构造

- 样本乘积空间: $\Omega := \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1 \text{且} \omega_2 \in \Omega_2\}$;
- 乘积 σ- 代数 $\mathcal{F}_1 \times \mathcal{F}_2$:
  - 可测矩形集类: $\mathcal{G} := \{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$;
  - $\mathcal{F}_1 \times \mathcal{F}_2 := \sigma(\mathcal{G})$
- 乘积概率测度:
  - 对于每个可测矩形 $A_1 \times A_2 \in \mathcal{G}$ 定义如下集函数:

$$P(A_1 \times A_2) = P_1(A_1)P_2(A_2), \quad A_i \in \mathcal{F}_i, i = 1, 2. \tag{19}$$

  - 理论上可以证明如上定义在 $\mathcal{G}$ 上的集函数 P 可唯一地扩张为 $\mathcal{F}_1 \times \mathcal{F}_2$ 上的概率测度，称之为 P_1 与 P_2 的乘积 (概率) 测度.
- 在上述乘积测度下

$$P(A_1 \times \Omega_2) = P_1(A_1), \quad P(\Omega_1 \times A_2) = P_2(A_2)$$

$$P\big((A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)\big) = P(A_1 \times A_2)$$

$$= P_1(A_1)P_2(A_2) = P(A_1 \times \Omega_1)P(\Omega_1 \times A_2)$$

- $(\Omega_i, \mathcal{F}_i, P_i)$ 的独立性取决于乘积样本空间 $\Omega_1 \times \Omega_2$ 上的概率是否取作由 (19) 所确定的乘积测度

定义 0.14 设有 n 个随机试验，第 i 个试验的概率空间为 $(\Omega_i, \mathcal{F}_i, P_i)$, $i = 1, \cdots, n$. 代表这 n 个试验的乘积样本空间 $\Omega = \Omega_1 \times \cdots \times \Omega_n$, $\mathcal{F} = \mathcal{F}_1 \times \cdots \times \mathcal{F}_n = \sigma(\mathcal{G})$, 其中 $\mathcal{G}$ 为形如 $B_1 \times \cdots \times B_n (B_i \in \mathcal{F}_i)$ 的可测矩形的全体。如果 $(\Omega, \mathcal{F})$ 上的概率测度 P 是 $P_1, \cdots, P_n$ 的乘积测度，即对任何 $B_1 \times \cdots \times B_n \in \mathcal{G}$ 满足

$$P(B_1 \times \cdots \times B_n) = P_1(B_1) \cdots P(B_n),$$

则称这 n 个度验独立. 如果现设

$$\Omega_i \equiv \Omega_0, \mathcal{F}_i \equiv \mathcal{F}_0, P_i \equiv P_0, i = 1, \cdots, n,$$

即 n 个试验有相同的概率空间，则称它们为 n 个 (重) 独立重复试验. 如果在 n 个独立重复实验中，每次试验的可能结果为两个:A 或 $\overline{A}$, 则称这种试验为 n 重伯努利试验.

彩票问题

例 0.57 某彩票每周开奖一次，每次提供十万分之一的中奖机会，且各周开奖是独立的。若你每周买一张彩票，坚持十年（每年按 52 周计算），试求未中奖的概率. 解: 依假设，每次中奖的概率为 10^{-5}, 于是每次不中奖的概率是 $1-10^{-5}$. 另外十年一共购买 520 次彩票，而每次开奖都是独立的，相当于进行了 520 次独立重复试验. 若记 A_i 为 “第 i 次开奖不中奖”，则 $A_1,\cdots,A_{520}$ 相互独立，从而

$$P(A_1A_2\cdots A_{520})=(1-10^{-5})^{520}=0.9948$$

第八讲：随机变量及其分布

Email: xzhangseu@seu.edu.cn

东南大学　　数学学院

随机变量的引入

- 赌徒输光：甲和乙初始资金分别为 $i, a-i$ 元，每一局甲赢的概率为 p
- 关注的问题
  - 甲最终获胜的概率
  - 甲乙两人在任意时刻的剩余资产:k 轮赌博后恰好剩下 j 元
  - k 轮赌博后甲乙两人资产的差额 Z
  - 赌博持续时间 R
- 表示方法:
  - E :={甲最终获胜 }, $Q_i := P(\mathrm{E})$
  - A_{jk} :={甲在 k 轮赌博后恰好剩下 j 元 }
  - B_{jk} :={乙在 k 轮赌博后恰好剩下 j 元 }
  - k 轮赌博后甲乙两人资产的差额如何表示?
  - 赌博持续时间 R 如何表示?
  - 很难用事件来表示或者表示很复杂

随机变量的引入

- $X_k :=$ 甲在 k 轮赌博后的资产
  - 乙在 k 轮赌博后的资产 $Y_k = a - X_k$
  - 资产差额:$Z = X_k - Y_k = 2X_k - a$
  - 持续时间:$R = \min\{n : X_n = 0, 或 Y_n = 0\}$
- X_k 取值的特点
  - 依赖于前面 k 次赌博这一“随机试验”的结果
  - 在“随机试验”完成之前，X_k 取值不确定，因此具有不确定性
  - k 次赌博“随机试验”一旦完成，X_k 的值必然确定
  - 记 k 次赌博“随机试验”样本空间为 Ω, 则给定 $\omega \in \Omega$, 则 X_k 值确定
- 以 $k = 2$ 为例，看一下 X_2 的取值情况，设 $i \geq 2$
  - $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, 其中
    $\omega_1 = (胜，胜), \omega_2 = (胜，败), \omega_1 = (败，胜), \omega_1 = (败，败)$
  - $X_2(\omega_1) = i + 2,\ X_2(\omega_2) = X_2(\omega_3) = i,\ X_2(\omega_4) = i - 2$
- X_k 可以看作定义在样本空间 Ω 上的函数，即
  $X_k : \Omega \to \{0, 1, \cdots, a\}$
- 一般的，随机变量可以看作从样本空间到实数的映射:$X : \Omega \to R$

定义 1.1 (直观定义) 称 X 为随机变量，如果 X 是从样本空间 Ω 到实数的映射，即 $X : \Omega \to R$.

例 1.1 考虑硬币抛掷两次的随机试验，其样本空间为

$$\Omega = \{HH, HT, TH, TT\}.$$

- 令 X 表示正面朝上的次数，则 X 是一个随机变量，相应的映射如下
$$X(TT) = 0, X(HT) = X(TH) = 1, X(HH) = 2.$$
- 令 Y 表示反面朝上的次数，则 $Y = 2 - X$, 对于任意的 $\omega \in \Omega$, 有 $Y(\omega) = 2 - X(\omega)$.
- 设 I 是由第一次掷硬币的结果决定的随机变量：若第一次硬币正面朝上则 $I = 1$, 反之 $I = 0$, 即
$$I(HH) = I(HT) = 1, \quad I(TH) = I(TT) = 0.$$
- 若正面记 1, 反面记 0, 此时 $\Omega = \{(1,1),(1,0),(0,1),(0,0)\}$. 上述的 X, Y 和 I 可表示为:
$$X(\omega_1, \omega_2) = \omega_1 + \omega_2, Y(\omega_1, \omega_2) = 2 - \omega_1 - \omega_2, I(\omega_1, \omega_2) = \omega_1.$$

与函数对比分析

- 数学分析中的函数 $f(x)$
  - $f(x): D \to R$, 其中 $D \subset R$
  - R 上可定义距离 $d(x,y) = |x-y|$
  - 可根据距离 d 定义函数的连续性
- 概率论中的随机变量 X
  - $X(\omega): \Omega \to R$
  - 定义域 Ω 没有距离的定义，但有事件域 $\mathcal{F}$ 及定义其上的概率 P, 即具有结构 $(\Omega, \mathcal{F}, P)$
  - 值域 R 有距离，但因定义域无距离，故无法考虑随机变量的连续性
  - 值域 R 还有 σ- 代数 $\mathcal{B} := \mathcal{B}(R)$, 有可测空间结构 $(R, \mathcal{B})$
- 是否需要对 $X(\omega): \Omega \to R$ 做一些额外的限定，以便更好的研究 X?
- 从赌徒输光问题可以看出，对随机变量 X, 我们会关注
  - $X(\omega) = x$ 的概率，$X(\omega) \le x, X(\omega) \in [b,c]$ 的概率
  - 更一般的，$\forall B \in \mathcal{B}(R), X(\omega) \in B$ 的概率
- 概率的定义域是 $\mathcal{F}$, 要想计算 $X(\omega) \in B$ 的概率，当且仅当

$$\forall B \in \mathcal{B}(R),\ X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}$$

定义 1.2 (可测映射) 设 $(\Omega, \mathcal{F})$ 和 $(E, \mathcal{E})$ 为两个可测空间并令 X 为从样本空间 Ω 到 E 的映射，即 $X(\omega): \Omega \to E$. 若对任意的 $B \in \mathcal{E}$ 均有

$$X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}$$

则称 X 为 $(\Omega, \mathcal{F})$ 到 $(E, \mathcal{E})$ 的可测映射.

若将上述定义中的可测空间 $(E, \mathcal{E})$ 更换为 $(R, \mathcal{B})$，则

定义 1.3 (可测函数或随机变量) 设 $(\Omega, \mathcal{F})$ 是可测空间，X 为从样本空间 Ω 到实数集 R 的映射，即 $X(\omega): \Omega \to R$. 如果对 $\forall B \in \mathcal{B}$ 均有

$$X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}.$$

则称 X 为可测空间 $(\Omega, \mathcal{F})$ 上的可测函数或随机变量.

定义 1.4 (随机变量的另一定义) 设 $(\Omega, \mathcal{F})$ 是可测空间，X 为从样本空间 Ω 到实数集 R 的映射，即 $X(\omega): \Omega \to R$. 如果对任意的 $x \in R$ 均有

$$\{\omega : X(\omega) \leq x\} \in \mathcal{F} \text{ 或} \{\omega : X(\omega) < x\} \in \mathcal{F}$$

则称 $X(\omega)$ 是可测空间 $(\Omega, \mathcal{F})$ 上的随机变量，简称随机变量.

由定义 1.4 推定义 1.3: 仅需说明若定义 1.4 成立，则对任意 $B \in \mathcal{B}$ 均有

$$X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}.$$

即只需说明以下集合包含关系成立即可

$$\mathcal{A} := \{A : X^{-1}(A) \in \mathcal{F}\} \supset \mathcal{B}$$

欲证上面包含关系成立，我们只需说明以下两点即可:

1. $\mathcal{A}$ 是 σ 代数;
2. $O_1 := \{(-\infty, x] : x \in R\} \subset \mathcal{A}$.

再由 $\mathcal{B} := \sigma(O_1)$ 知 $\mathcal{B} \subset \mathcal{A}$.

$\mathcal{A} := \{A : X^{-1}(A) \in \mathcal{F}\}$ 为 σ 代数

- $X^{-1}(R) = \{\omega : X(\omega) \in R\} = \Omega \in \mathcal{F}$, 故 $R \in \mathcal{A}$;
- 若 $A \in \mathcal{A}$, 即 $X^{-1}(A) \in \mathcal{F}$, 则

$$
\begin{aligned}
X^{-1}(\overline{A}) &= \{\omega : X(\omega) \in \overline{A}\} = \{\omega : X(\omega) \notin A\} \\
&= \overline{\{\omega : X(\omega) \in A\}} = \overline{X^{-1}(A)} \\
&\in \mathcal{F}
\end{aligned}
$$

- 对于 $A_j \in \mathcal{A}, j = 1, 2, \cdots$, 有 $X^{-1}(A_j) \in \mathcal{F}, j = 1, 2, \cdots$. 从而

$$
\begin{aligned}
X^{-1}(\cup_{j=1}^{\infty} A_j) &= \{\omega : X(\omega) \in \cup_{j=1}^{\infty} A_j\} = \cup_{j=1}^{\infty} \{\omega : X(\omega) \in A_j\} \\
&= \cup_{j=1}^{\infty} X^{-1}(A_j) \\
&\in \mathcal{F}
\end{aligned}
$$

随机变量的分类

- 随机变量 X 是从样本空间 Ω 到实数 R 的映射，故根据其值域集合可粗略的分为两大类
  - 离散型随机变量：其值域集合是有限点集或可数点集即

$$X(\Omega) := \{X(\omega) : \omega \in \Omega\} = \{a_n\}_{n\geq 1}$$

  - 非离散型随机变量：其值域集合不是有限点集或可数点集

示性随机变量

例 1.2 设 Ω 是某随机试验的样本空间，$\mathcal{F}$ 为其事件域 (σ 代数)，则对于任意的 $A \in \mathcal{F}$，示性函数 $I_A(\omega) := \begin{cases} 0, & \omega \notin A \\ 1, & \omega \in A \end{cases}$ 是随机变量.

解: 由示性函数的定义知:

$$\{\omega : I_A(\omega) \leq x\} = \begin{cases} \emptyset, & x < 0, \\ \overline{A}, & x \in [0,1), \\ \Omega, & x \geq 1. \end{cases}$$

显然，无论 x 取何值，均有 $\{\omega : I_A(\omega) \leq x\} \in \mathcal{F}$

定理 1.1 若 $X, Y, \{X_n, n \geq 1\}$ 都为概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量，则

1. $|X|$，$aX + bY, (a, b \in R)$ 均为随机变量;
2. $X^+ := X \vee 0, X^- := (-X) \vee 0$ 均为随机变量;
3. XY 为随机变量;
4. 若 X/Y 处处有意义，则 X/Y 为随机变量;
5. $\inf_n X_n, \sup_n X_n, \liminf_{n\to\infty} X_n, \limsup_{n\to\infty} X_n$ 均为随机变量.

证明:

1.
   - $\{\omega : |X| < x\} = \{\omega : -x < X < x\} = \{\omega : X < x\} \cap \overline{\{\omega : X \leq -x\}} \in \mathcal{F}$;
   - $\{\omega : aX < x\} = \{\omega : X < \dfrac{x}{a}\} \in \mathcal{F}$, (当 $a > 0$ 时);
   - 设 Q 为有理数集，则

$$
\begin{aligned}
\{\omega : X + Y < x\} &= \{\omega : X < x - Y\} = \cup_{r \in Q} \{\omega : X < r < x - Y\} \\
&= \cup_{r \in Q} \{\omega : X < r, Y < x - r\} \\
&= \cup_{r \in Q} (\{\omega : X < r\} \cap \{\omega : Y < x - r\}) \\
&\in \mathcal{F}
\end{aligned}
$$

2. 注意到 $X^+=\dfrac{|X|+X}{2}, X^-=\dfrac{|X|-X}{2}$，易得 X^+, X^- 均为随机变量;
3. 首先假定 X,Y 非负，则对任意的 $x>0$ 有

$$\begin{aligned}\{XY<x\} &= \{X=0\}\cup\{Y=0\}\cup\left(\cup_{r\in Q_+}\left[\{X<r\}\cap\{Y<\frac{x}{r}\}\right]\right)\\ &\in \mathcal{F}.\end{aligned}$$

对一般的 X,Y, 由 X^+,X^-,Y^+,Y^- 为随机变量，可得

$$XY=(X^+-X^-)(Y^+-Y^-)=(X^+Y^++X^-Y^-)-(X^+Y^-+X^-Y^+)$$

为随机变量.
4. 设 $|Y|>0$ 处处成立，易证 $\dfrac{1}{Y}$ 是随机变量，故 $\dfrac{X}{Y}=X\cdot\dfrac{1}{Y}$ 为随机变量.
5. 对任意的 $x\in R$, 我们有

$$\{\inf_n X_n<x\}=\cup_n\{X_n<x\},\quad \{\sup_n X_n\le x\}=\cap_n\{X_n\le x\}$$

例 1.3 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间，$A_i \in \mathcal{F}, i = 1, \cdots, n$ 为 Ω 的一个分割，$a_i, i = 1, \cdots, n$ 为 n 个不同的实数，则

$$X(\omega) := \sum_{i=1}^{n} a_i I_{A_i}(\omega) \tag{20}$$

作为 n 个示性随机变量的线性组合，仍为随机变量。我们称形如 (20) 的 $X(\omega)$ 为简单随机变量.

定理 1.2 设 X 是可测空间 $(\Omega, \mathcal{F})$ 上的随机变量，$g(x)$ 为 $(R, \mathcal{B}) \to (R, \mathcal{B})$ 上的可测函数，证 $Y := g(X)$ 为 $(\Omega, \mathcal{F})$ 上的随机变量.

解: 注意到，对任意的 $B \in \mathcal{B}$, $g^{-1}(B) \in \mathcal{B}$, 故

$$\begin{aligned} Y^{-1}(B) &= \{\omega : Y(\omega) \in B\} \\ &= \{\omega : g(X(\omega)) \in B\} \\ &= \{\omega : X(\omega) \in g^{-1}(B)\} \\ &= X^{-1}(g^{-1}(B)) \\ &\in \mathcal{F} \end{aligned}$$

对于 $(\Omega,\mathcal{F},P)$ 上的任意非负随机变量 X 及自然数 n，我们可将 Ω 按 X 的取值进行分割。即令

$$\begin{aligned} A_k(\omega) &:= \{\omega : \frac{k}{2^n} \le X(\omega) < \frac{k+1}{2^n}\}, k = 0,1,\cdots,n2^n-1 \\ A_{n2^n}(\omega) &:= \{\omega : X(\omega) \ge n\} \end{aligned}$$

则

$$X_n(\omega) := \sum_{k=0}^{n2^n} \frac{k}{2^n} I_{A_k}(\omega)$$

为简单随机变量且随机变量序列 $\{X_n, n \ge 1\}$ 满足

$$0 \le X_1(\omega) \le X_2(\omega) \le \cdots \le X_n(\omega) \to X(\omega)$$

注意到对任意的 $k=0,1,\cdots,n2^n-1$,

$$\begin{aligned}
A_k(\omega) &= \{\omega : \frac{k}{2^n} \le X(\omega) < \frac{k+1}{2^n}\} = \{\omega : \frac{2k}{2^{n+1}} \le X(\omega) < \frac{2(k+1)}{2^{n+1}}\} \\
&= \{\omega : \frac{2k}{2^{n+1}} \le X(\omega) < \frac{2k+1}{2^{n+1}}\} \cup \{\omega : \frac{2k+1}{2^{n+1}} \le X(\omega) < \frac{2k+2)}{2^{n+1}}\} \\
&= A_k^1(\omega) \cup A_k^2(\omega)
\end{aligned}$$

故在集合 $A_k(\omega), k=0,1,\cdots,n2^n-1$ 上,

$$X_{n+1}(\omega) = \begin{cases} \frac{2k}{2^{n+1}}, & \omega \in A_k^1(\omega) \\ \frac{2k+1}{2^{n+1}}, & \omega \in A_k^2(\omega) \end{cases} \quad (\ge \frac{k}{2^n} = X_n(\omega))$$

而在 $A_{n2^n}(\omega) := \{\omega : X(\omega) \ge n\} = \{\omega : X(\omega) \ge \frac{n2^{n+1}}{2^{n+1}}\}$ 上，显然有

$$X_{n+1}(\omega) \ge n = X_n(\omega).$$

$X_n(\omega)$ 的收敛性

注意到，对任意的 ω, 必定存在 k 使得

$$\frac{k}{2^n} \le X(\omega) < \frac{k+1}{2^n},$$

从而

$$0 \le X(\omega) - X_n(\omega) \le \frac{1}{2^n}$$

显然有

$$\lim_{n\to\infty} X_n(\omega) = X(\omega).$$

定理 1.3 对 $(\Omega, \mathcal{F}, P)$ 上的实值变量 $X(\omega)$ 为随机变量的充要条件是：存在简单随机变量序列 $\{X_n(\omega), n \geq 1\}$ 使得

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \quad \forall \omega \in \Omega$$

而且当 X 非负时，还可选取 $\{X_n(\omega), n \geq 1\}$ 为非负单调不减的简单随机变量序列.

定理 1.4 设 X 为 $(\Omega, \mathcal{F}, P)$ 上的随机变量，对于 Borel 集 B, 定义集函数 $\mathbf{F}(B)$ 如下：

$$\mathbf{F}(B) := P(X^{-1}(B)) = P \circ X^{-1}(B) = P(X \in B) \quad (21)$$

则 $\mathbf{F}(\cdot)$ 为 $(R, \mathcal{B})$ 上的概率，称之为随机变量 X 的诱导概率测度.

定义 1.5 称由 (21) 式定义在 $(R, \mathcal{B})$ 上的概率测度 $\mathbf{F}(\cdot)$ 为随机变量 X 的概率分布，简称分布.

- 给定概率空间 $(\Omega, \mathcal{F}, P)$，任给随机变量均可在 $(R, \mathcal{B})$ 上诱导出一个概率测度。由此可见，在同一个可测空间上可以定义不同的概率测度.
- 对于任意的 $B \in \mathcal{B}$，随机变量 X 落入 B 中的概率可通过 B 的概率测度 $\mathbf{F}(B)$ 得出。这也就是说，概率分布 $\mathbf{F}(\cdot)$ 完全刻画了随机变量 X 取值的概率规律.

如果我们将 $(R, \mathcal{B})$ 上的测度仅局限于集类 $\mathcal{P} := \{(-\infty, x], x \in R\}$ 上，由于 $\mathcal{P}$ 中的每条半直线被它的右端点 x 所决定，于是集函数 F 就化为 R 上的点函数.

定义 1.6 对于随机变量 X 而言，称 x 的函数

$$F(x) := \mathbf{F}((-\infty, x]) = P(X \leq x)$$

为 X 的概率分布函数或累积分布函数，简称分布函数并记作 $X \sim F(x)$，有时也以 $F_X(x)$ 表明是 X 的分布函数.

注 1.1 也有一些教材按如下方式定义分布函数:

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

分布函数的性质

定理 1.5 任一分布函数 $F(x)$ 都具有以下三条基本性质

1. 单调性非降性：$F(x)$ 是单调非减函数即对任意的 $x_1 < x_2$，有 $F(x_1) \le F(x_2)$；
2. 右连续性：$F(x)$ 是 x 的右连续函数，即

$$F(x_0) = F(x_0+) := \lim_{x \to x_0+} F(x)$$

3. 规范性：对任意的 x 有，$0 \le F(x) \le 1$ 且

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

分布函数性质的证明

1. 对任意的 $x<y$, $F(y)-F(x)=P(x<X\le y)\ge 0$;
2. 因 $F(x)$ 是单调有界非降函数，所以其任一点 x_0 的右极限 $F(x_0+)$ 必存在，为证其连续性，只需证对单调上下降且收敛至 x_0 的数列 $\{x_n\}$ 有 $\lim_{n\to\infty}F(x_n)=F(x_0)$ 即可。注意到

$$
\begin{aligned}
\lim_{n\to\infty}F(x_n) &= \lim_{n\to\infty}P(X\le x_n)=P(\cap_{n=1}^{\infty}\{X\le x_n\})=P(X\le x_0)\\
&= F(x_0)
\end{aligned}
$$

3. 由 F 的单调性及概率的连续性可知

$$
\begin{aligned}
F(+\infty) &= \lim_{n\to+\infty}F(n)=\lim_{n\to+\infty}P(X\le n)\\
&= P(\cup_{n=1}^{\infty}\{X\le n\})=P(X<\infty)=1
\end{aligned}
$$

同理可证 $F(-\infty)=0$.

事件概率的分布函数表示

- $P(X > b) = 1 - F(b) := \mathbf{F}((b, \infty))$;
- $P(a < X \leq b) = F(b) - F(a) := \mathbf{F}((a, b])$;
- $P(X < a) = F(a-) := \mathbf{F}((-\infty, a))$;
- $P(X = a) = F(a) - F(a-) := \mathbf{F}(\{a\})$;
- $P(X \geq b) = 1 - F(b-) := \mathbf{F}([b, \infty))$;
- $P(a \leq X < b) = F(b-) - F(a-) := \mathbf{F}([a, b))$;
- $P(a \leq X \leq b) = F(b) - F(a-) := \mathbf{F}([a, b])$;
- $P(a < X < b) = F(b-) - F(a) := \mathbf{F}((a, b))$;

- 对于不交区间并 $\cup_{i=1}^{n}[a_i, b_i)$,
  $P(X \in \cup_{i=1}^{n}(a_i, b_i]) = \sum_{i=1}^{n}[F(b_i) - F(a_i)] := \mathbf{F}(\cup_{i=1}^{n}(a_i, b_i])$
- 一般的，$P(X \in B) = \mathbf{F}(B) = \int_B dF(x)$;
- 事实上根据测度扩张定理，由分布函数所确定的定义在 $\mathcal{P}$ 上的集函数 $\mathbf{F}((-\infty, x]) := F(x)$ 可以唯一的扩张到 $\mathcal{B} := \sigma(\mathcal{P})$ 上，成为 $\mathcal{B}$ 上的概率测度，扩张后的概率测度称之为分布函数 $F(x)$ 所引出的勒贝格 - 斯蒂尔吉斯测度。实际上这个 $\mathbf{F}$ 正好是我们前面引进的概率分布.

定义 1.7 (离散型随机变量) 如果随机变量 X 只取有限个值 $x_1, x_2, \cdots, x_n$ 或可列个值 $x_1, x_2, \cdots$, 就称 X 为离散型随机变量，简称离散随机变量，其分布函数称之为离散型的.

定义 1.8 (离散型随机变量的分布列或概率质量函数) 对于离散型随机变量 X, 称 X 取值 x_k 的概率

$$p_k := p(x_k) = P(X = x_k), k = 1, 2, \cdots,$$

为 X 的概率分布列或简称为分布列，记 $X \sim \{p_k\}$. 分布列也常用下面的矩阵来表示

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

容易验证，分布列有以下性质

1. 非负性: $p_k \geq 0, k = 1, 2, \cdots$;
2. 正则性: $\sum_k p_k = 1$

- 由概率分布的定义，对任意的 $B \in \mathcal{B}$，我们有

$$\begin{aligned}\mathbf{F}(B) &= P(X \in B) = P(\cup_{k:x_k \in B}\{X = x_k\}) \\ &= \sum_{k:x_k \in B} P(X = x_k) = \sum_{k:x_k \in B} p_k\end{aligned}$$

- 由分布函数的定义知，

$$\begin{aligned}F(x) &= P(X \le x) = P(\cup_{k:x_k \le x}\{X = x_k\}) \\ &= \sum_{k:x_k \le x} P(X = x_k) = \sum_{k:x_k \le x} p_k\end{aligned}$$

- 易见离散型随机变量 X 的分布函数是一个纯跳跃函数：在 X 的每个可能取值 x_k 上有跃度 p_k，在每个不含 x_k 的区间上恒取常值.

退化随机变量

例 1.4 常数 c 可看作仅取一个值的随机变量 X, 即

$$P(X=c)=1$$

这个分布常称为 单点分布 或 退化分布，其分布函数为

$$F(x)=\begin{cases}0, & x<c\\ 1, & x\geq c\end{cases}$$

例 1.5 计算例 1.1 中的所有随机变量的分布列或概率质量函数.

- X 表示正面朝上的次数，其概率质量函数 p_X 为:

$$p_X(0)=P(X=0)=1/4,\quad p_X(1)=P(X=1)=1/2,$$
$$p_X(2)=P(X=2)=1/4,\quad p_X(x)=P(X=x)=0, x\neq 0,1,2.$$

- $Y=2-X$, 表示反面朝上的次数。注意到

$$P(Y=y)=P(2-X=y)=P(X=2-y)=p_X(2-y),$$

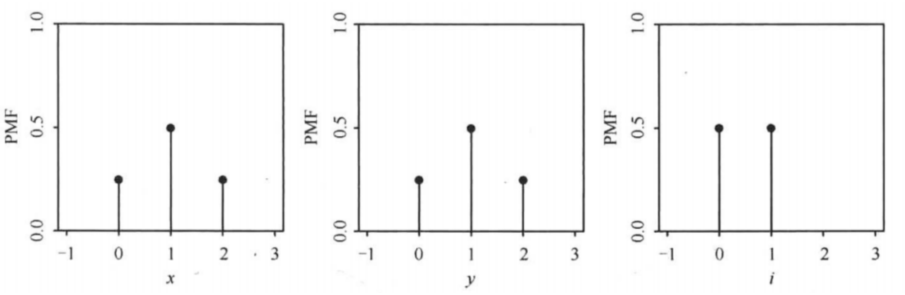
因此，随机变量 Y 的概率质量函数为

$$p_Y(0)=P(Y=0)=1/4,\quad p_Y(1)=P(Y=1)=1/2,$$
$$p_Y(2)=P(Y=2)=1/4,\quad p_Y(y)=P(Y=y)=0, y\neq 0,1,2.$$

- I 表示第一次是否正面朝上的示性随机变量.

$$p_I(0)=P(I=0)=1/2,\quad p_I(1)=P(I=1)=1/2,$$
$$p_I(i)=P(I=i)=0, i\neq 0,1.$$

X, Y 和 I 的概率质量函数图



例 1.6 掷两颗骰子，其样本空间 Ω 含有 36 个等可能的样本点

$$\Omega = \{(x, y) : x, y = 1, 2, \cdots, 6\}$$

令 X 和 Y 表示每个骰子分别出现的点数。试求下面随机变量的分布列:

1. $T_1 := X + Y =$ 骰子点数之和;
2. $T_2 := 14 - (X + Y)$;
3. $T_3 :=$ 点数为 6 点的骰子的个数;
4. $T_4 := \max\{X, Y\} =$ 骰子的最大点数

- T_1, T_2 的概率分布列为

$$\begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{pmatrix}$$

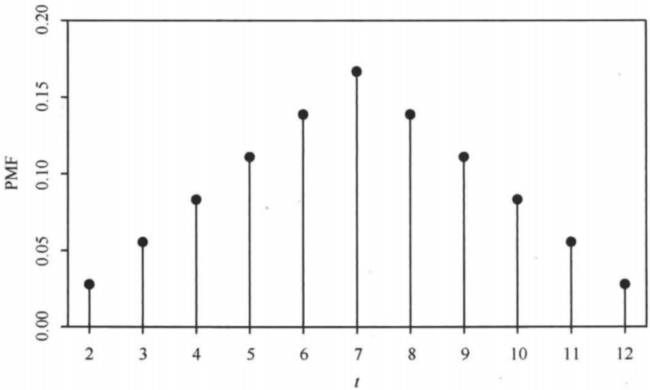


图 3.4　两个骰子的点数之和的概率质量函数。

- T_3 的概率分布列为

$$\begin{pmatrix} 0 & 1 & 2 \\ \frac{25}{36} & \frac{10}{36} & \frac{1}{36} \end{pmatrix}$$

- T_4 的概率分布列为

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{36} & \frac{3}{36} & \frac{5}{36} & \frac{7}{36} & \frac{9}{36} & \frac{11}{36} \end{pmatrix}$$

定义 1.9 (连续型随机变量) 设 X 为一随机变量，$F(x)$ 为随机变量 X 的分布函数，如果存在非负可积函数 $p(x)$ 使得

$$F(x) = \int_{-\infty}^{x} p(y)dy \tag{22}$$

则称 X 为连续型随机变量，其分布函数称之为连续型分布函数，函数 $p(x)$ 称为 X 的概率密度函数，简称密度函数或密度.

注 1.2

- 能够表为 (22) 式变上限积分的函数 $F(x)$ 在分析中称为绝对连续函数。绝对连续函数必为连续函数.
- 在若干个点上或零测集上改变密度函数 $p(x)$ 的值并不影响其积分的值，从而不影响分布函数 $F(x)$ 的值，这意味着连续分布的密度函数不唯一.

容易验证，随机变量 X 的密度函数有以下性质

1. 非负性：$p(x) \geq 0$;
2. 正则性：$\int_{-\infty}^{\infty} p(x)dx = 1$.

连续型随机变量的一些常见性质

- $p(x) = F'(x)$;
- $P(a < X \le b) = F(b) - F(a) = \int_a^b p(x)dx$;
- $0 \le P(X = a) \le P(X \in (a-\epsilon, a)) = \int_{a-\epsilon}^a p(x)dx \xrightarrow{\epsilon \to 0} 0$, 故 $P(X = a) = 0$，即连续型随机变量取值单点的概率为 0;
- $P(a < X \le b) = P(a \le X < b) = P(a \le X \le b) = P(a < X < b)$;
- 对任意的 Borel 集 B,

$$P(X \in B) = \int_B p(x)dx$$

- $P(X \in [x, x+\Delta x]) = \int_x^{x+\Delta x} p(y)dy = p(\xi)\Delta x \approx p(x)\Delta x$

例 1.7 定义函数 $F(x)$ 如下

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

试说明

1. $F(x)$ 为分布函数;
2. $F(x)$ 既非离散型也非连续型分布;
3. $F(x)$ 可分解为

$$F(x)=\frac{1}{2}F_1(x)+\frac{1}{2}F_2(x)$$

其中

$$F_1(x)=\begin{cases}0, & x<0,\\ 1, & x\ge 0.\end{cases}\qquad F_2(x)=\begin{cases}0, & x<0,\\ x, & 0\le x<1,\\ 1, & x\ge 1.\end{cases}$$

定理 1.6 (勒贝格分解) 对任一分布函数 $F(x)$ 有如下分解

$$F(x) = c_1F_1(x) + c_2F_2(x) + c_3F_3(x),$$

其中常数 $c_1, c_2, c_3 \geq 0, c_1 + c_2 + c_3 = 1$, 而 $F_1(x), F_2(x), F_3(x)$ 都是分布函数，$F_1(x)$ 为纯跳跃函数，$F_2(x)$ 为绝对连续函数，$F_3(x)$ 为奇异函数.

- 上述定理中奇异函数的含义及定理的证明可参见一般的实变函数论教科书，这里我们不再详述，仅指出几种特殊情况:
  - 在分解式中取 $c_1 = 1, c_2 = c_3 = 0$ 便得到我们所讨论的离散型分布函数;
  - 在分解式中取 $c_2 = 1, c_1 = c_3 = 0$ 便得到连续型分布函数;
  - 若取 $c_3 = 0, c_1 \neq 0, c_2 \neq 0, c_1 + c_2 = 1$ 便得到离散与连续混合分布
- 从上面分析可看出，随机变量除了离散型与连续型外还有很多其他类型.

第九讲：常见的离散型随机变量

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

伯努利实验

- 只有两种可能结果的试验称为伯努利试验；例如抽检产品，可能是合格品，也可能是次品；掷两颗骰子，可能得到同点，也可能得到不同点，等等都是伯努利试验.
- 伯努利试验的样本空间 Ω 并不一定只含有两个样本点，有时只是把我们所关心的一部分样本点归结为一种结果 A, 同时把其余的样本点的集合看作另一种结果 $\bar{A}$;
- 在上述掷骰子的试验中，样本空间 Ω 共含有 36 个样本点，如果我们只关心同点是否发生，就可以把其中的 6 个样本点组成的事件 $A := \{(i,i) : i = 1, \cdots, 6\}$ 视为一种结果，而其余的 30 个样本点组成另一结果 $\bar{A} := \{$不同点$\}$;
- 此外我们不再关心由 Ω 的其他非空子集组成的事件，于是对于伯努利试验而言，事件 σ 代数应取为 $\mathcal{F} = \{\emptyset, A, \bar{A}, \Omega\}$;
- 通常把结果 A 称作"成功", 而把 $\bar{A}$ 称作"失败";
- 再取定成功失败的概率 $p = P(A), q = P(\bar{A})$ $(p > 0, q > 0$ 且 $p + q = 1)$, 则建立了一次伯努利试验的概率空间 $(\Omega, \mathcal{F}, P)$.

- 在概率论的理论与应用中，经常以一系列独立重复的伯努利试验作为概率模型;
- 所谓重复，粗略的说即各次试验的概率空间都是上述的 $(\Omega, \mathcal{F}, P)$;
- 而 n 个试验的独立性则是指各次试验的结果互不影响，即对于第 i 次试验的任何结果 $E_i(i=1,\cdots,n)$ 均有

$$P(E_1E_2\cdots E_n)=P(E_1)P(E_2)\cdots P(E_n)$$

- 将一次伯努利试验独立重复 n 次，称作 n 重伯努利试验;
- 将一次伯努利试验独立地重复下去所得到的一系列试验，称为可列重伯努利试验.

- X:n 重伯努利试验中成功 (事件 A 发生) 的次数;
- X 的所有可能取值为: $0, 1, \cdots, n$;
- 下面我们考虑 X 的分布列
  - n 重伯努利试验的样本空间:
    $\Omega := \{\omega = (\omega_1, \cdots, \omega_n) : \omega_i\text{或者为}A\text{或者为}\bar{A}\}$
  - 样本空间样本点的个数为 2^n 个;
  - $\{X = k\} = \{\omega = (\omega_1, \cdots, \omega_n) : \omega_1, \cdots, \omega_n\text{中有}k\text{个}A\}$, 共包含 C_n^k 个样本点;
  - 若任给样本点 $\omega = (\omega_1, \cdots, \omega_n) \in \{X = k\}$, 则意味着 $\omega_1, \cdots, \omega_n$ 中有 k 个 A, $n - k$ 个 $\bar{A}$, 故由独立性可知

$$P(\omega) = p^k(1-p)^{n-k}$$

  - 而事件 $\{X = k\}$ 中共有 C_n^k 个类似的 ω, 故

$$P(X = k) = C_n^k p^k (1-p)^{n-k} := b(k; n, p), k = 0, 1, \cdots, n$$

    这个分布常称为二项分布，记为 $X \sim B(n, p)$.
  - 容易验证，$\sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$

两点分布

- $n=1$ 时的二项分布 $B(1,p)$ 称为二点分布，或 $0-1$ 分布，或称伯努利分布，其分布列为

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

或记为

$$\begin{pmatrix} 0, & 1 \\ 1-p, & p \end{pmatrix}$$

- $B(1,p)$ 主要用于描述一次伯努利试验中成功的次数 (0 或 1);
- 若记 X_i 表示第 i 次伯努利试验中成功的次数，则 X_i 相互独立且有

$$X=X_1+X_2+\cdots X_n$$

即二项分布随机变量可写为 n 个独立同为两点分布随机变量的和.

二项分布的性质

- 对 $k \geq 1$, 考虑比值

$$
\begin{aligned}
\frac{b(k;n,p)}{b(k-1;n,p)} &= \frac{C_n^k p^k (1-p)^{n-k}}{C_n^{k-1} p^{k-1} (1-p)^{n-k+1}} = \frac{(n-k+1)p}{k(1-p)} \\
&= \frac{k(1-p) + (n+1)p - k}{k(1-p)} = 1 + \frac{(n+1)p - k}{k(1-p)}
\end{aligned}
$$

- 当 $(n+1)p > k$ 时，$b(k;n,p) > b(k-1;n,p)$;
- 当 $(n+1)p < k$ 时，$b(k;n,p) < b(k-1;n,p)$;
- 从而，对于固定的 $n, p, \{X = k\}$ 的概率 $b(k;n,p)$ 先随 k 增大而增大，再随 k 增大而减小，故必有最大值:
  - 如果 $m := (n+1)p$ 为整数，则 $b(m;n,p) = b(m-1;n,p)$ 同为 $b(k;n,p)$ 的最大值
  - 如果 $(n+1)p$ 不为整数，则 $b(k;n,p)$ 在 $m = [(n+1)p]$ 处取到最大值 (此处 $[a]$ 表示不超过 a 的最大整数)
- 我们称使得 $b(k;n,p)$ 取得最大值的 m 为二项分布随机变量的最可能值，或称为 n 重伯努利试验中最可能的成功次数.

二项分布的性质

定理 1.7 设 $X \sim B(n,p)$, 且 $q=1-p$(通常用 q 表示伯努利试验失败的概率), 则有 $n-X \sim B(n,q)$.

- 借用二项分布的直观定义：将 X 为 n 次独立伯努利试验成功的次数，则 $n-X$ 为这些试验中失败的次数.
- 互相交换成功与失败的角色，可知 $n-X \sim B(n,q)$.
- 也可从分布列 (概率质量函数) 的角度出发得到 $n-X \sim B(n,q)$.
- 令 $Y=n-X$, 则 Y 的分布列 (概率质量函数) 为

$$
\begin{aligned}
P(Y=k) = & P(n-X=k) = P(X=n-k) \\
= & \binom{n}{n-k} p^{n-k} q^k = \binom{n}{k} q^k p^{n-k},
\end{aligned}
$$

定理 1.8 设 $X \sim B(n,p)$, 其中 n 为偶数，$p=1/2$, 则 X 的分布关于 $n/2$ 对称，即对任意的非负整数 j, 均有

$$P(X=\frac{n}{2}+j)=P(X=\frac{n}{2}-j).$$

解: 由定理 1.7 可知，$n-X$ 同样服从 $B(n,1/2)$. 因此对任意非负整数 k 均有

$$P(X=k)=P(n-X=k)=P(X=n-k).$$

令 $k=n/2+j$, 即可得证.

例 1.8 设每台自动机床在运行过程中需要维修的概率均为 $p = 0.01$, 并且各机床是否需要维修相互独立。如果:

1. 每名维修工人负责看管 20 台机床;
2. 3 名维修工人负责看管 80 台机床;

求机床不能及维修的概率.

解: 1. 这是 $n = 20$ 重伯努利试验，参数 $p = 0.01$, 故需要维修的机床数 X 服从 $B(20, 0.01)$ 分布。故不能及时维修的概率为

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) \\ &= 1 - C_{20}^{0} 0.01^{0} (1 - 0.01)^{20} - C_{20}^{1} 0.01 (1 - 0.01)^{20-1} \approx 0.0169 \end{aligned}$$

2. 此时需要维修的机床数 X 服从 $B(80, 0.01)$ 分布，类似可得不能及时维修的概率为

$$P(X > 3) = 1 - \sum_{k=0}^{3} b(k; 80, 0.01) \approx 0.0087$$

例 1.9 在可列重伯努利试验中，求事件 $E := \{试验终将成功\}$ 的概率.

解: 考虑所求概率事件的反面即 $\overline{E} := \{试验永不成功\}$.若我们记

$$F_n := \{前n次试验均失败\},$$

则易知，$\{F_n\}$ 为单调下降事件序列，且

$$\lim_{n\to\infty} F_n = \cap_{n=1}^{\infty} F_n = \overline{E}$$

从而

$$P(\overline{E}) = P(\lim_{n\to\infty} F_n) = \lim_{n\to\infty} P(F_n) = \lim_{n\to\infty} C_n^0 p^0 (1-p)^n = 0$$

故

$$P(E) = 1 - P(\overline{E}) = 1 - 0 = 1$$

无论成功的概率有多小，但是试验最终成功的概率为 1, 也就是说小概率事件终将发生的概率为 1.

- 记 X 为可列重伯努利试验中首次成功的等待时间即首次成功所需要试验的次数;
- $\{X=k\} = \{\underbrace{\overline{A}\cdots\overline{A}}_{k-1\text{个}}A\}$, 故

$$P(X=k) = (1-p)^{k-1}p := g(k;p), k = 1, \cdots,$$

这个分布常称为几何分布，记为 $X \sim G(p)$.

定理 1.9 取值自然数的随机变量 X 为几何分布当且仅当 X 有无记忆性:

$$P(X>m+n|X>m)=P(X>n),\text{对任意的}m,n\geq 1. \tag{23}$$

证明: 若 X 为几何分布，则

$$P(X>m+n|X>m)=\frac{P(X>m+n,X>m)}{P(X>m)}=\frac{P(X>m+n)}{P(X>m)}$$

而

$$P(X>n)=\sum_{k=n+1}^{\infty}(1-p)^{k-1}p=\frac{(1-p)^n p}{1-(1-p)}=(1-p)^n$$

故

$$P(X>m+n|X>m)=\frac{(1-p)^{m+n}}{(1-p)^m}=(1-p)^n=P(X>n)$$

若 X 具有无记忆性，则由 ~ (23) 知

$$Q_n := P(X > n) > 0, \quad \text{对任意的} n \geq 1$$

并且有

$$Q_{m+n} = P(X > m+n) = P(X > m)P(X > m+n|X > m) = Q_m Q_n$$

从而

$$Q_m = Q_1^m$$

注意到 $Q_1 \in (0,1)$, 事实上,

- $Q_1 = P(X > 1) > 0$ 显然;
- 若 $Q_1 = 1$, 则对一切的 m 均有 $Q_m = P(X > m) = 1$, 这与 X 取自然数矛盾，故 $Q_1 \in (0,1)$.

故取 $p = 1 - Q_1 \in (0,1)$, 且对任意的 $k \geq 1$ 有

$$\begin{aligned} P(X = k) &= P(X > k-1) - P(X > k) = Q_{k-1} - Q_k \\ &= (1-p)^{k-1} - (1-p)^k = (1-p)^{k-1} p \end{aligned}$$

无记忆性

$$P(X > m + n | X > m) = P(X > n)$$

- 上述的无记忆性表明：已知试验了 m 次未获得成功，再加做 n 次试验仍不成功的概率，等于从开始算起做 n 次试验都不成成功的概率.
- 换句话放，已做过的 m 次失败的试验被忘记了；
- 产生几何分布的这种无记忆性的根本原因在于，我们进行的是独立重复试验，这是不学习，不总结经验的试验，已经做过的试验当然不会留下记忆.

例 1.10 10 把外形相同的钥匙中只有一把能打开门。现一一试开，试对每次试毕放回与不放回两种情形，分别求事件 $E := \{至多试3次能打开门\}$ 的概率.

解: 1. 放回情形是独立重复试验，属伯努利概型。以 X 表示首次打开门的等待时间，则 X 服从几何分布 $G(0.1)$. 故所求概率为

$$P(E) = P(X \leq 3) = \sum_{k=1}^{3} (1-0.1)^{k-1} 0.1 = 0.271$$

2. 不放回情形不再是独立重复试验，适用于古典概型。样本点总数 $n(\Omega) = C_{10}^3$, 而 $n(\overline{E}) = C_9^3$. 故

$$P(E) = 1 - P(\overline{E}) = 1 - \frac{C_9^3}{C_{10}^3} = 0.3$$

或令 $A_i := \{第i次取到能开门的钥匙\}$, 则 A_1, A_2, A_3 互不相容，由可加性及抽签的公平性可得

$$P(E) = P(\cup_{i=1}^3 A_i) = \sum_{i=1}^{3} P(A_i) = 0.3$$

帕斯卡 (Pascal) 分布：第 r 次成功的等待时间

- 记 X_r 为可列重伯努利试验中第 r 次成功的等待时间即第 r 次成功所需要试验的次数;
- 易见 X 的所有可能取值为 $k = r, r+1, \cdots$, 并且有

$$\begin{aligned}
\{X_r = k\} &= \{\text{前}k-1\text{次试验恰有}r-1\text{次成功且第}k\text{次成功}\} \\
P(X_r = k) &= P(\text{前}k-1\text{次试验恰有}r-1\text{次成功})P(\text{第}k\text{次成功}) \\
&= C_{k-1}^{r-1}p^{r-1}(1-p)^{k-1-(r-1)}p \\
&= C_{k-1}^{r-1}p^{r}(1-p)^{k-r} := f(k;r,p), \quad k = r, r+1, \cdots
\end{aligned}$$

这个分布常称为帕斯卡 (Pascal) 分布或负二项分布.

- $f(k;r,p), k = r, r+1, \cdots$, 可以成为离散型分布的密度，事实上: $f(k;r,p) > 0$ 显然，其和

$$\begin{aligned}
\sum_{k=r}^{\infty} f(k;r,p) &= \sum_{k=r}^{\infty} C_{k-1}^{r-1}p^{r}(1-p)^{k-r} \overset{k-r=i}{\underset{q=1-p}{=\!=\!=}} \sum_{i=0}^{\infty} C_{r+i-1}^{r-1}p^{r}q^{r+i-r} \\
&= \sum_{i=0}^{\infty} C_{r+i-1}^{i}p^{r}q^{i} = \sum_{i=0}^{\infty} C_{-r}^{i}p^{r}(-q)^{i} = p^{r}(1-q)^{-r} = 1
\end{aligned}$$

帕斯卡分布的性质

- 若帕斯卡分布中的 $r=1$, 则此时的帕斯卡分布即为几何分布;
- 如果记

$$\tau_1 = X_1, \quad \tau_n = X_n - X_{n-1}, \quad n > 1,$$

则随机变量 τ_n 是第 $n-1$ 次成功到第 n 次成功的间隔时间。显然有

$$X_r = \tau_1 + \tau_2 + \cdots + \tau_r$$

- 以后我们会看到: $\tau_1, \cdots, \tau_r$ 是 r 个相互独立的随机变量且每个 τ_k 均服从几何分布.

例 1.11 某人口袋中有两盒火柴，开始时每盒各装 n 根。每次他从口袋中任取一盒使用其中的一根火柴。求此人掏出一盒发现已空，而另一盒尚余 r 根的概率.

解: 记

$$E = \{掏出甲盒已空而乙盒尚余r根\}$$

则由对称性可知所求概率为 $2P(E)$. 若我们以取出甲盒为" 成功", 这便是一个成功率 $p = 1/2$ 的独立重复伯努利试验. 而

$$E = \{第n+1次成功发生在第2n-r+1次试验\}$$

故所求概率为

$$\begin{aligned} 2P(E) &= 2f(2n-r+1; n+1, 1/2) = 2C_{2n-r}^{n}\left(\frac{1}{2}\right)^{n+1}\left(\frac{1}{2}\right)^{2n-r-n} \\ &= C_{2n-r}^{n}2^{r-2n} \end{aligned}$$

例 1.12 1654 年，当时的职业赌徒 DeMere 爵士向法国的大数学家 Pascal 提出如下问题：甲乙两人各下赌注 m 元，商定先胜三局者取得全部赌金。假定在每一局中二人获胜的机会相等，且各局胜负相互独立。如果当甲胜一局而乙胜零局时赌博被迫中止，问赌注如何分?

- 为解决这个问题，Pascal 与当时声望很高的数学家 Fermat 建立了通信联系。他们进行了卓有成效的讨论，不仅完满的回答了分赌注问题，而且为解决其他概率问题建立起了框架，极大的促进了概率论的建立与发展;
- Pascal 令人信服的指出，赌金的分法应当取决于若赌博能继续进行下去甲乙各自获胜的概率，这个概率即为在 $p = 0.5$ 的可列重伯努利试验中 2 次成功发生在 3 次失败之前的概率;
- 更一般的，下面我们求一下 n 次成功发生在 m 次失败之前的概率.

例 1.13 在可列重伯努利试验中，求下面事件的概率:

$$E = \{n\text{次成功发生在}m\text{次失败之前}\}$$

解: 记 $F_k = \{\text{第}n\text{次成功发生在第}k\text{次试验}\}$, 则

$$E = \cup_{k=n}^{n+m-1} F_k$$

从而由 F_k 的互不相容性可得

$$P(E) = \sum_{k=n}^{n+m-1} P(F_k) = \sum_{k=n}^{n+m-1} C_{k-1}^{n-1} p^n (1-p)^{k-n}$$

利用上面的公式可计算 $n = 2, m = 3, p = 1/2$ 时，其相应的概率为

$$P(\text{甲胜}) = P(E) \xlongequal{n=2,m=3,p=1/2} \frac{11}{16}$$

故赌注应以 $11 : 5$ 的比例分配给甲乙两人.

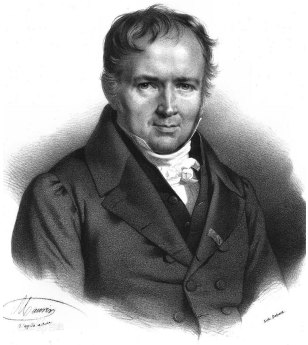


图 1: 泊松

- 西莫恩 德尼 泊松：法国数学家、几何学家和物理学家；
- 泊松的科学生涯开始于研究微分方程及其在摆的运动和声学理论中的应用；
- 对积分理论、行星运动理论、热物理、电磁理论、位势理论和概率论都有重要贡献；
- 19 世纪概率统计领域里的卓越人物，改进了概率论的运用方法，特别是用于统计方面的方法，建立了描述随机现象的一种概率分布 泊松分布；
- 推广了“大数定律”，并导出了在概率论与数理方程中有重要应用的泊松积分。

泊松 (Poisson) 分布

- Poisson 分布是概率论中一种重要的离散型分布，它在理论与实践中都有广泛的应用，常与单位时间 (面积) 内上的计数过程相联系;
  - 一天内到达某商场的顾客数;
  - 单位时间内，电路受外界电磁波的冲击次数;
  - 一定时期内，某放射性物质放射出来的粒子数等
- 在二项分布中，当参数 n 较大时，计算二项概率 $b(k;n,p)$ 会非常麻烦.

定理 1.10 设有一列二项分布 $\{b(k;n,p_n)\}$, 若其参数列 p_n 满足

$$\lim_{n\to\infty} np_n = \lambda > 0,$$

则对任何非负整数 k 有

$$\lim_{n\to\infty} b(k;n,p_n) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

证明: 记 $\lambda_n := np_n$, 则

$$
\begin{aligned}
b(k;n,p_n) &= C_n^k p_n^k (1-p_n)^{n-k} = \frac{n!}{k!(n-k)!}(\frac{\lambda_n}{n})^k(1-\frac{\lambda_n}{n})^{n-k} \\
&= \frac{\lambda_n^k}{k!}(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})(1-\frac{\lambda_n}{n})^{n-k}
\end{aligned}
$$

注意到

$$
\lim_{n\to\infty}\lambda_n^k = \lambda^k,
$$

$$
\lim_{n\to\infty}(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n}) = 1
$$

$$
\lim_{n\to\infty}(1-\frac{\lambda_n}{n})^{n-k} = \lim_{n\to\infty} e^{(n-k)\ln(1-\frac{\lambda_n}{n})}
$$

$$
= e^{\lim_{n\to\infty}(n-k)\ln(1-\frac{\lambda_n}{n})} = e^{\lim_{n\to\infty}(n-k)(-\frac{\lambda_n}{n}+o(\frac{1}{n}))}
$$

$$
= e^{\lim_{n\to\infty}(-\lambda_n+\frac{k\lambda_n}{n}+(n-k)o(\frac{1}{n}))} = e^{-\lambda}
$$

从而 $$\lim_{n\to\infty} b(k;n,p_n) = \frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,2,\cdots$$

有了上述定理，当 n 很大而 p 很小时，可以用近似公式计算二项概率

$$b(k;n,p) \approx \frac{(np)^k}{k!}e^{-np}$$

这里我们要求 p 很小，为保证 n 很大时，乘积 np 有适度的大小.

定义 1.10 (泊松分布) 对参数 $\lambda > 0$, 记

$$p(k;\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0, 1, \cdots .$$

易见 $p(k;\lambda) > 0$ 且

$$\sum_{k=0}^{\infty} p(k;\lambda) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} = e^{\lambda}e^{-\lambda} = 1$$

即 $\{p(k;\lambda)\}$ 可以看成离散型分布的密度，我们就把它称为以 λ 为参数的泊松分布，记作 $P(\lambda)$.

泊松分布的性质

- 泊松分布列 $p(k;\lambda)$ 随 k 变化情况与二项分布相似，事实上，考虑比值

$$\frac{p(k;\lambda)}{p(k-1;\lambda)} = \frac{\lambda^k(k-1)!e^{-\lambda}}{\lambda^{k-1}k!e^{-\lambda}} = \frac{\lambda}{k}, k \geq 1$$

  - 当 $k < \lambda$ 时有，$p(k;\lambda) > p(k-1;\lambda)$;
  - 当 $k > \lambda$ 时有，$p(k;\lambda) < p(k-1;\lambda)$;
  - 因此 $p(k;\lambda)$ 随 k 先升后降，在 $m = [\lambda]$ 处达到最大值，而 λ 为整数时，$p(k;\lambda)$ 在 $m = \lambda, \lambda - 1$ 处同时取到最大值.
- 泊松分布在随机选择下的不变性：假设某块放射性物质在单位时间内发射出的粒子数 X 服从 $P(\lambda)$ 分布。而每个粒子被记录下来的概率为 p 即粒子有 $1-p$ 的概率被计数器遗漏，如果各粒子是否被记录相互独立，试求记录下的粒子数 Y 的分布.

$$
\begin{aligned}
P(Y=k) &= \sum_{n=0}^{\infty} P(X=n, Y=k) = \sum_{n=0}^{\infty} P(X=n)P(Y=k|X=n) \\
&= \sum_{n=k}^{\infty} P(X=n)P(Y=k|X=n) = \sum_{n=k}^{\infty} p(n;\lambda)B(k;n,p) \\
&= \sum_{n=k}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} C_n^k p^k (1-p)^{n-k} = \sum_{n=k}^{\infty} \frac{[\lambda(1-p)]^{n-k}}{(n-k)!} \frac{1}{k!} e^{-\lambda} (\lambda p)^k \\
&= \frac{1}{k!} e^{-\lambda p} (\lambda p)^k
\end{aligned}
$$

超几何分布

假设某个罐子中有 w 个白球和 b 个黑球，考虑如下两种取球方式

- 有放回的抓取 n 个球
  - $X := n$ 个球中白球的数量;
  - $X \sim B(n, \dfrac{w}{w+b})$
- 不放回的抓取 n 个球
  - $X := n$ 个球中白球的数量;
  - X 服从超几何分布，记为 $X \sim H(w, b, n)$;
  - 易知
    $$P(X = k) = \frac{C_w^k C_b^{n-k}}{C_{w+b}^n} := h(k, w, b, n), k = 0, 1, \cdots, r,$$
    其中 $r := \min\{w, n\}$.
  - 若要验证以上给出的确实为一个概率分布列，只需注意到下面的组合等式成立
    $$\sum_{k=0}^{r} C_w^k C_b^{n-k} = C_{w+b}^n$$

超几何分布：两个例子

- 不合格品抽检问题
  - 设有 N 件产品，其中有 M 件不合格品，从中不放回的抽取 n 件;
  - 令 $X := n$ 件抽取的产品中不合格品的件数;
  - $X \sim H(M, N-M, n)$
- 麋鹿的捕获 - 再捕获问题
  - 森林中共有 N 头麋鹿。某天，捕获了 M 头麋鹿，标记后将这 M 头麋鹿再放回野外.
  - 几天后，又重新随机地捕获 n 头麋鹿。假设重新捕获的麋鹿也同样可能是之前捕获的麋鹿.
  - 令 $X :=$ 再次被捕获的麋鹿的数量;
  - $X \sim H(M, N-M, n)$;
  - 第一次捕获的 M 头麋鹿相当于前面例子中白球的总数，第一次未捕获的 $N-M$ 头麋鹿相当于前面例子中黑球的总数，再次被捕获的 n 头麋鹿相当于前面例子中抽样的数量.

超几何分布的适用基础

- 除上面例子外，超几何分布还可出现在许多情况下;
- 超几何分布的适用基础是总体根据两套标签进行分类:
  - 在罐子的示例中，每个球不是白色就是黑色 (第一套标签);
  - 每个球要么是样本要么不是样本 (第二套标签);
- 两套标签中至少有一个是被完全随机分配的 (在罐子的例子中，球是随机抽样的);
- X 代表被两套标签都标记的数量：在罐子的例子中，关注的是既被抽样又是白色的球。
- X 服从超几何分布.

超几何分布之间的对称性

定理 1.11 $X \sim H(w, b, n), Y \sim H(n, w+b-n, w)$, 则 X 和 Y 是同分布的.

证明:

- 考虑一个由 w 个白球和 b 个黑球充满的罐子，现在随机不放回地从罐子里抓取 n 个球;
- 白球或者黑球看作第一套标签，有没有被抽取看作第二套标签;
- X 表示抽取的样本球中白球的数量 $\sim H(w, b, n)$;
- 若将是否被抽取看作第一套标签，白球或者黑球看作第二套标签;
- Y 表示所有白球中被抽中的数量 $\sim H(n, w+b-n, w)$;
- 显然 X 和 Y 都是表示被抽取的白球数量，所以它们有相同的分布;
- 也可以用代数的方法来检查 X 和 Y 是否具有相同的概率质量函数，即验证 $P(X=k) = P(Y=k)$.

定理 1.12 如果 $X\sim B(n,p), Y\sim B(m,p)$, 且 X 和 Y 相互独立，则当给定条件 $X+Y=r$ 时，X 的条件分布为超几何分布 $H(n,m,r)$.

定理 1.13 如果 $X\sim H(M,b,n)$, 且当 $N=M+b\to\infty$ 时 $p=\dfrac{M}{N}$ 保持不变，则 X 的分布列 (概率质量函数) 收敛到 $B(n,p)$ 的分布列.

超几何分布的二项近似

注意到

$$h(k;M,b,n)=\frac{C_M^kC_b^{n-k}}{C_{M+b}^n}=\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}$$

$$=\frac{\dfrac{M!}{k!(M-k)!}\dfrac{(N-M)!}{(n-k)!(N-M-(n-k))!}}{\dfrac{N!}{n!(N-n)!}}$$

$$=\frac{n!}{k!(n-k)!}\frac{M!}{(M-k)!}\frac{(N-M)!}{(N-M-(n-k))!}\frac{(N-n)!}{N!}$$

$$=\frac{n!}{k!(n-k)!}\frac{M(M-1)\cdots(M-(k-1))}{N(N-1)\cdots(N-(k-1))}\frac{(N-M)\cdots(N-M-(n-k)+1)}{(N-k)\cdots(N-n+1)}$$

$$=\frac{n!}{k!(n-k)!}\frac{\frac{M}{N}(\frac{M}{N}-\frac{1}{N})\cdots(\frac{M}{N}-\frac{(k-1)}{N})}{1(1-\frac{1}{N})\cdots(1-\frac{(k-1)}{N})}\frac{(1-\frac{M}{N})\cdots(1-\frac{M}{N}-\frac{(n-k)-1}{N})}{(1-\frac{k}{N})\cdots(1-\frac{n-1}{N})}$$

$$\to C_n^k\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}=b(k;n,p)$$

第十讲：常见的连续型分布

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

均匀分布

定义 1.11 任给参数 $a<b$, 函数

$$p(x)=\frac{1}{b-a},\quad a<x<b$$

满足密度函数的两个性质即:$p(x)\geq 0$ 且 $\int_{-\infty}^{\infty}p(x)dx=1$. 我们称以上式中的 $p(x)$ 为密度的连续型分布为区间 (a,b) 上的均匀分布，记作 $U(a,b)$.

- 易见，均匀分布的分布函数为

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

- 如果 X 服从 $U(a,b)$ 分布，则对任何 $a\leq x<y\leq b$ 有

$$P(x<X\leq y)=\int_x^y\frac{1}{b-a}dt=\frac{y-x}{b-a}$$


GU5672972S2
DEUTSCHE BUNDESBANK
Banknote
10
Carl Friedr. Gauß
1777–1855
ZEHN DEUTSCHE MARK
Deutsche Bundesbank
Frankfurt am Main
1. September 1999
GU5672972S2
ZEHN DEUTSCHE MARK
10
Banknote
Zehn
Deutsche Mark
10
DEUTSCHE BUNDESBANK

图 2: 高斯

- 卡尔。弗里德里希。高斯 1777 年出生于布伦瑞克;
- 11 岁时就发现了二项式定理;
- 19 岁发现正十七边形的尺规作图法，解决了困扰数学家们两千多年的难题;
- 1809 年提出“最小二乘法”并在此基础上建立的正态分布方程，是概率统计中一个非常重要的工具，广泛应用于数学、物理学等领域;
- 一生发表了 155 篇论文，对数论、代数学、非欧几何、复变函数和微分几何等领域都做出了开创性的贡献，被誉为数学王子.

正态分布 (高斯分布)

考虑下述函数

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

下面我们验证 $p_{\mu,\sigma}(x)$ 满足密度函数的两个性质即非负性与正则性：非负性显然；正则性则需要计算积分 $I := \int_{-\infty}^{\infty} p_{\mu,\sigma}(x)dx$.

由于 $p_{\mu,\sigma}(x)$ 的原函数不是初等函数，我们无法通过微积分基本公式计算 I. 我们考虑计算 I^2:

$$
\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy \\
&= \frac{1}{2\pi} \int_{0}^{2\pi} \int_{0}^{\infty} e^{-\frac{r^2}{2}} rdrd\theta \\
&= 1
\end{aligned}
$$

正态分布的定义

定义 1.12 若随机变量 X 的密度函数为

$$p_{\mu,\sigma}(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$

则称 X 服从参数为 (μ,σ^2) 的正态分布，记作 $X \sim N(\mu,\sigma^2)$. 特别的称 $N(0,1)$ 分布为标准正态分布，并简写 $p_{0,1}(x)$ 为 $\varphi(x)$. 正态分布又称高斯分布.

正态分布的性质

- 分布密度关于参数 μ 对称，即有
$$p_{\mu,\sigma}(\mu-x)=p_{\mu,\sigma}(\mu+x),\quad \forall x\in R$$
特别的，$N(0,1)$ 分布的密度函数为偶函数.
- $p_{\mu,\sigma}(x)$ 在 $x=\mu$ 处取得最大值 $\dfrac{1}{\sqrt{2\pi}\sigma}$. 注意到密度曲线下的面积应保持等于 1, 并且密度函数在 x 处的值反映了此分布取值 x 附近的概率大小。故 σ^2 越小，密度的曲线越尖陡，分布取值越集中；σ^2 越大，密度曲线越平缓，分布取值越分散.
- 正态分布的分布函数为
$$\Phi_{\mu,\sigma}(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}e^{-\dfrac{(t-\mu)^2}{2\sigma^2}}dt$$
特别的记标准正态分布函数为 $\Phi(x)$. 其与 $\Phi_{\mu,\sigma}(x)$ 有以下关系:
$$\Phi_{\mu,\sigma}(x)=\Phi(\frac{x-\mu}{\sigma})$$

例 1.14 (3σ 原则) 设随机变量 X 服从 $N(\mu,\sigma^2)$ 分布，求 $P(|X-\mu|<3\sigma)$.

解: 注意到

$$
\begin{aligned}
P(|X-\mu|<3\sigma) &= P(\mu-3\sigma<X<\mu+3\sigma) \\
&= \Phi_{\mu,\sigma}(\mu+3\sigma)-\Phi_{\mu,\sigma}(\mu-3\sigma) \\
&= \Phi(3)-\Phi(-3)=\Phi(3)-(1-\Phi(3)) \\
&= 2\Phi(3)-1=0.9973
\end{aligned}
$$

这表明，尽管正态分布的取值范围是 $(-\infty,+\infty)$, 但是其取值落在区间 $(\mu-3\sigma,\mu+3\sigma)$ 的概率高达 99.73%

例 1.15 若 $X\sim N(\mu,\sigma^2)$, 则 $Y:=\dfrac{X-\mu}{\sigma}\sim N(0,1)$; 反之，若 $X\sim N(0,1)$, 则 $Y:=\sigma X+\mu\sim N(\mu,\sigma^2)$.

解: 我们仅给出第一种情况的证明。事实上，我们只需要说明 Y 的分布密度为

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

即可. 事实上，考虑 Y 的分布函数

$$\begin{aligned}
F_Y(x) &= P(Y\le x)=P\left(\frac{X-\mu}{\sigma}\le x\right)=P(X\le \sigma x+\mu)\\
&= \int_{-\infty}^{\sigma x+\mu}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy
\end{aligned}$$

从而 Y 的分布密度为

$$p_Y(x)=F_Y'(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}=\varphi(x)$$

正态分布函数的简单性质

记 $\Phi(x)$ 为标准正态分布的分布函数，$U \sim N(0,1)$, $\Phi_{\mu,\sigma}(x)$ 为 $N(\mu,\sigma^2)$ 分布的分布函数，$X \sim N(\mu,\sigma^2)$, 则

- $\Phi(0) = \dfrac{1}{2}$;
- $\Phi_{\mu,\sigma}(\mu) = \dfrac{1}{2}$;
- $\Phi(-x) = 1 - \Phi(x)$;
- $P(|U| < x) = P(-x < U < x) = P(U < x) - P(U \le -x)$
  $= \Phi(x) - (1 - \Phi(x)) = 2\Phi(x) - 1$;
- $\Phi_{\mu,\sigma}(x) = P(X \le x) = P(\dfrac{X-\mu}{\sigma} \le \dfrac{x-\mu}{\sigma}) = \Phi(\dfrac{x-\mu}{\sigma})$
- $P(a < X \le b) = \Phi(\dfrac{b-\mu}{\sigma}) - \Phi(\dfrac{a-\mu}{\sigma})$.

定义 1.13 称以下函数

$$\Gamma(\alpha) := \int_0^{\infty} x^{\alpha-1}e^{-x}dx, \quad \alpha > 0,$$

为 Gamma 函数. Gamma 函数有以下性质:

- $\Gamma(1) = \int_0^{\infty} x^{1-1}e^{-x}dx = 1$;
- $\Gamma(\frac{1}{2}) = \int_0^{\infty} x^{\frac{1}{2}-1}e^{-x}dx = 2\sqrt{\pi}\int_0^{\infty} \frac{1}{\sqrt{2\pi}}e^{-x}d\sqrt{2}x^{\frac{1}{2}}$
  $= 2\sqrt{\pi}\int_0^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt = \sqrt{\pi}$;
- $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, 故 $\Gamma(n+1) = n\Gamma(n) = \cdots = n!$; 事实上,

$$\begin{aligned}\Gamma(\alpha+1) &= \int_0^{\infty} x^{\alpha}e^{-x}dx = \int_0^{\infty} -x^{\alpha}de^{-x} \\ &= -x^{\alpha}e^{-x}|_0^{\infty} + \int_0^{\infty} \alpha x^{\alpha-1}e^{-x}dx = \alpha\Gamma(\alpha)\end{aligned}$$

伽玛分布

定义 1.14 (Gamma 分布) 若随机变量 X 的密度函数为

$$p(x)=\begin{cases}\dfrac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases}$$

则称 X 服从参数为 (α,λ) 的 Gamma 分布，记作 $X\sim\Gamma(\alpha,\lambda)$.

指数分布与 χ^2(卡方) 分布

定义 1.15 称 $\Gamma(1,\lambda)$ 分布为指数分布，其密度函数与分布函数分别为

$$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0,\\ 0, & x<0.\end{cases}\qquad F(x)=\begin{cases}1-e^{-\lambda x}, & x\ge 0,\\ 0, & x<0.\end{cases}$$

定义 1.16 称 $\Gamma(\frac{n}{2},\frac{1}{2})$ 分布为 χ^2 分布，记作 $\chi^2(n)$, 其密度函数为

$$p(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x\ge 0,\\ 0, & x<0.\end{cases}$$

例 1.16 假定有一个于随机时刻陆续到来的质点流，我们若以 $N(t)$ 表示在 $[0,t]$ 时间内到来的质点的个数，并假设其服从参数为 λt 的泊松分布。试证明第 n 个质点到达的时间 S_n 服从 $\Gamma(n,\lambda)$ 分布.

证明: 首先我们考虑一下 S_n 的分布函数

$$\begin{aligned} P(S_n \le t) &= P(N(t) \ge n) = \sum_{k=n}^{\infty} P(N(t)=k) \\ &= \sum_{k=n}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \end{aligned}$$

故其密度 $p(t)$ 为

$$\begin{aligned} p(t) = F'(t) &= \sum_{k=n}^{\infty} \left[\frac{\lambda(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} - \frac{\lambda(\lambda t)^k}{k!} e^{-\lambda t} \right] \\ &= \frac{\lambda(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t} \end{aligned}$$