概率

主讲人：刘宏志

liuhz@ss.pku.edu.cn

概率论

- 研究对象：
  - 结果具有随机性的现象
- 典型应用：
  - 遗传学中构建基因突变模型
  - 运筹学中建立货物存货需求的随机化模型
  - 保险公司使用的精算科学
  - 概率论是金融学的奠基石

……

样本空间

- 随机试验
  - 随机现象的实现和对它的观察
  - 例：抛一颗骰子，观察出现的点数
- 随机事件
  - 随机试验的每一可能结果称为一个基本事件
  - 一个或一组基本事件统称随机事件，或简称事件
- 样本空间
  - 一个随机试验的所有可能试验结果组成的集合称为该随机试验的样本空间，记为Ω，其元素记为ω
  - 样本空间又称为基本事件空间
  - 例：抛一颗骰子，出现点数的样本空间为

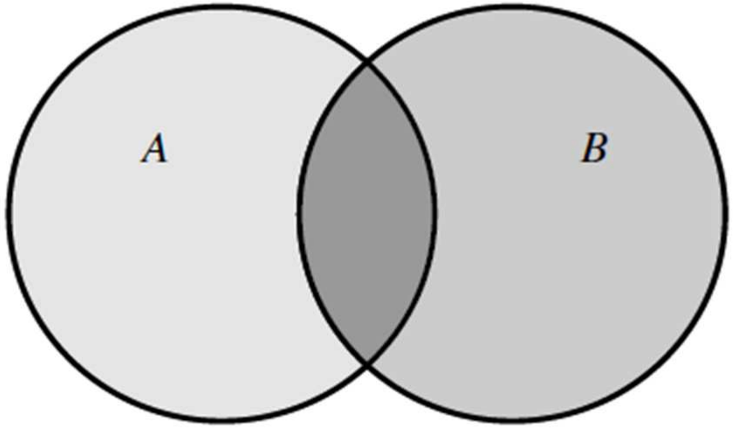
  $\Omega=\{1, 2, 3, 4, 5, 6\}$

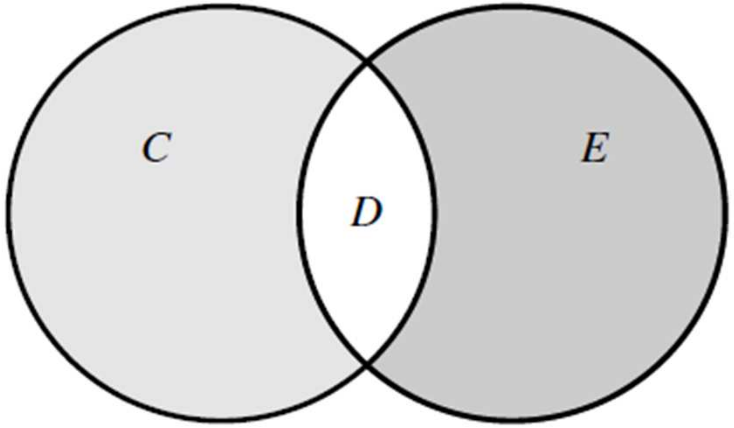
概率测度

- 概率
  - 对随机事件发生的可能性的度量
  - 定义在样本空间Ω子集上的实函数
- 公理
  - $P(\Omega) = 1$
  - 如果$A \subset \Omega$，则$P(A) \geq 0$
  - 如果 $A \cap B = \varnothing$，则$P(A \cup B) = P(A) + P(B)$

概率测度的基本性质

- $P(A^c) = 1 - P(A)$
  - ➢因为：$A \cap A^c = \varnothing$ 且$A \cup A^c = \Omega$
- $P(\varnothing) = 0$
  - ➢因为：$\varnothing = \Omega^c$且$P(\Omega) = 1$
- $A \subset B \Rightarrow P(A) < P(B)$
  - ➢因为：$B = A + (A^c \cap B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$





案例：AIDS感染的统计风险

- 研究表明：平均来看，单次与AIDS携带者进行没有保护措施的性行为后被感染的风险为五百分之一。
- 如果与AIDS携带者进行100次性行为，感染的风险增加到五分之一。
- 从统计上来看，与同一感染者进行500次性行为将导致百分之百的感染率。

-- 《洛杉矶时报》(1987.8.24)

结论是否正确？ 该如何计算概率？

北京大学

概率计算：计数法

- 有限样本空间：$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$
- 事件A发生的概率 = A包含的基本事件ω_i发生概率的加和
- 例：一次掷色子的点数小于3的概率是多少？
  - $\Omega = \{1,2,3,4,5,6\}$
  - $A = \{1,2\}$，$P(A)=1/3$
- 若Ω中的元素具有等概率性，事件A通过多个互斥途径中的任一种方式方法，则

$$P(A)=\frac{\text{导致}A\text{发生的方式个数}}{\text{所有试验结果个数}}$$

注意：此公式仅当所有实验结果是等可能发生时才成立

案例：抽奖游戏

- 游戏1
  - 一黑盒中装有5个红球和6个绿球，
  - 一白盒中装有3个红球和4个绿球。
  - 允许你选择一个盒子并随机从中选一个球，
  - 如果选中红球，则会有一份奖品。
  - 你会选择从哪个盒子中取球？
- 游戏2
  - 一个黑盒中装有6个红球和3个绿球，
  - 一个白盒中装有9个红球和5个绿球。
  - 规则不变，你会选择从哪个盒子中取球？
- 游戏3
  - 将游戏2中黑盒和白盒中球倒入游戏1中对应颜色的盒中。
  - 规则不变，你会选择从哪个盒子中取球？

辛普森悖论：在某个条件下的两组数据，分别讨论时都会满足某种性质，可是一旦合并考虑，却可能导致相反的结论。

案例：生日问题

- 一个30人的班里，至少有两人生日在相同的概率是多少？
- 设：至少有两个人的生日在同一天的事件用A表示
- 样本空间的可能结果有365^{30}种
- A发生有很多种方式，而A^c发生的方式相对简单
- 先计算A^c（所有人生日都不相同）的概率：
  - 第一个人的生日是 365选365
  - 第二个人的生日是 365选364

    ⋮
  - 第30个人的生日是 365选365-(30-1)
  - $P(A^c)=\frac{365}{365}\frac{364}{365}\cdots\frac{336}{365}=0.29$

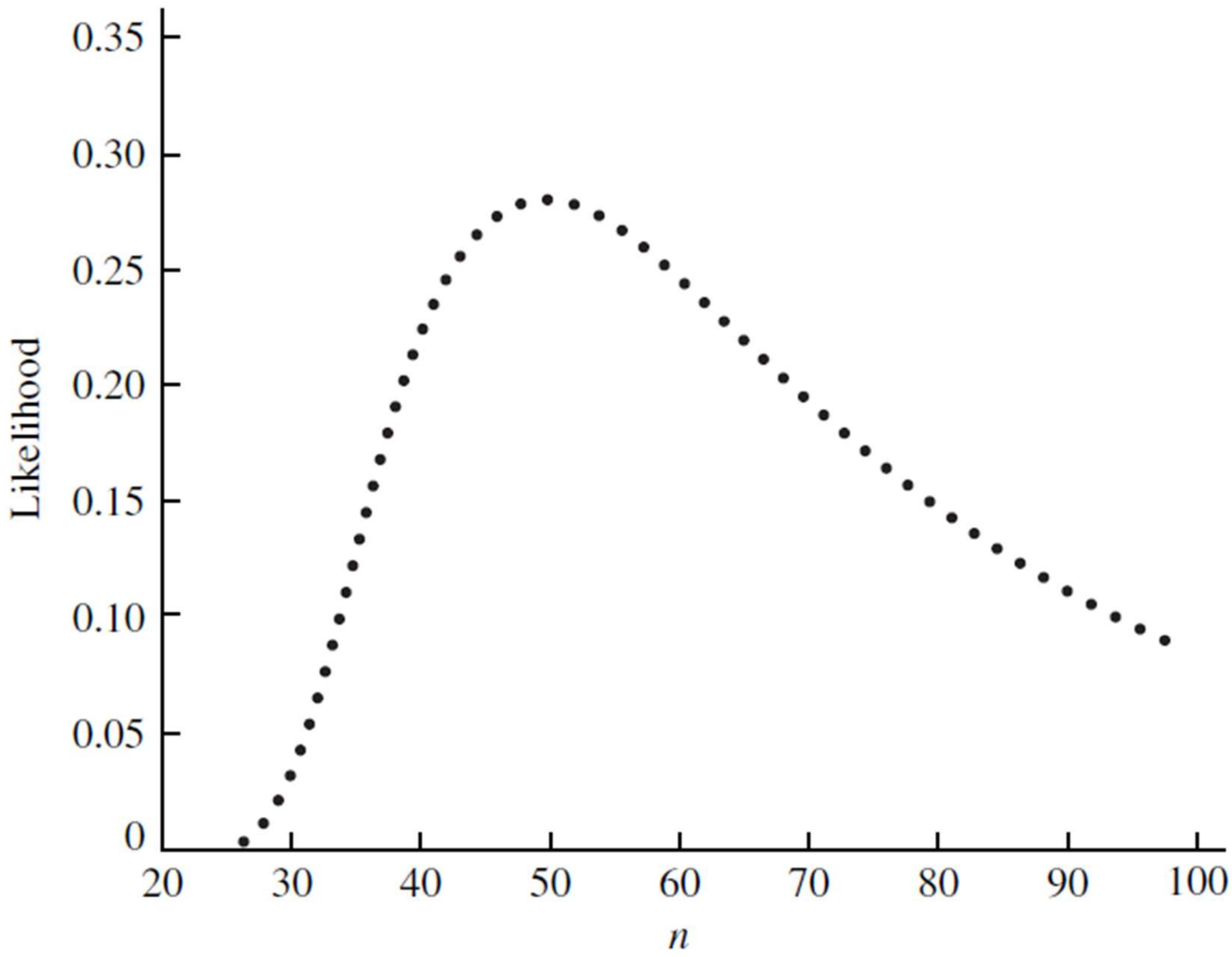
案例：野生动物总数估算

- 假设捕捉10个动物，将它们做上标记后释放。这之后，再捕捉20个动物，发现有4个带有标记。动物的总数是多少？
- 设动物总数为n，则第2次捕捉中含4个带标记动物的概率为：

$$\frac{\binom{10}{4}\binom{n-10}{16}}{\binom{n}{20}}$$

- 观察结果发生的概率是待估参数n的函数，称为似然
- 最大似然估计：将使观测结果出现可能性最大的n作为估计值

标记重捕法：假设（1）调查期间数量稳定；（2）标记个体均匀分布在全部个体之中；（3）标记操作不影响动物的行为和死亡。



$$L_n = \frac{\binom{10}{4}\binom{n-10}{16}}{\binom{n}{20}}$$

条件概率

- 事件A在另一事件B已经发生条件下发生的概率

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

- 乘法定律：$P(A \cap B) = P(A|B)P(B)$
- 全概率定律：

令$B_1, B_2, \ldots, B_n$满足$\bigcup_{i=1}^{n} B_i = \Omega, B_i \cap B_j = \emptyset, i \neq j$, 且对所有的$i$，$P(B_i)>0$。则对任意A，

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

条件概率

- 贝叶斯公式：

  令A和$B_1, B_2, \ldots, B_n$是事件，其中$\bigcup_{i=1}^{n} B_i = \Omega$, $B_i \cap B_j = \emptyset, i \neq j$, 且对所有的$i$，$P(B_i)>0$。则

$$P\left(B_j \middle| A\right) = \frac{P(A|B_j)P(B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)\,P(B_i)}$$

案例：测谎仪

- 常用于例行管理具有敏感职位的员工或准员工
- 事件 +表示测谎仪显示积极信号，预示受测者撒谎
- T表示受测者说的是真话
- L表示受测者说的是假话
- 据测谎可靠性的研究：$P(+|L)=0.88$, $P(-|T)=0.86$
- 假设：在例行检查时，对某些特定问题，大部分人没有撒谎的理由，因此$P(T)=0.99$，而$P(L)=0.01$
- $P(L|+)=\dfrac{P(+|L)P(L)}{P(+|L)P(L)+P(+|T)P(T)}$

$$=\frac{(0.88)(0.01)}{(0.88)(0.01)+(0.14)(0.99)}$$

$$=0.06$$

事件独立性

- 两事件相互独立：
  - A，B是两事件，P(AB)=P(A)P(B)
- 两两独立：
  - A,B,C三个事件，P(AB)=P(A)P(B)，P(AC)=P(A)P(C)，P(BC)=P(B)P(C)
- 事件集相互独立：
  - 事件集$A_1,A_2,\ldots,A_n$,任意事件子集$A_{i1},A_{i2},\ldots,A_{im}$,满足$P(A_{i1}A_{i2}\ldots A_{im})=P(A_{i1})P(A_{i2})\ldots P(A_{im})$

案例：AIDS感染的统计风险

- 研究表明：平均来看，单次与AIDS携带者进行没有保护措施的性行为后被感染的风险为五百分之一。
- 如果与AIDS携带者进行100次性行为，感染的风险增加到五分之一。
- 从统计上来看，与同一感染者进行500次性行为将导致百分之百的感染率。

--《洛杉矶时报》(1987.8.24)

令C_i表示第i次性行为没有发生病毒传染的事件：

$$P(C_1C_2\ldots C_{100})=(1-1/500)^{100}=0.82$$

$$P(C_1C_2\ldots C_{500})=(1-1/500)^{500}=0.37$$