概率论的四大基石之一：随机

- 随机 Random，不可预测
- 理想随机和伪随机

随机性和不确定性？可能出现的选项是否可知

随机性：可选择选项可知，灰犀牛，结果可知，不知道何时

不确定性：无法预知，黑天鹅事件

所以概率论讨论的是随机性，不是不确定性

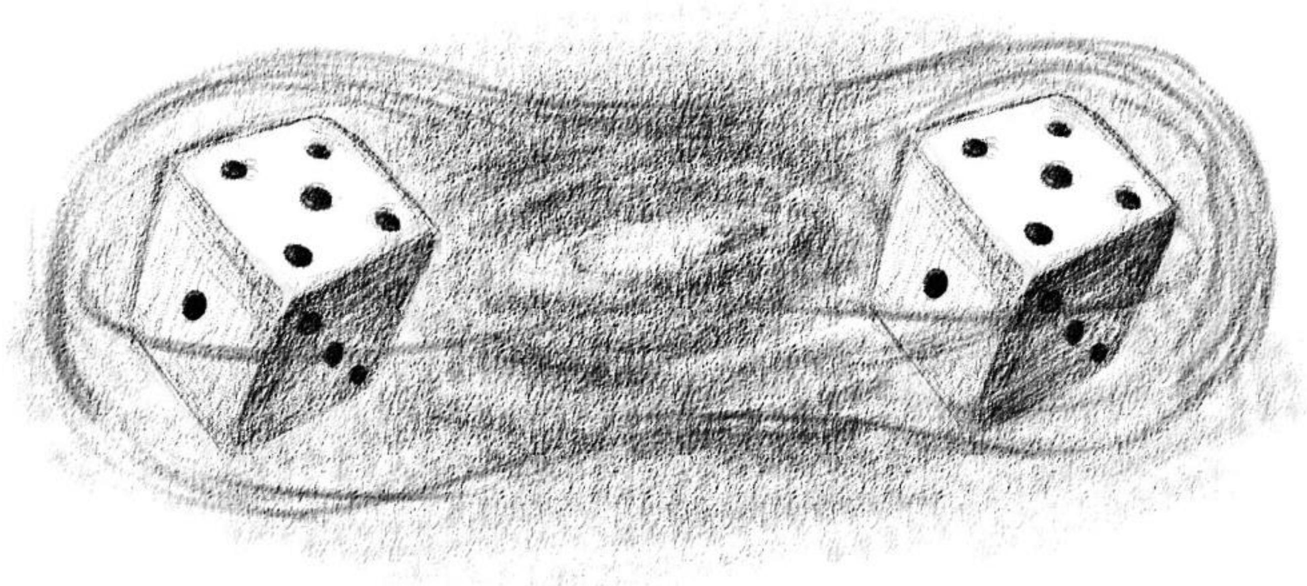
两者可以转化

今天会遇到谁？ → 今天遇到的人我认识不认识？

随机

- 经典随机：掷骰子

- 量子随机：隐函数理论？纠缠的骰子

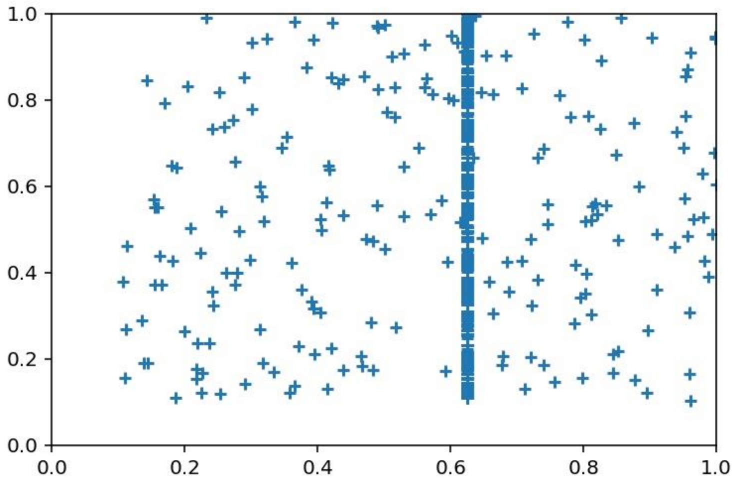


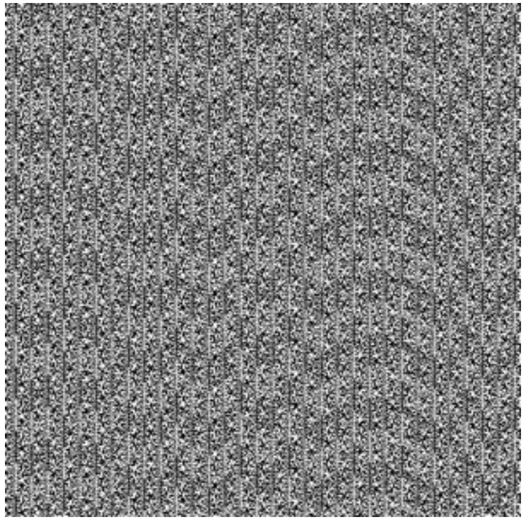
生成伪随机数

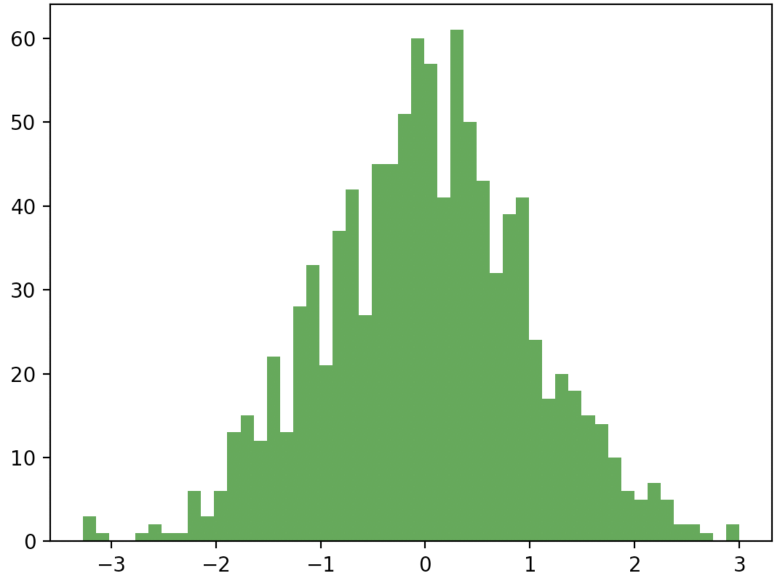
```c
/* 使用 ANSI C 可移植算法 */
static unsigned long int next = 1; // 种子
int rand(void) // 生成伪随机数
{
 next = next * 1103515245 + 12345;
return (unsigned int) (next / 65536) % 32768;
 }

 void srand(unsigned int seed) // 修改种子
{ next = seed; }
```

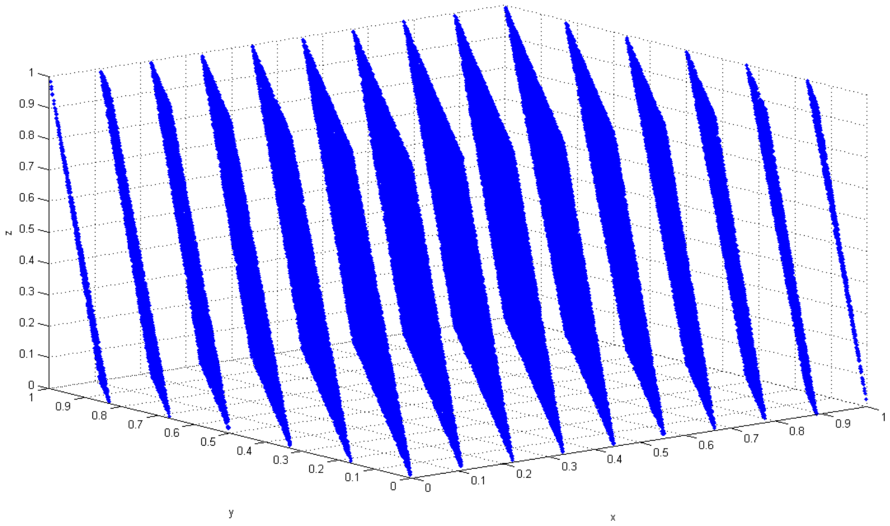
- 线性同余方法
- 平方取中法
- M-sequence
- 梅森旋转算法
- 伪随机数二进制数列

伪随机函数的概率分布





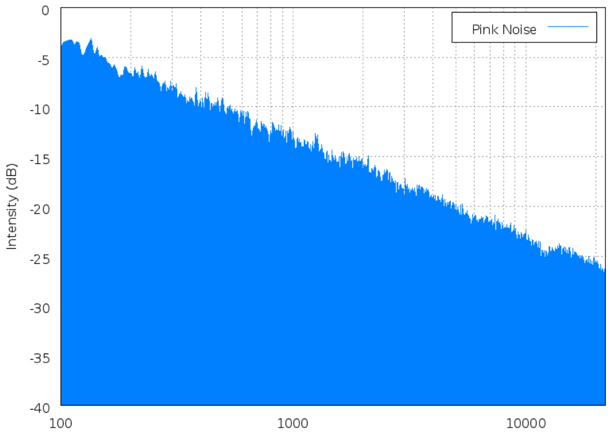


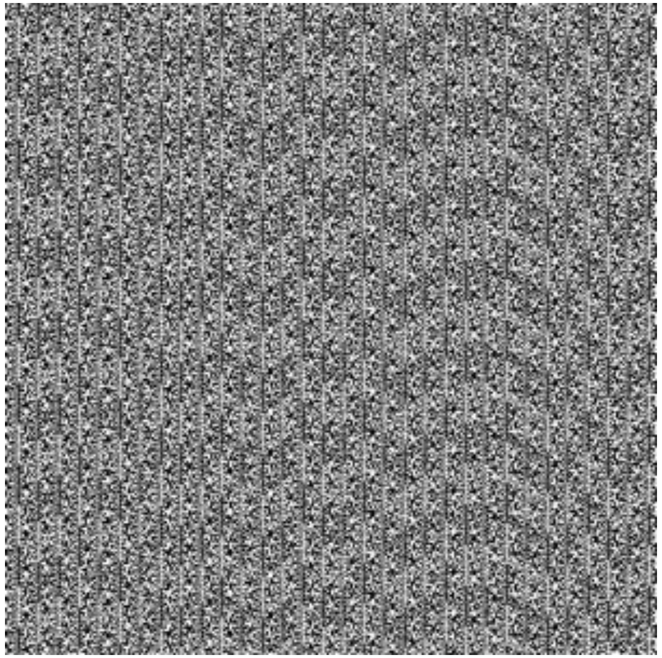


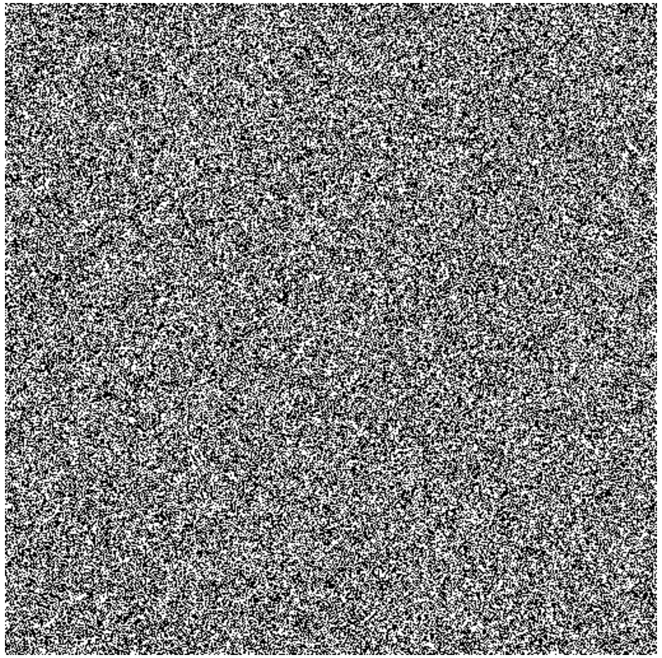
其他随机来源

比如掷钱币、骰子、转轮、使用电子元件的噪音、核裂变等等，这样的随机数发生器叫做物理性随机数发生器，它们的缺点是技术要求比较高。

- 白噪声
- 粉噪声
- Shot noise







概率论基石之二：概率

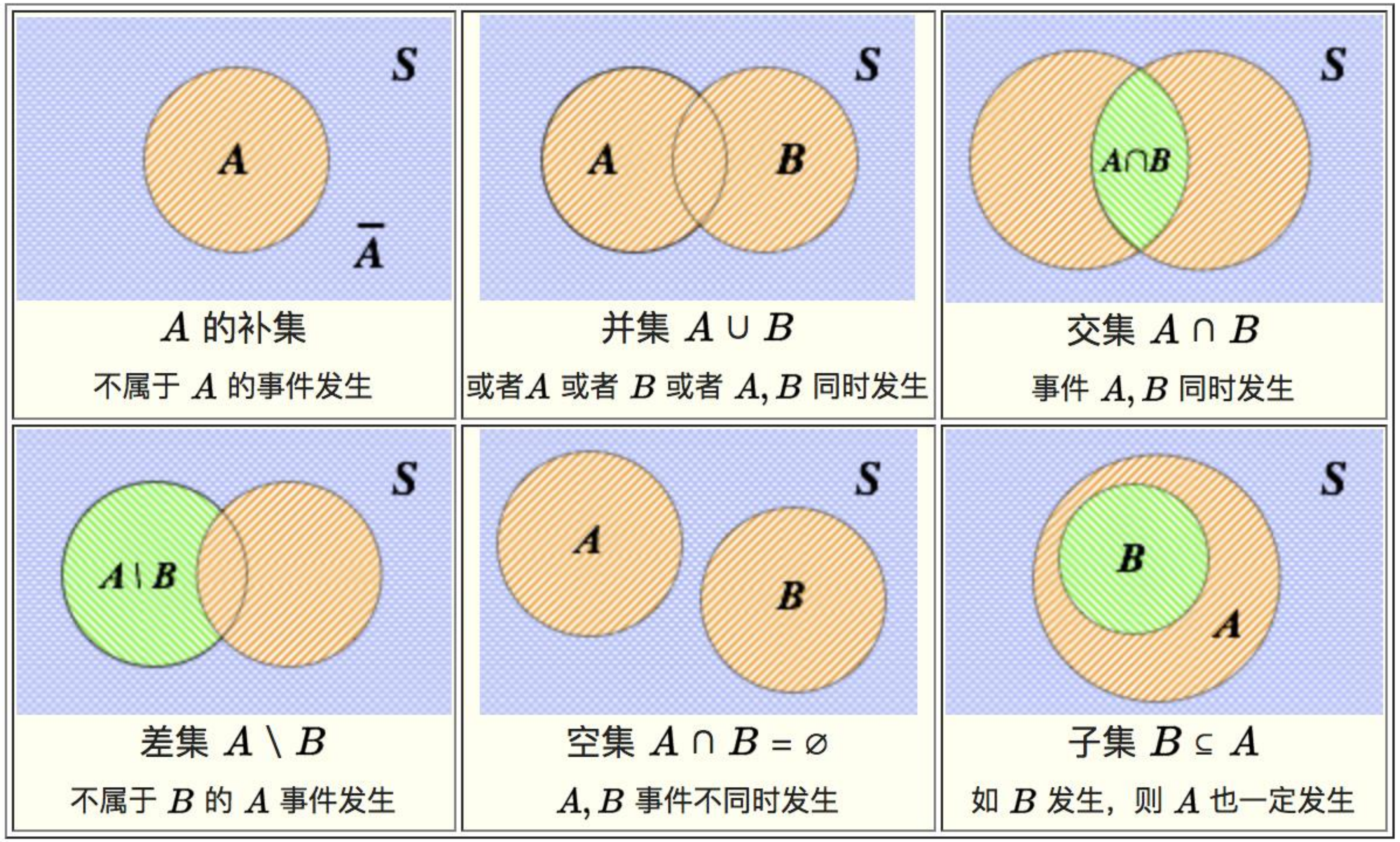
- 定义：设*E*是随机试验，*S*是它的样本空间，对于*E*的每一事件A赋予一个实数，记为*P(A)*，称为事件 *A* 的概率

概率是随机事件发生可能性大小的定量描述：对于任何事情，只要设定一个条件，从可能性的角度出发，对某一个发生结果进行陈述，就可以转化为随机事件，然后度量概率。

1. 限定一个条件
2. 可能性：要不没发生，要不不知道
3. 对发生结果的陈述：必须是随机结果，而不是不确定性结果

概率是随机事件在样本空间的比率

- Venn图



概率的完备性有可能不能被满足！对世界的认识，是对样本空间完备性的认识

概率公理（Kolmogorov公理，柯尔莫果洛夫公理）

第一公理(非负性)
对于任意一个集合$A \in \xi$，即对于任意的事件
$P(A) \geq 0$ 即，任一事件的概率都可以用0到1区间上的一个实数来表示。

第二公理(归一化)

$$P(\Omega) = 1$$

第三公理(可加性)
任意两两不相交事件$E_1, E_2, ..$的可数序列满足
$P(E_1 \cup E_2 \cup \cdots) = P(E_i)$。
不相交子集的并的事件集合的概率为那些子集的概率的和。这也被称为是σ可加性。如果存在子集间的重叠，这一关系不成立。

概率论引理

从柯尔莫果洛夫公理可以推导出另外一些对计算概率有用的法则。
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$，
$P(Q - E) = 1 - P(E)$，
$P(A \cap B) = P(A) \cdot P(B|A)$，
这一关系给出了贝叶斯定理。以此可以得出A和B是独立的当且仅当

$$P(A \cap B) = P(A) \cdot P(B)$$

概率论四大基石之三：独立性

- 随机事件之间的相互关系
- 独立性：随机事件之间没有任何关系，一个随机事件对另外一个事件没有影响

赌场的彷徨：

赌徒谬误：连着很多次，总该轮着了吧，“运气用完！”

热手谬误：既然已经这么多次，这次也该，“运气！”

双黄蛋：事情是独立的吗？

连着6个鸡蛋都是双黄蛋，1000个鸡蛋里面有一个双黄蛋，则这样的事情几乎不会发生，但实际上发生过，而且连着几次。

看似独立，但有其它原因作为解释：

1. 鸡的年龄和批次
2. 大鸡蛋趋向于放在一起
3. 激素

我们有可能对“独立”事件是不够了解的，困难如同完备性。

两两独立和多组分完全独立

两两独立: 如果事件U是事件总集S中的子集，当且仅当U里的事件A、B满足

P(A∩B)=P(A) x P(B)

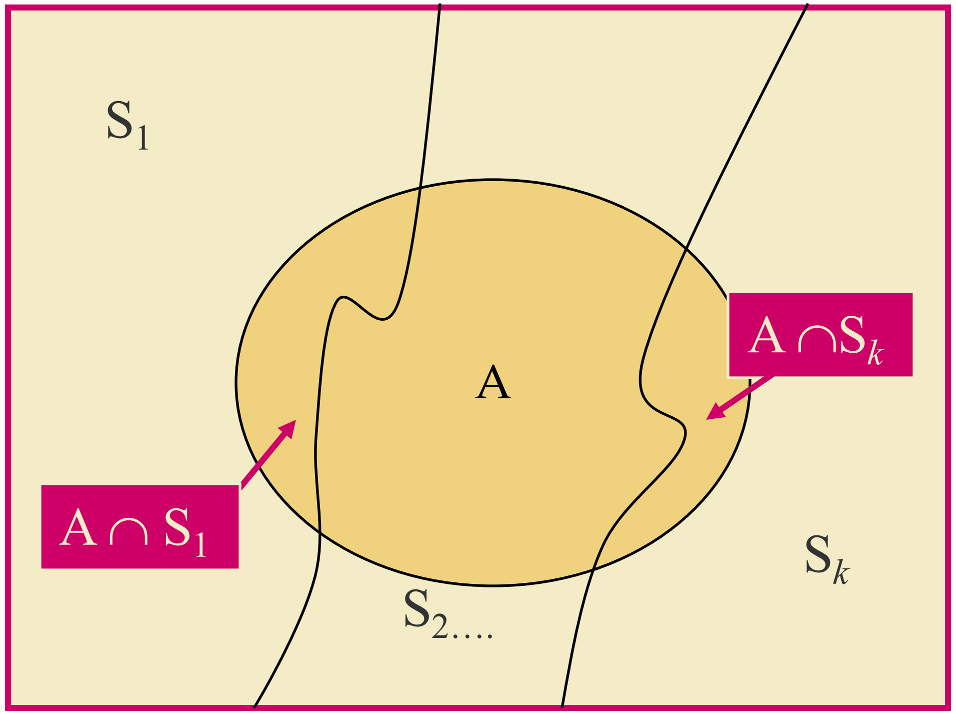
则事件集U两两独立。

多组分完全独立: 一个事件集，当且仅当它的每个事件发生概率为各个事件发生概率的乘积，则这个事件集U是多组分完全独立:

$$P(A_{i1} \cap A_{i2}, \dots, \cap A_{ik}) = \prod_{j=1}^{k} P(A_{ij})$$

完全概率定律

如果一个事件空间可以被分割为独立的事件组份



Let S_1 , S_2 , S_3 ,..., S_k 是互斥的事件，并且是所有事件可能，则某一事件A的发生概率

$$\mathbf{P(A) = P(A \cap S_1) + P(A \cap S_2) + \ldots + P(A \cap S_k)}$$

$$\mathbf{= P(S_1)P(A|S_1) + P(S_2)P(A|S_2) + \ldots + P(S_k)P(A|S_k)}$$

$$=\sum_{i=1}^{k} P(S_i)P(A|S_i)$$

加法法则

一个随机事件发生或者另外一个随机事件发生的概率，也就是两者发生其一的概率等
两个随机事件各自发生的和。前提：
两个事件必须是互斥的

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

若A和B是互不相容事件，则变为

$$P(A\cup B)=P(A)+P(B)$$

乘法规则

如果事件有A和事件B

$$P(A\ \text{and}\ B)=P(A\cap B)=P(A)\cdot P(B/A)$$

如果事件A和事件B是相互独立的

$$P(A\ \text{and}\ B)=P(A\cap B)=P(A)\cdot P(B)$$

概率计算总结

事件	概率
A	$P(A) \in [0,1]$
非A	$P(A^c) = 1 - P(A)$
A或B	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $P(A \cup B) = P(A) + P(B) \qquad \text{if A and B are mutually exclusive}$
A和B	$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$ $P(A \cap B) = P(A)P(B) \qquad \text{if A and B are independent}$
B的情况下A的概率	$P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(B \mid A)P(A)}{P(B)}$

大数定理—概率论的黄金定律

雅各布.伯努利Jakob I. Bernoulli，1654年－1705年

随着试验数据的不断积累，频率和概率的差距会越来越小。只要重复实验或者观测的数据足够大，随机事件发生的频率会无限接近它的概率。

--在相同环境、重复实验的条件下，用历史数据预测未来是可行的。

问题：足够多是多少？

概率是对发生频率的计算

弱大数定理：一个随机事件的发生，是**存在**一个真实的、客观的概率的，是可以通过大量实验和统计无限接近这个真实、客观的概率的。

硬币扔了1000次，524次H，476次T，那么硬币5/5开，你能证明硬币均匀吗？

记得原子的精细常数？真的不一样有可能潜伏在很高很高的精度上

大数定律告诉我们，数据越多，频率就会越接近概率。
在真实概率上下浮动。这种浮动的范围就是“**精度误差**”。
抛硬币的结果，并不是刚好等于理想的50%，这个和理想值之间的差距，就是精度误差。
假如你抛出正面朝上的频率是在47%-53%之间，那么精度误差就是± 3%。
针对这±3%的误差率，100组试验样本，如果有95组样本算出来的频率，正好在这个精度误差的范围之内，我们就称之为95%的“**置信度**”。
比如，99.9%的置信度和2%的精度误差，就可以把重复的次数从无限降低到7000次左右；如果把置信度下降到95%，重复次数可以降低到2500次左右；如果再放宽点标准，把精度误差从2%变成3%，试验次数可以下降到1000次左右。

强大数定理

整体概率稳定的可能性很大和一定稳定还是有些差别的。只有一定、100%的稳定，才是真正的确定性。

1930年代苏联数学家、概率论的先驱柯尔莫哥洛夫(Kolmogorov)在雅各布的基础上，做出了更加严密的证明，也就是“强大数定律”。

随着数据越来越多，频率接近概率不仅是可能性越来越大，而是一定的。

也就是说，随着数据越来越多，频率最终**一定会**接近真实概率。

到此为止，我们先用**弱大数定律找到了整体**，又用**强大数定律确定了整体一定是稳定的**。大数定律又被称为“黄金定理”，它让我们真正能用整体的确定性来对抗局部的随机性。

数学期望

- 数学期望是对随机结果长期价值的数值化衡量，是判断一件事值不值得做的整体量化指标
- 对随机事件的不同结果的概率加权求平均，对事件长期价值的数字化衡量
- 计算数学期望要把所有的随机结果数值化，只有赋予每个结果一个具体值，才能进行数学期望的计算
- 对一同一个结果，个体的数学期望值也可能是不一样的，计算时需要加入个体对价值的主观考量

魔球理论

某队员：篮下2分 55%命中，中距离2分，45%命中，远距离3分35%命中

期望数值化?

篮下期望值：0.55 X 2 = 1.1

中距期望值：0.45 X 2 = 0.9

远距离期望值：0.35 X 3 = 1.05

所以该队员尽量打篮下

对所有队员分析得分能力，可以量化最佳组合

数学期望不能完成描述一个随机事件

- 投资方案一：收益稳定，100%挣5万；
- 投资方案二：收益不稳定，50%机会赚20万，另外50%可能赔10万

方差来表达涨落

$$\sum (x_i - \bar{x})^2$$

赌博游戏对方差的利用

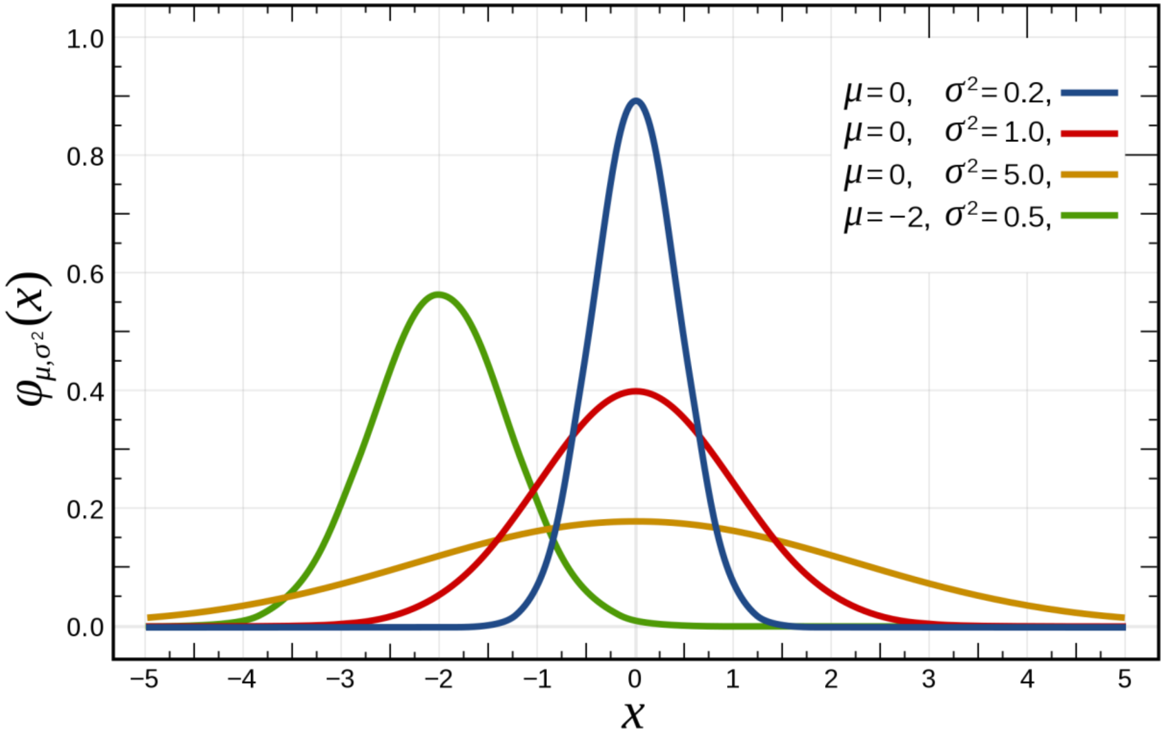
现实世界的概率模型

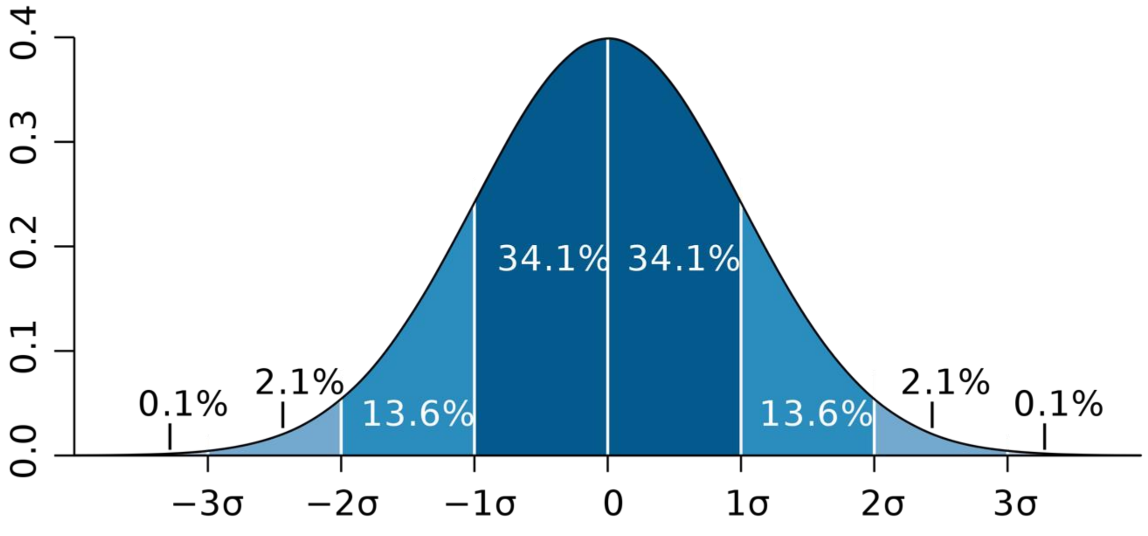
- 正（常）态分布

normal distribution

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

期望值就是平均值：表明好坏
极端之很少：小概率小影响
标准差决定胖瘦：表明波动





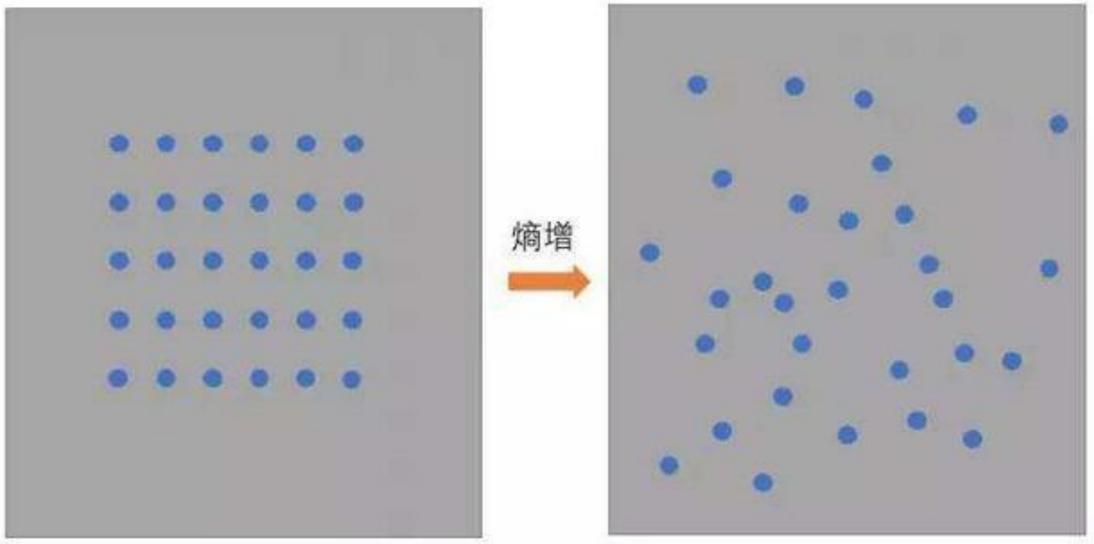
拉普拉斯中央极限定理

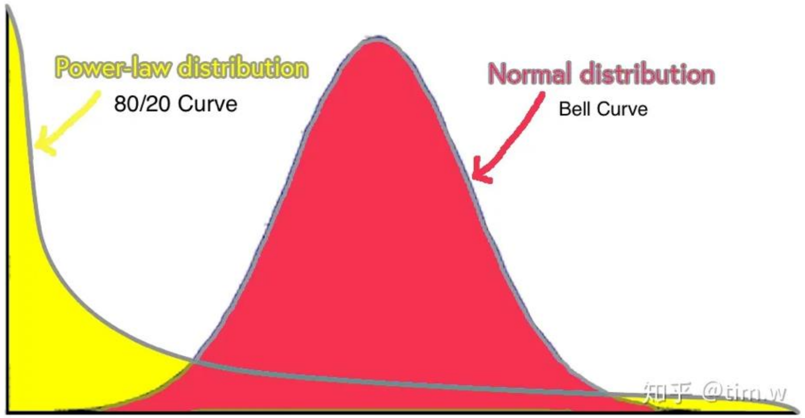
在特定条件下，大量统计独立**的随机变量的平均值的分布趋于正态分布**。

中央极限定理的重要意义在于，根据这一定理的结论，其他概率分布可以用正态分布作为近似。

- 正态分布普遍存在
- 所有分布最终都会演化成正态分布

熵增加原理：正态分布意味着平衡态





$$H(x) = -\sum_{i=1}^{n} p_i \mathrm{log} p_i$$

$$0 \le p_i \le 1, i = 1,2, \dots, n$$

$$\sum_{i=1}^{n} p_i = 1$$

香农熵：对正态分布的偏离程度

A=00，B=01，C=10，D=11。平均2比特

假设字母 A、B、C、D 的出现概率分别是 1/2、1/4、1/8 和 1/8。

将字母 A 编码为单个比特 0，对于字母 B，持其编码仍为 1对于字母 C 和 D，如 10 和 11。

字母 A：1/2×1 比特 = 1/2 比特

字母 B：1/4×1 比特 = 1/4 比特

字母 C：1/8×2 比特 = 1/4 比特

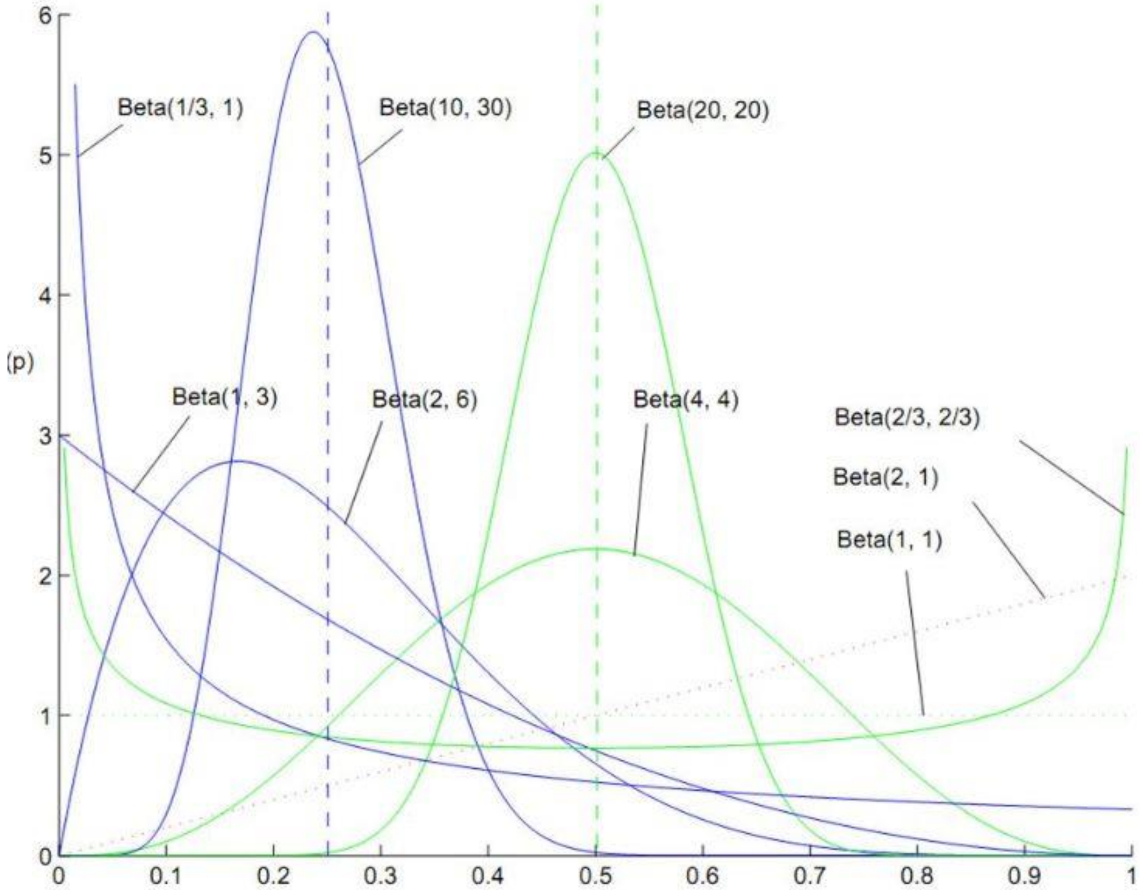
字母 D：1/8×2 比特 = 1/4 比特

平均1.25比特

$$H(x) = -\sum_{i=1}^{n} p_i \log p_i$$

常用的概率分布

- 均匀分布
- 伯努利分布
- 二项分布
- 高斯分布
- 拉普拉斯分布
- 泊松分布
- 幂律分布
- 伽马分布
- 贝塔分布
- 狄拉克分布
- 多项式分布与狄里克雷分布



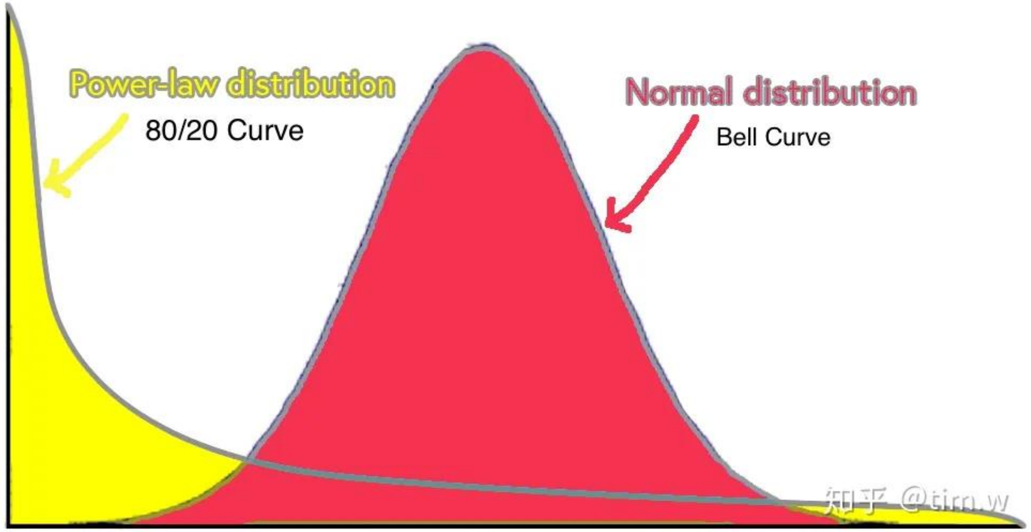
$$p(X,\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}X^{\alpha-1}(1-X)^{\beta-1}=\frac{1}{B(\alpha,\beta)}X^{\alpha-1}(1-X)^{\beta-1},\quad 0\le X<1$$

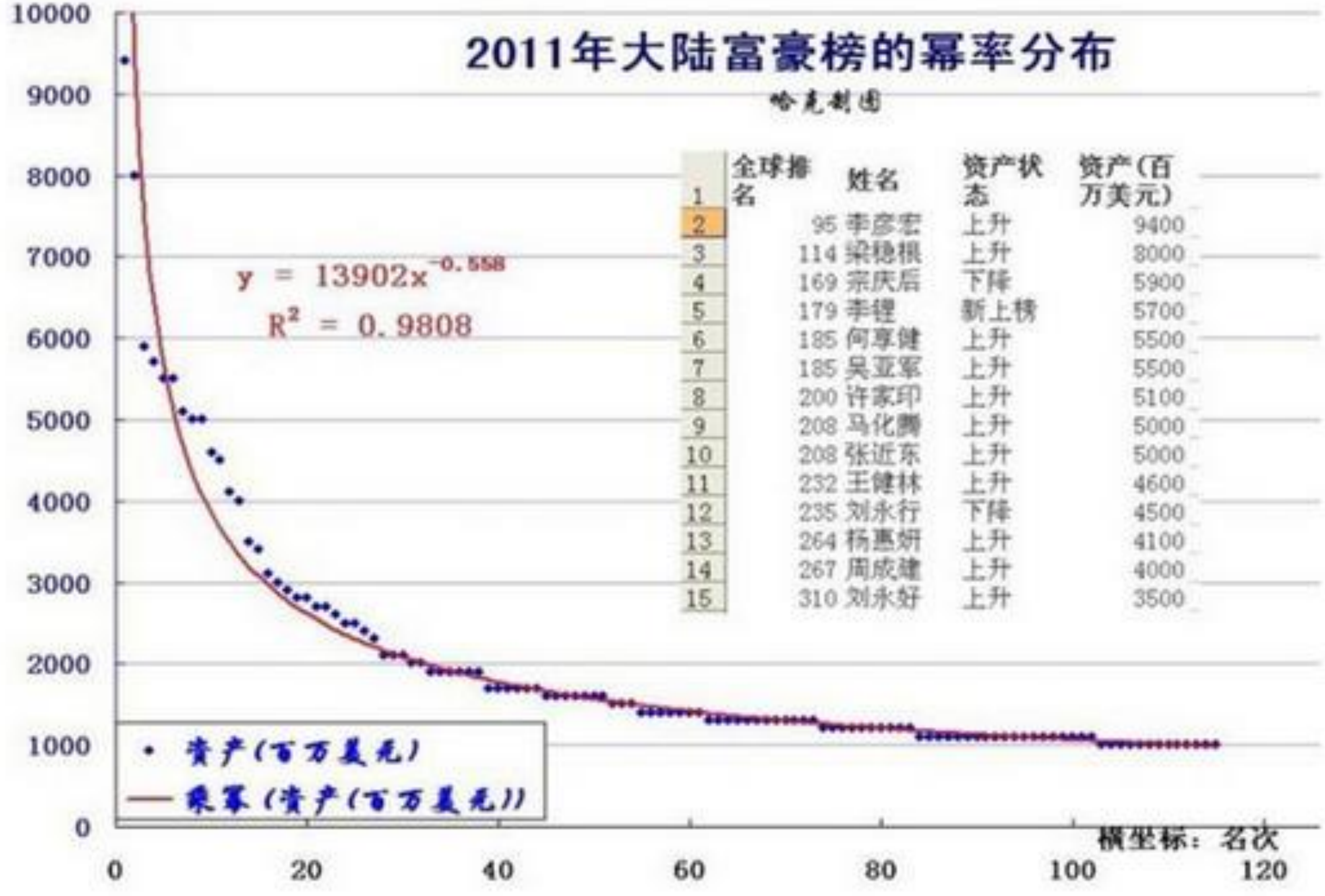
幂律分布：二八法则

• 百分之八十的生意来自百分之二十的消费者

大偏差理论(theory of large deviations)的研究联系起来（也称为**极值理论(extreme value theory)**），它探究了诸如股市崩盘和大型自然灾害等极其罕见的事件的发生频率。在统计分布的研究中更倾向于称之为“幂律”。

$$p(t;\lambda)=\begin{cases}0, & t<0\\ \dfrac{\lambda}{\exp(\lambda t)}, & t\geq 0\end{cases}$$

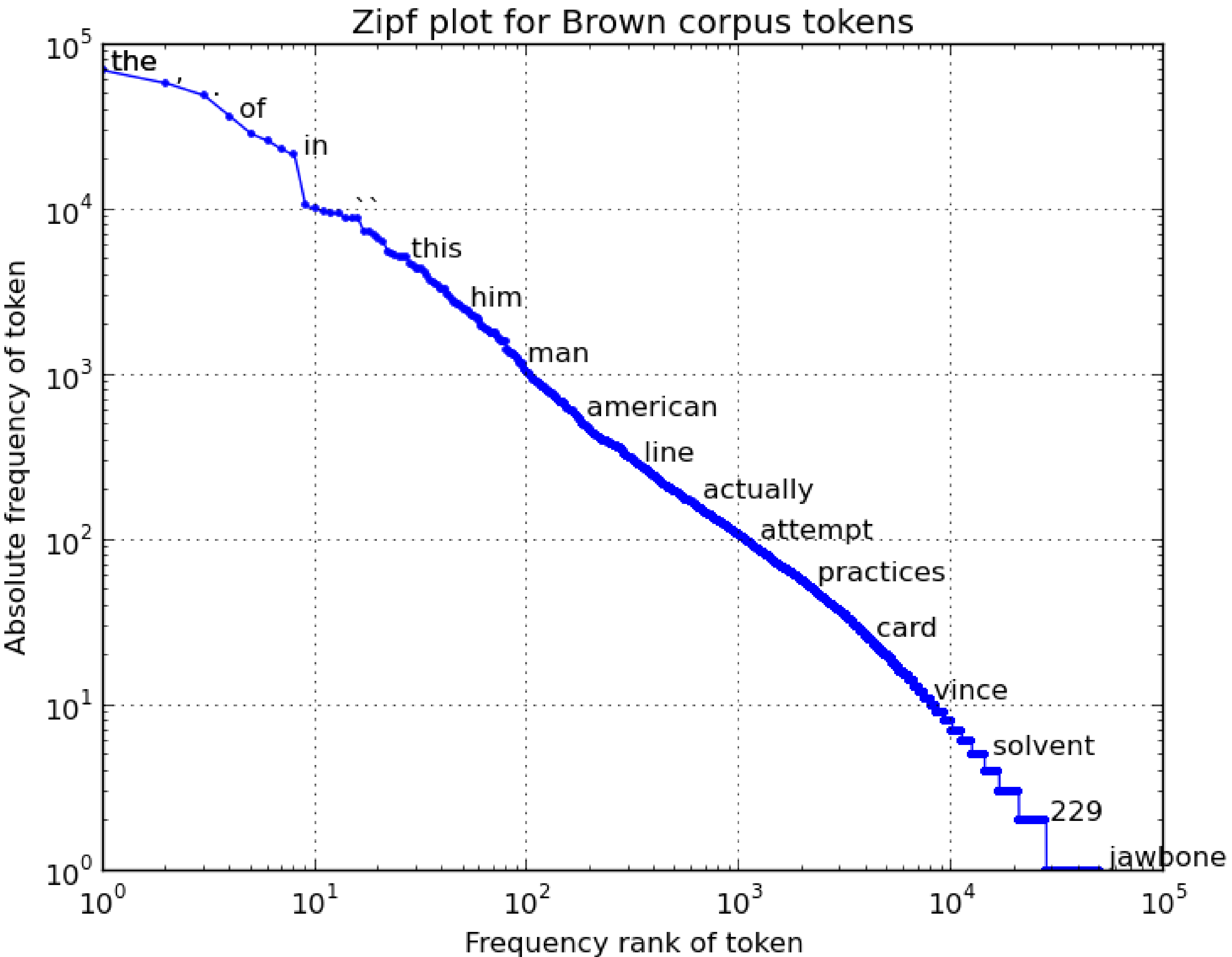




幂律分布特点

1. 平均值无意义
2. 小概率大影响
3. 无标度

Zipf定律



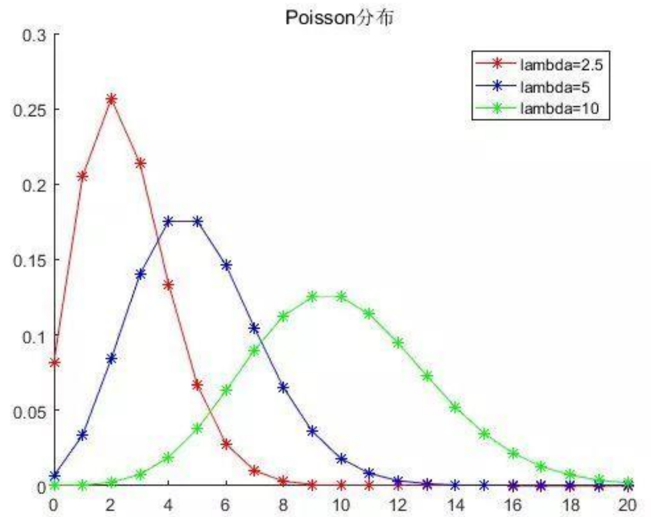
Zipf plot for Brown corpus tokens
Absolute frequency of token
Frequency rank of token
10^0
10^1
10^2
10^3
10^4
10^5
the
,
.
of
in
``
this
him
man
american
line
actually
attempt
practices
card
vince
solvent
229
jawbone

泊松分布

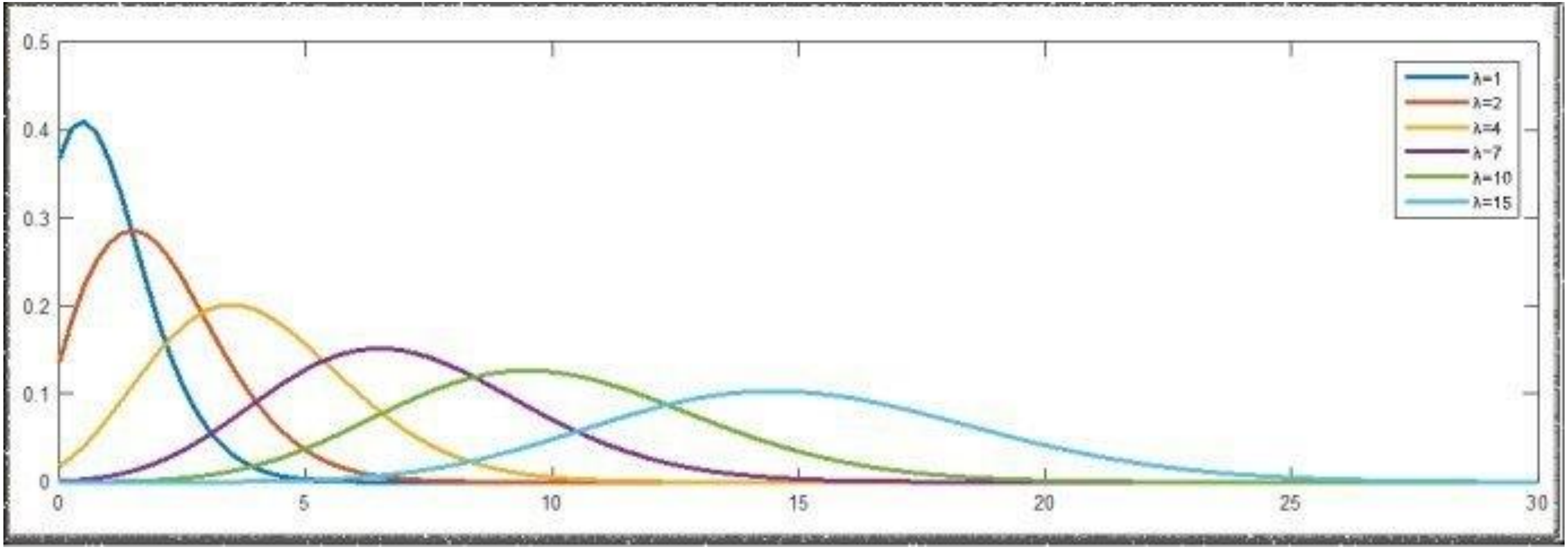
假设已知事件在单位时间（或者单位面积）内发生的**平均**次数为 λ，则泊松分布描述了：事件在单位时间（或者单位面积）内发生的具体次数为 *k* 的概率。

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

- 基础是正态分布
- 随机事件间隔无记忆



泊松分布过渡到正态分布



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

暴雨问题（灾难性问题常常是泊松分布的）

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

λ是整体概率与要求解问题匹配之后对应的数值。

整体概率是50年1次，也就是1/50。

接下来50年这个时间段暴雨次数的概率分布，λ =1。

100年，那λ就是1/50乘以100，λ=2。

5年呢？1/50乘以5，λ=0.1。

接下来50年，这时候 λ=1。

k=0, 接下来的50年1次大暴雨都不发生的概率：37%。

k=1, 接下来的50年发生1次大雨的概率：37%。

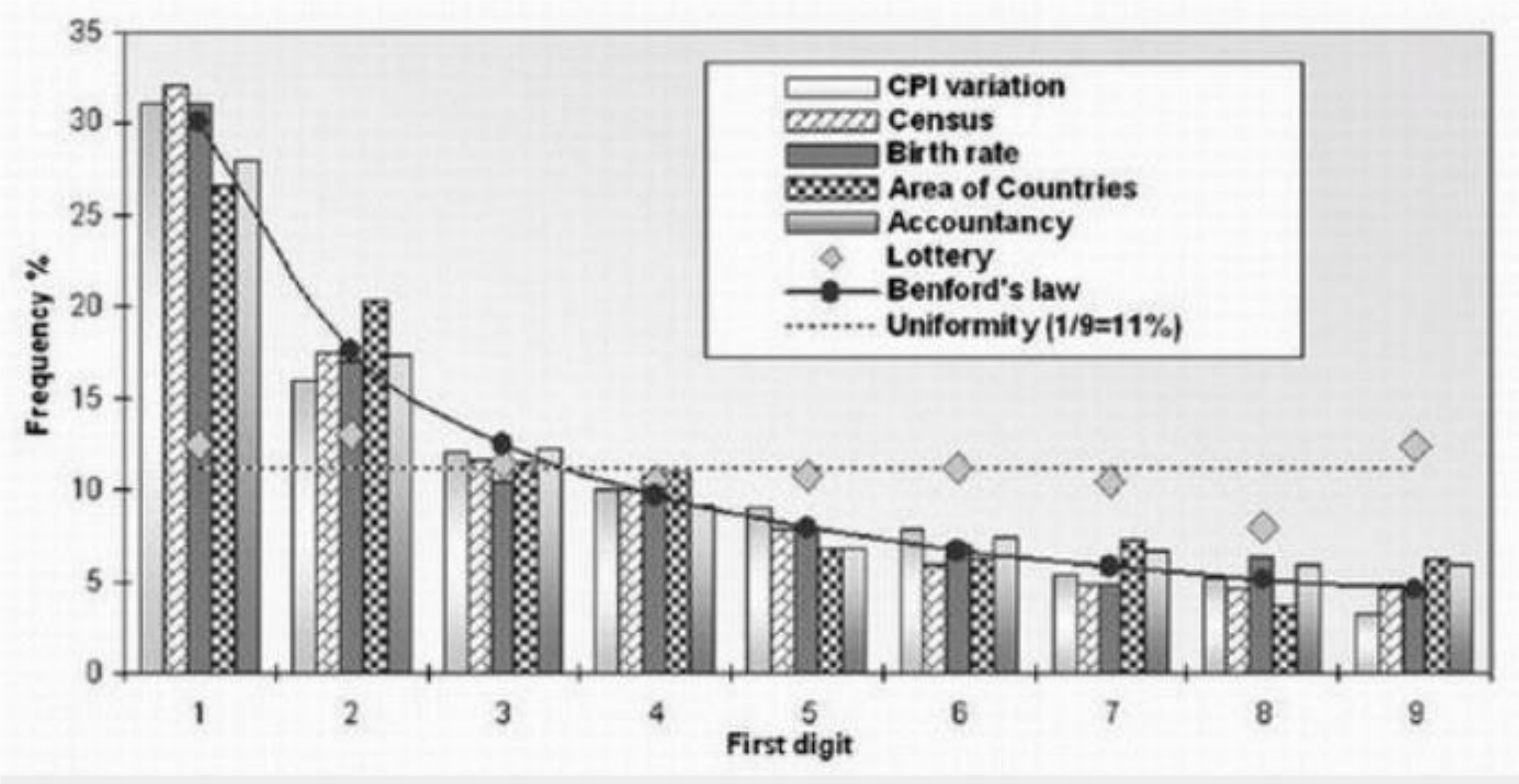
k=2, 接下来的50年发生2次大暴雨的概率：18%。

接下来50年发生2次和2次以上“五十年一遇”的大暴雨的概率是多少？

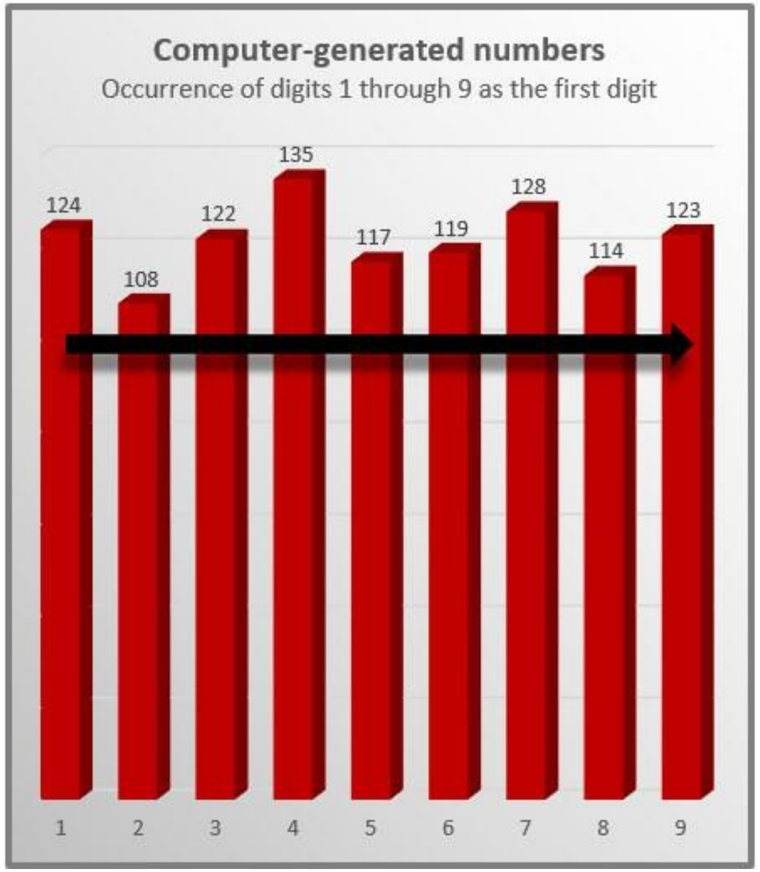
1减去发生0次的概率和发生1次的概率，1-37%-37%=26%

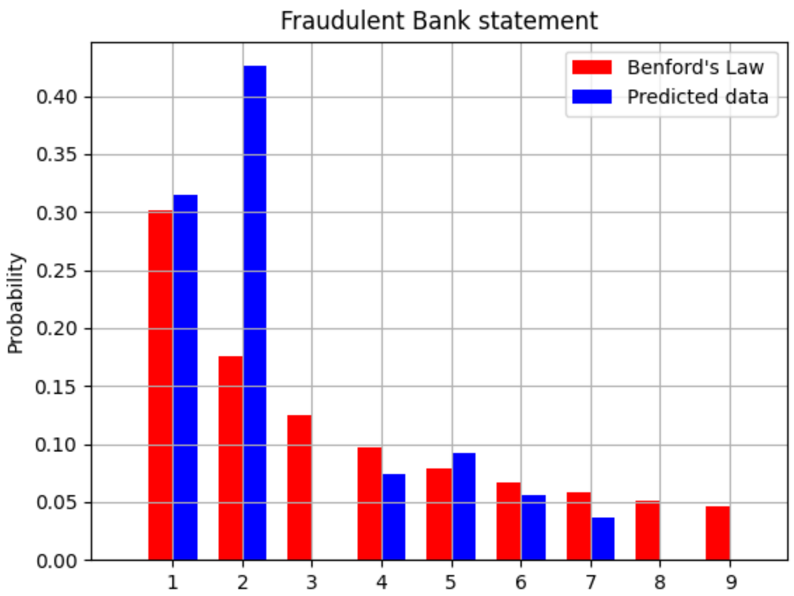
即：在“五十年一遇”的整体概率下，在接下来50年里发生2次或2次以上大暴雨的概率是26%。

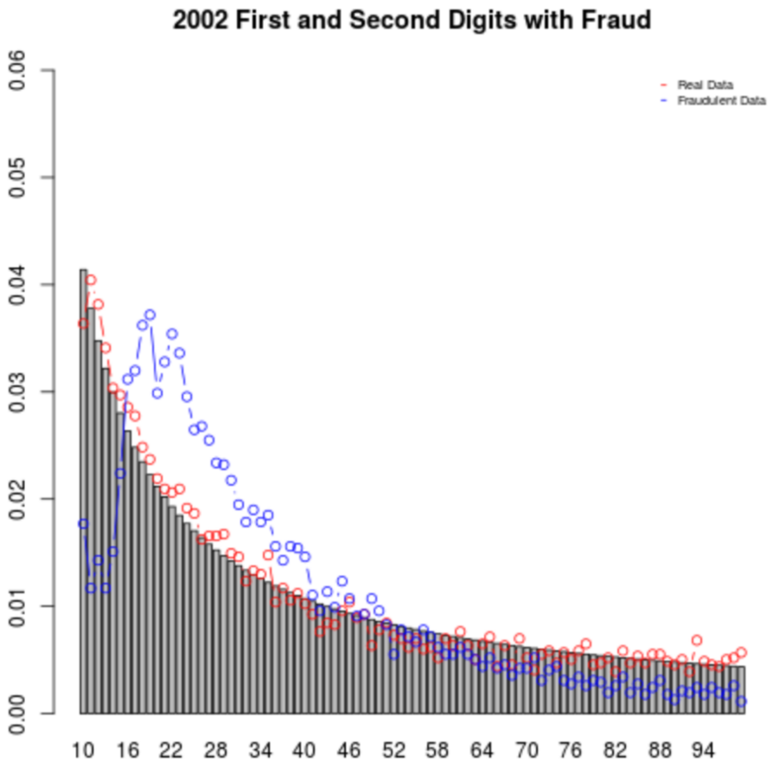
Benford定律



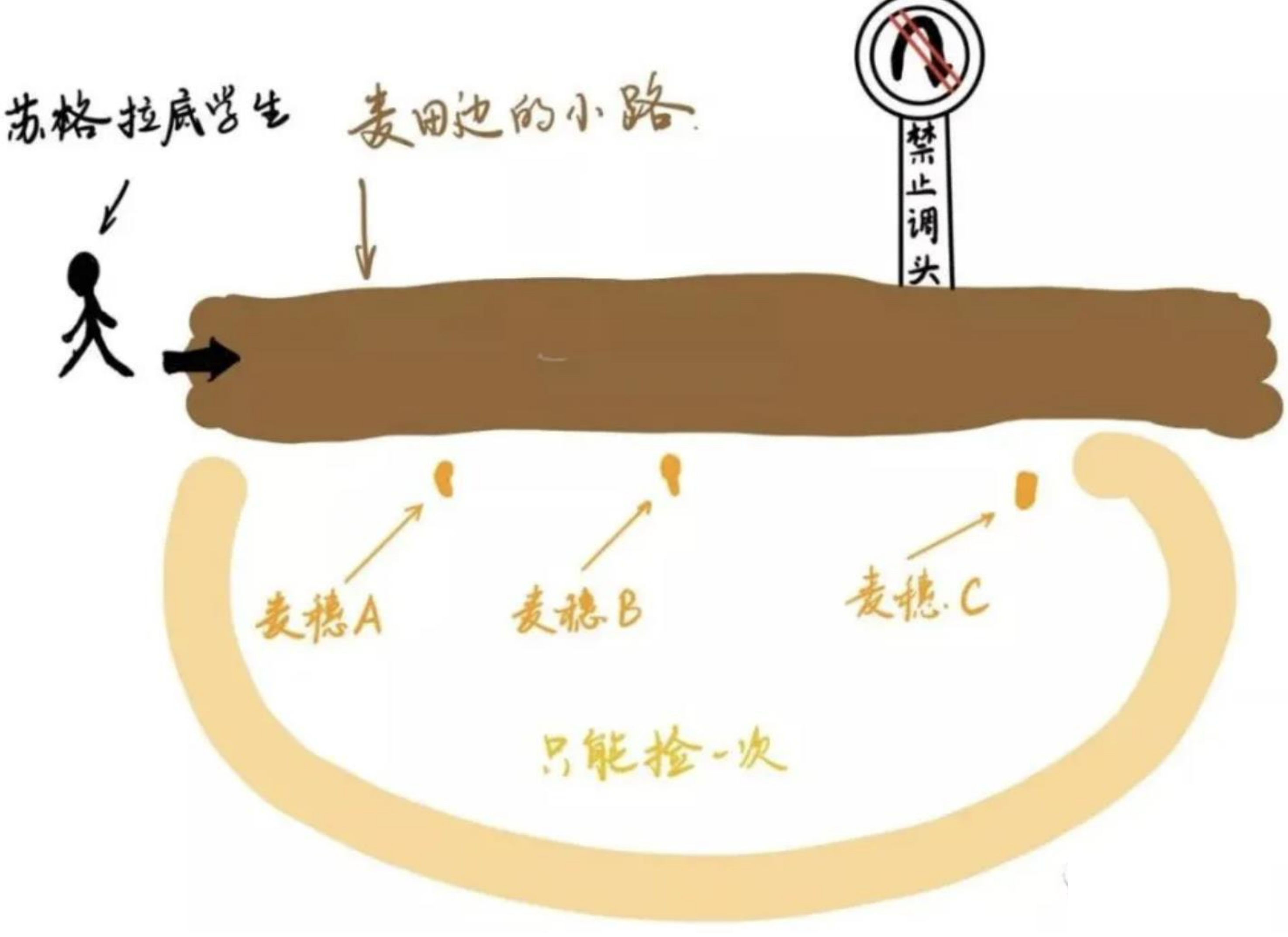
CPI variation
Census
Birth rate
Area of Countries
Accountancy
Lottery
Benford's law
Uniformity (1/9=11%)
Frequency %
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
First digit


Computer-generated numbers
Occurrence of digits 1 through 9 as the first digit
124
108
122
135
117
119
128
114
123
1
2
3
4
5
6
7
8
9


Fraudulent Bank statement
Benford's Law
Predicted data
Probability
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1
2
3
4
5
6
7
8
9


2002 First and Second Digits with Fraud
Real Data
Fraudulent Data
0.06
0.05
0.04
0.03
0.02
0.01
0.00
10
16
22
28
34
40
46
52
58
64
70
76
82
88
94

例1 麦穗问题（熊瞎子掰玉米问题）



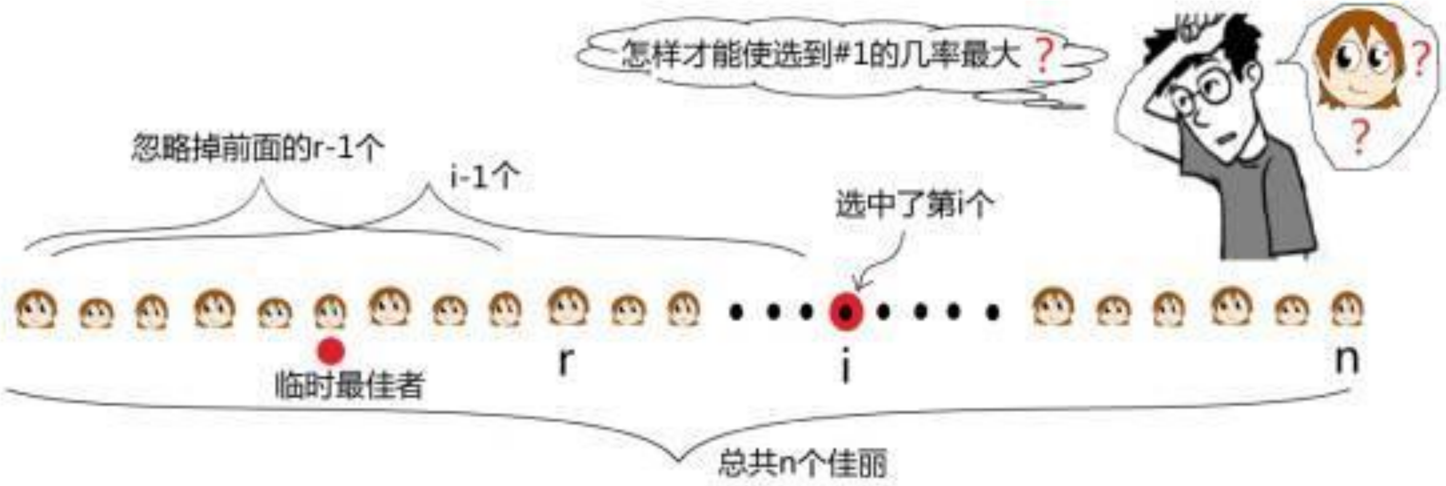
若策略为以A,B作参照，则能选到最大麦穗的次数为2:

A	小	小	中	中	大	大
B	中	大	小	大	小	中
C	大	中	大	小	中	小

- 第一个看到就选：1/3可能选到最大的
- 直接选第三个，后悔也没办，1/3选大最大的
- 只要比前面的大才选，有1/2可能性选到最大的，会选第1列、第2列、第3列、第4列、其中选项2、3、4中都可以选出最大的，但显然是至少要A作为参照

例2 追女孩的理想概率

某人相亲，总共假期一个月，约见100位女生。这100位女生一个一个见，每位只能见一次面，见面之后立即给出“不喜欢”或“喜欢”的答案。如果“不喜欢”，则以后再无机会面试该位女子，如果答案是“喜欢”，则意味着选中了这位女子，相亲过程便到此结束。

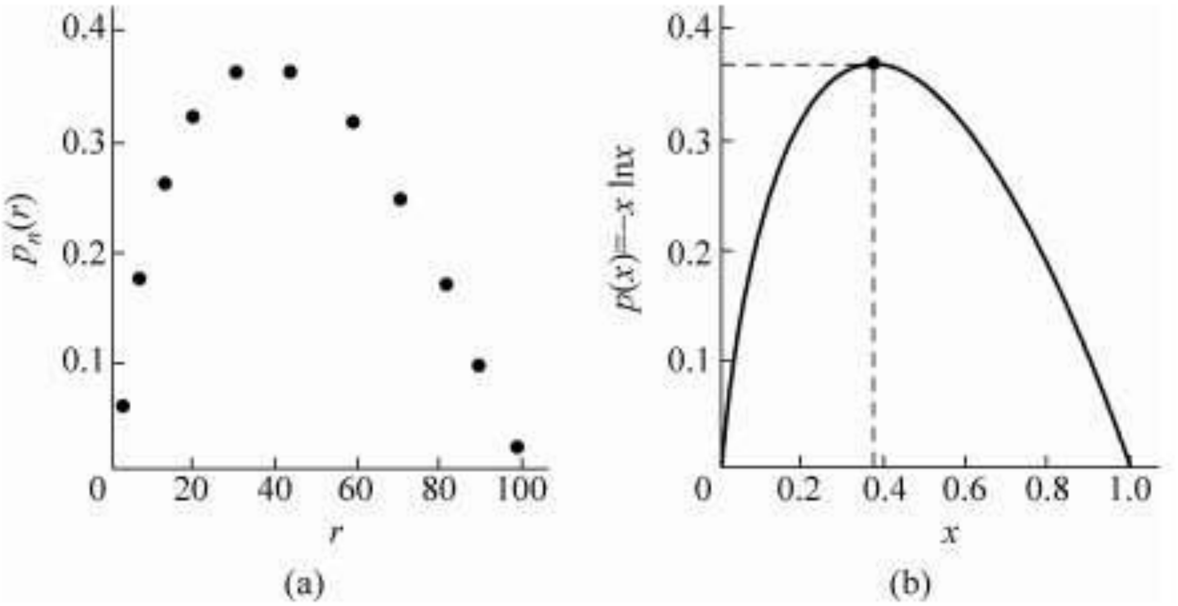


$$P(r)=\sum_{i=r}^{n}P(i)=\sum_{i=r}^{n}P(\text{选中了第i个})\times P(\text{第i个是\#1})=$$

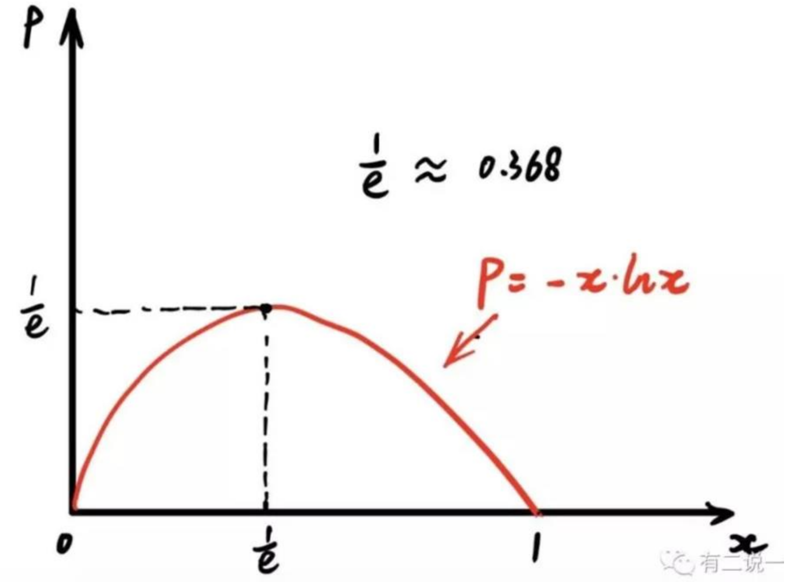
$$=\sum_{i=r}^{n}\frac{r-1}{i-1}\times\frac{1}{n}=\frac{r-1}{n}\sum_{i=r}^{n}\frac{1}{i-1}$$

$$\int\frac{1}{x}=lnx+C$$

这个事件发生说明从第*r*个到第*i-1*个人都不比参考组里面的临时最优者更好，否则会提前结束面试，而不会等到第*i*个人出场了。换句话说，“第*i*个人被面试”事件发生的条件是：前面*i-1*个人中的临时最优者必在参考组*r-1*个人里面，此概率为*(r-1)/(i-1)*。



0.4
0.3
0.2
0.1
$p_n(r)$
0
20
40
60
80
100
r
(a)
0.4
0.3
0.2
0.1
$p(x)=-x\ln x$
0
0.2
0.4
0.6
0.8
1.0
x
(b)

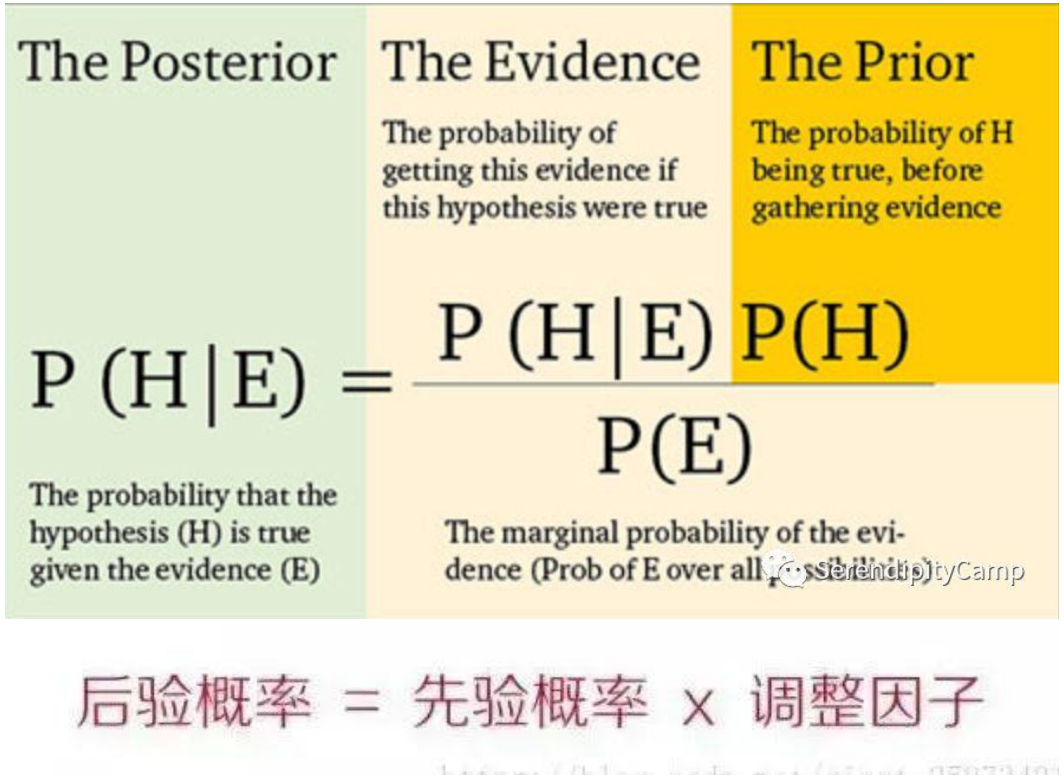


P
$\frac{1}{e} \approx 0.368$
$\frac{1}{e}$
$P = -x \cdot \ln x$
0
$\frac{1}{e}$
1
x
有二说一

条件概率和贝叶斯统计：给怀疑留下空间

"A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule."

贝叶斯定理，是控制证据强度的概率法则，告诉我们当知晓了一个新的事实或观察到新的证据时，该如何修改概率（改变我们的想法）。



若没有其他可作为依据的数据，可以把主观预测当作数据使用。
然后再根据新的信息，不断更新概率，结果就会越来越准确。

频率派与贝叶斯派各自不同的思考方式：

•频率派把需要推断的参数θ看做是固定的未知常数，即概率θ虽然是未知的，但最起码是确定的一个值，同时，样本X 是随机的，所以频率派重点研究样本空间，大部分的概率计算都是针对**样本 X 的分布**；

•贝叶斯派的观点则截然相反，他们认为参数θ是随机变量，而样本X 是固定的，由于样本是固定的，所以他们重点研究的是**参数θ的分布**。

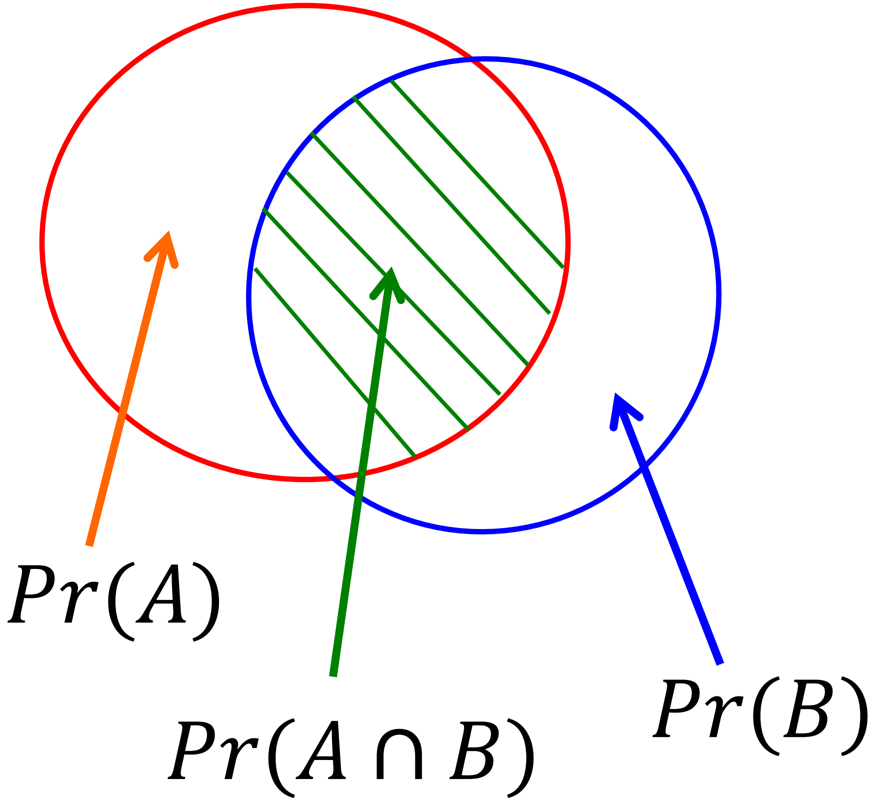
先验分布 **+ 样本信息** $\Rightarrow$ **后验分布**

$\pi(\theta)$ $\quad X \quad$ $\pi(\theta|x)$

条件概率：任何事件都是有条件的

在A事件发生的前提下，B发生的可能。

事件A情况下的条件概率 $P(B|A)$



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

$$Pr(A|B) + Pr(A'|B) = 1$$

贝叶斯定理

- 对易： $A \cap B = B \cap A$

$$\Pr(A \cap B) = \Pr(B \cap A)$$

- 由乘法规则:

$$\Pr(A \cap B) = \Pr(A|B)\Pr(B)$$
$$\Pr(B \cap A) = \Pr(B|A)\Pr(A)$$

- 则$\Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$

在某些情况下，条件概率需要反向计算，我们已知在B发生的情况下A发生的概率，但A发生的时候B发生的概率未必明显

$$\boxed{\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}}$$

Bayes' Theorem

扩展版贝叶斯定理

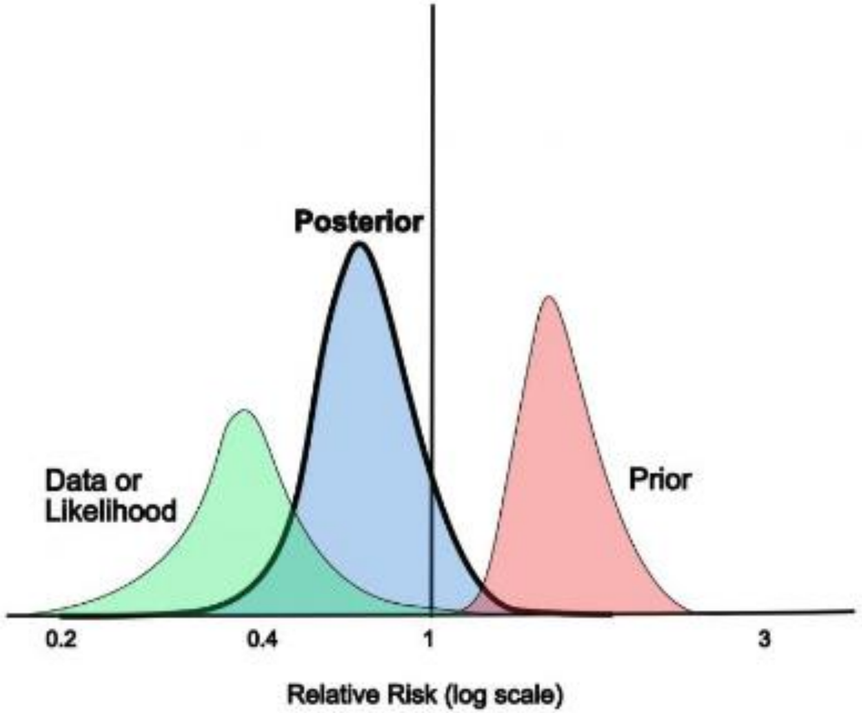
- 扩展版的贝叶斯定理

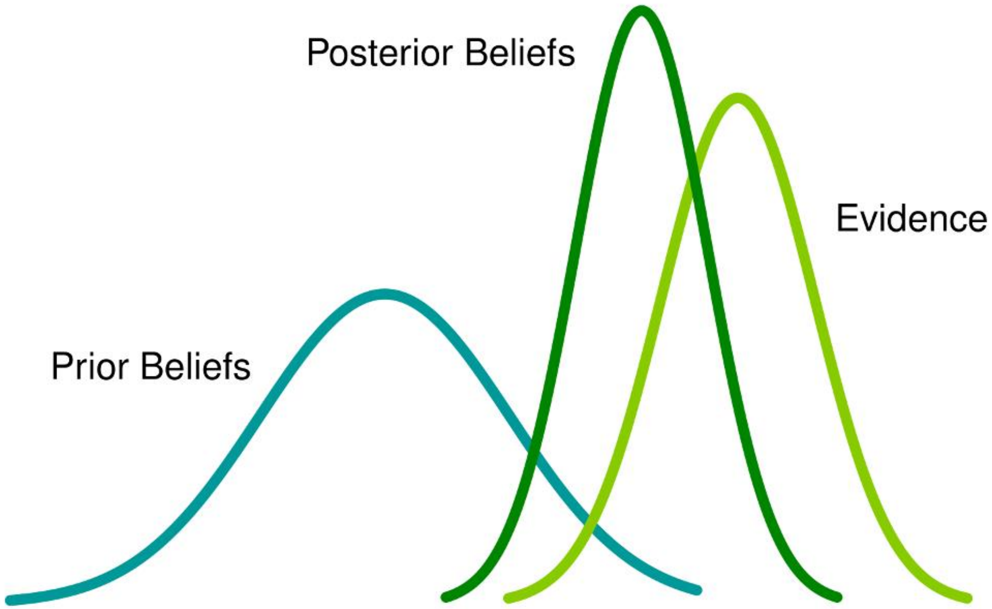
$$\Pr(B_i|A) = \frac{\Pr(A|B_i)}{\Pr(A)}$$

全概率定理

$$B_1 \cup B_2 \cup \cdots \cup B_m = \Omega$$

$$\boxed{\Pr(B_i|A) = \frac{\Pr(A|B_i)}{\sum_{j=1}^{m} \Pr\left(A\middle|B_j\right)\Pr(B_j)}}$$





似然函数

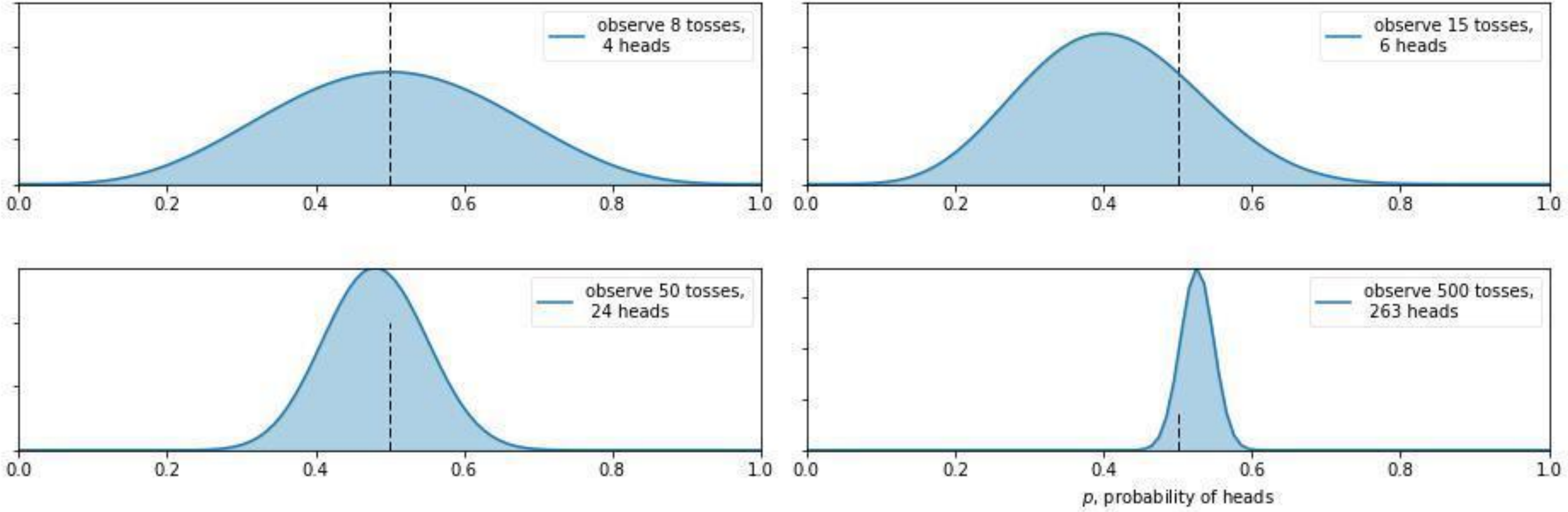
先验概率

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j)P(\omega_j)}{p(x)}$$

后验概率

证据因子（evidence）

$$p(x) = \sum_{j=1}^{2} p(x \mid \omega_j)P(\omega_j)$$



$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{\int f(y|\theta)f(\theta)d\theta}.$$

贝叶斯统计举例

知道结果求原因

例一

色盲是一种常见的遗传性视觉障碍，通常与X染色体上的基因突变有关。由于男性具有一个X染色体和一个Y染色体，而女性具有两个X染色体，因此男性更容易表现出色盲的性状。我们知道男性中有5%的人是色盲，而女性中只有0.25%的人是色盲。

在这个问题中，我们要计算的是，如果一个人被检测出患有色盲，那么他是男性的概率有多大。

贝叶斯定理：

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

计算的是一个人患有色盲（事件B）给定的条件下，他是男性（事件A）的概率，即 $P(A|B)$。

- $P(A)$ 是男性的先验概率，即在不考虑色盲的情况下，随机选择的一个人是男性的概率，这里假设为50%。
- $P(B|A)$ 是男性患有色盲的概率，即在已知一个人是男性的条件下，他患有色盲的概率，这里为5%。
- $P(B)$ 是一个人患有色盲的总概率，需要考虑男性和女性色盲的概率。

为了计算 $P(B)$，我们需要知道男性和女性色盲的概率，并结合他们的人口比例。由于男性人口比例为50%，女性为50%，男性中有5%是色盲，女性中有0.25%是色盲，我们可以得到：

$$P(B) = (0.05 \cdot 0.5) + (0.0025 \cdot 0.5) = 0.025 + 0.00125 = 0.00625$$

现在我们可以将这些值代入贝叶斯定理的公式中：

$$P(A|B) = \frac{0.05 \cdot 0.5}{0.00625} \approx 0.95238$$

- 例2、

Covid-19某一变种每一千个人里有一例。假设检测手段有99%的可靠性。同时，检测手段对没有病的人有2%的假阳性。如果一个人经过检测得阳性，那么他得病的概率实际是多少？

假设新冠检测试剂的准确率是99%，这意味着它有1%的误报率（即假阳性率）。如果疫情稳定，感染率P(A)是千分之一，即0.1%。现在我们要计算的是，一个人检测出阳性后，他实际感染新冠病毒的概率P(A|B)，其中B表示检测结果为阳性。在讨论新冠病毒检测的问题时，我们需要考虑几种不同的情况，这些情况涉及到检测的准确性和实际感染状态之间的关系。以下是四种可能的情况：

- 1. 真阴性（True Negative, TN）：个体没有感染病毒，检测结果也为阴性。
- 2. 真阳性（True Positive, TP）：个体感染了病毒，检测结果也为阳性。
- 3. 假阴性（False Negative, FN）：个体感染了病毒，但检测结果为阴性，即漏检。
- 4. 假阳性（False Positive, FP）：个体没有感染病毒，但检测结果为阳性，即误报。

应用贝叶斯公式：

- $P(B|A)$ 是在已知感染的情况下，检测结果为阳性的概率，即真阳性率，这里为99%。
- $P(A)$ 是感染新冠病毒的先验概率，即在没有任何检测结果的情况下，一个人感染新冠病毒的概率，这里是千分之一。
- $P(B)$ 是一个人检测结果为阳性的总概率，它包括了真阳性和假阳性的概率。

为了计算 $P(B)$，我们需要考虑两种情况：真阳性和假阳性。真阳性的概率是感染率乘以真阳性率，即$P(A)\cdot P(B|A)$。假阳性的概率是没有感染但检测结果为阳性的概率，即 $\big(1-P(A)\big)\cdot P(B|A')$，其中 A' 表示未感染，$P(B|A')$是假阳性率，即1%。

所以$P(B)$是：

$$P(B)=P(A)\cdot P(B|A)+\big(1-P(A)\big)\cdot P(B|A')$$

代入已知数值：

$$P(B)=0.001\cdot 0.99+(1-0.001)\cdot 0.01=0.01098$$

现在我们可以计算 $P(A|B)$：

$$P\big(A\big|B\big)=\frac{0.99\cdot 0.001}{0.01098}\approx 0.0902$$

例题 3 股票预测

有2个预测器，每个预测器在晚上会显示“涨”或者“跌”，来预测明天股市是涨还是跌。

根据历史统计，每个预测器预测的准确率都是0.7，并且预测器之间的预测结果是独立的。

今天晚上，2个预测器，都显示“涨”。

请问：明天股市涨的概率是多少？

设定以下术语：

A：明天股市涨

B1：第一个预测器显示"涨"

B2：第二个预测器显示"涨"

我们要计算的是后验概率 $P(A|B1,B2)$，即在两个预测器都显示"涨"的情况下，股市涨的概率。

根据贝叶斯定理：

$$P(A|B1,B2) = \frac{P(B1,B2 \mid A) \times P(A)}{P(B1,B2)}$$

我们知道：

$P(B1,B2|A)$：在股市涨的情况下，两个预测器都显示"涨"的概率。由于预测器之间的预测结果是独立的，所以这个概率可以计算为 P(B1|A) * P(B2|A)。

$P(A)$：明天股市涨的先验概率。由于没有提供先验信息，我们假设此概率为 0.5。

$P(B1,B2)$：两个预测器都显示"涨"的概率。这可以计算为

$P(B1,B2|A) * P(A) + P(B1,B2|A') * P(A')$，其中$A'$是股市跌的情况。

现在我们可以计算这些概率：

$$P(B1, B2|A) = P(B1|A) * P(B2|A) = 0.7 * 0.7 = 0.49$$

$$P(B1, B2) = P(B1, B2|A) * P(A) + P(B1, B2|A') * P(A')$$
$$= 0.49 * 0.5 + \left(1 - P(B1|A')\right) * \left(1 - P(B2|A')\right) * \left(1 - P(A)\right)$$
$$= 0.49 * 0.5 + 0.3 * 0.3 * 0.5 = 0.245 + 0.045 = 0.29$$

现在，我们可以计算后验概率：

$$P(A|B1, B2) = \frac{0.49 \times 0.5}{0.29} \approx \frac{0.245}{0.29} \approx 0.845$$

因此，在两个预测器都显示″涨″的情况下，明天股市涨的概率约为 0.845，或者约为 84.5%。

Sun Rise?

我们按照Bayesian推断的方式来讨论这个问题。采用如下记号。

$$SunRise = ``太阳升起"$$

$$NotRise = ``太阳没有升起"$$

假设$P\left(SunRise\right) = p$, 则$p \in \left[0,1\right]$, 但是我们并不知道$p$的准确值。事实上，$p$的准确值正是我们要推断的。既然我们不知道$p$到底等于多少，那么我们就需要对$\left[0,1\right]$之间的每一个数是否就是真实的$p$值进行评估。Bayesian先验分布就是这样的一个评估。假设先验分布密度函数为$w\left(p\right)$。这意味着我们的评估相当于说$p \in \left[a,b\right] \subseteq \left[0,1\right]$的可能性为$\int_a^b w\left(p\right) dp$。

按照Bayesian推断的方法。当我们观察到事实，太阳在过去的N天中升起了k次。那么我们需要将对于太阳升起的概率修正为

$$w(p|N,k) = \frac{w(p)p^k(1-p)^{N-k}}{\int_0^1 w(p)p^k(1-p)^{N-k}\, dp}.$$

关于p的一个评估，即在观察到事实(N, k)后，对于p的真实值的一个评估。我们需要给出一个具体的值，换言之，要作一个关于p的点估计。Bayesian的方法是计算平均值，即以

$$\hat{p} = \int_0^1 p\, w(p|N, k) dp,$$

作为``太阳升起"的主观概率。

为计算出来一个确定的值，需要知道先验分布$w(p)$.
接下来，我们需要选定一个先验分布从而可以计算出$\hat{p}$.

Laplace认为一个公平的假设是合理的，即p是[0,1]中的任意一个值都有相同的可能性。亦即先验分布密度为$w(p) = 1$. 这样可以算出

$$\hat{p} = \frac{k+1}{N+2}.$$

这个值称之为Laplace法则。
假设在人类的历史上太阳一直升起。这是合理的。再假设人类的文明史始于亚当，或者大洪水之后，或者公元前6000年，距离现在差不多

$$6000 \times 365 \approx 2 \times 10^6 \text{天}.$$

因而我们可以认为太阳明天不升起的概率为一百万分之五。

朴素贝叶斯分类

水果	长	甜	黄	总数
香蕉	400	350	450	500
苹果	0	150	300	300
梨子	100	150	50	200
总数	500	650	800	1000

数据表达中信息是相互独立的：**朴素**
问：一个长黄甜的东西最大可能是啥？

$$\Pr(B_i|A) = \frac{\Pr(A|B_i)}{\sum_{j=1}^{m} \Pr(A|B_j)\Pr(B_j)}$$

贝叶斯分类：水果 香蕉 苹果 梨 特点：长 黄 甜

$$P(LSY|B) = P(L|B) * P(S|B) * P(Y|B) = 0.8 * 0.7 * 0.9 = 0.504$$

$$P(B|LSY) = \frac{P(长黄甜|B)P(B)}{P(长黄甜)} = P(LSY|B)\frac{P(B)}{P(LSY)} = \frac{0.504 * 0.5}{P(LSY)}$$

$$P(LSY|Apple) = P(L|A) * P(S|A) * P(Y|A) = 0 * 0.5 * 1 = 0$$

$$P(LSY|Pear) = P(长|梨) * P(甜|梨) * P(黄|梨) = 0.5 * 0.75 * 0.25 = 0.09375$$

所以分母：所有长黄甜的可能性之和

$$P(长黄甜) = P(长黄甜|Banana)P(Banana) + P(长黄甜|Apple)P(Apple) + P(长黄甜|Pear)P(Pear)$$
$$= 0.504 * 0.5 + 0 * 0.3 + 0.09375 * 0.2 = 0.271$$

$$\therefore\ P(B|LSY) = \frac{0.504 * 0.5}{0.271} = 93\%$$

$$P(A|LSY) = 0$$

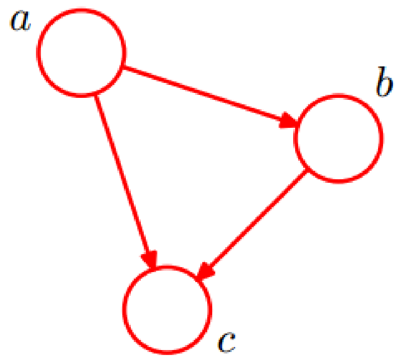
$$P\left(P|LSY\right) = \frac{0.09375*0.2}{0.27073} = 7\%$$

证据叠加和贝叶斯网络

假设节点E直接影响到节点H，即E→H，则用从E指向H的箭头建立结点E到结点H的有向弧(E,H)，权值(即连接强度)用条件概率P(H|E)来表示，如下图所示：

把某个研究系统中涉及的随机变量，根据是否条件独立绘制在一个有向图中，就形成了贝叶斯网络。其主要用来描述随机变量之间的条件依赖，用圈表示随机变量(random variables)，用箭头表示条件依赖(conditional dependencies)。

如下图所示，便是一个简单的贝叶斯网络：

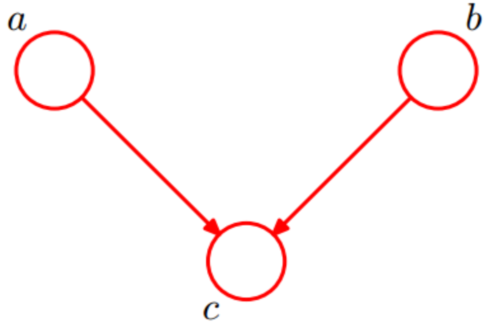


因为a导致b，a和b导致c，所以有:

$$p(a,b,c)=p(c|a,b)p(b|a)p(a)$$

形式1：head-to-head

贝叶斯网络的第一种结构形式如下图所示：



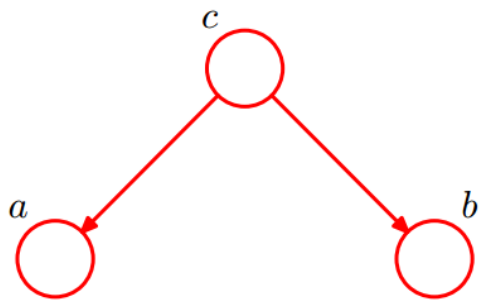
所以有：P(a,b,c) = P(a)*P(b)*P(c|a,b)成立，化简后可得：

$$\sum_{c} P(a,b,c) = \sum_{c} P(a)*P(b)*P(c \mid a,b)$$
$$\Rightarrow P(a,b) = P(a)*P(b)$$

即在c未知的条件下，a、b被阻断(blocked)，是独立的，称之为head-to-head条件独立，对应本节中最开始那张图中的“x1、x2独立”。

形式2：tail-to-tail

贝叶斯网络的第二种结构形式如下图所示:



考虑c未知，跟c已知这两种情况：

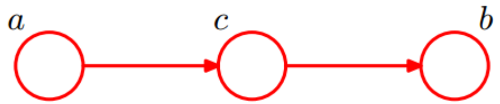
•在c未知的时候，有：P(a,b,c)=P(c)*P(a|c)*P(b|c)，此时，没法得出P(a,b) = P(a)P(b)，即c未知时，a、b不独立。

•在c已知的时候，有：P(a,b|c)=P(a,b,c)/P(c)，然后将P(a,b,c)=P(c)*P(a|c)*P(b|c)带入式子中，得到：P(a,b|c)=P(a,b,c)/P(c) = P(c)*P(a|c)*P(b|c) / P(c) = P(a|c)*P(b|c)，即c已知时，a、b独立。

所以，在c给定的条件下，a，b被阻断(blocked)，是独立的，称之为tail-to-tail条件独立，对应本节中最开始那张图中的"x6和x7在x4给定的条件下独立"。

形式3：head-to-tail

贝叶斯网络的第三种结构形式如下图所示：



还是分c未知跟c已知这两种情况：

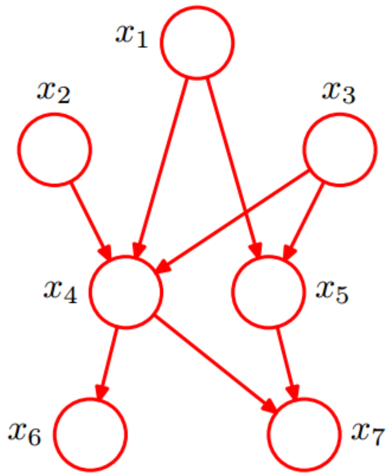
•c未知时，有：P(a,b,c)=P(a)*P(c|a)*P(b|c)，但无法推出P(a,b) = P(a)P(b)，即c未知时，a、b不独立。

•c已知时，有：P(a,b|c)=P(a,b,c)/P(c)，且根据P(a,c) = P(a)*P(c|a) = P(c)*P(a|c)，可化简得到：

$$
\begin{aligned}
&P(a,b|c)\\
&= P(a,b,c)/P(c)\\
&= P(a)*P(c|a)*P(b|c)\ /\ P(c)\\
&= P(a,c)*\ P(b|c)\ /\ P(c)\\
&= P(a|c)\ *\ P(b|c)
\end{aligned}
$$

所以，在c给定的条件下，a，b被阻断(blocked)，是独立的，称之为head-to-tail条件独立。

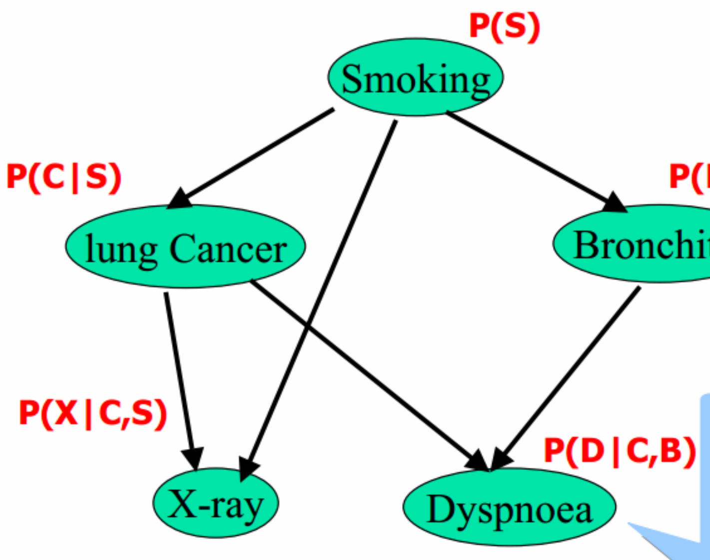
给定如下图所示的一个贝叶斯网络：



从图上可以比较直观的看出：
(1). x1,x2,…x7的联合分布为:

$$p(x_1)p(x_2)p(x_3)p(x_4|x_1,x_2,x_3)p(x_5|x_1,x_3)p(x_6|x_4)p(x_7|x_4,x_5)$$

(2). x1和x2独立（对应head-to-head）；
(3). x6和x7在x4给定的条件下独立（对应tail-to-tail）。

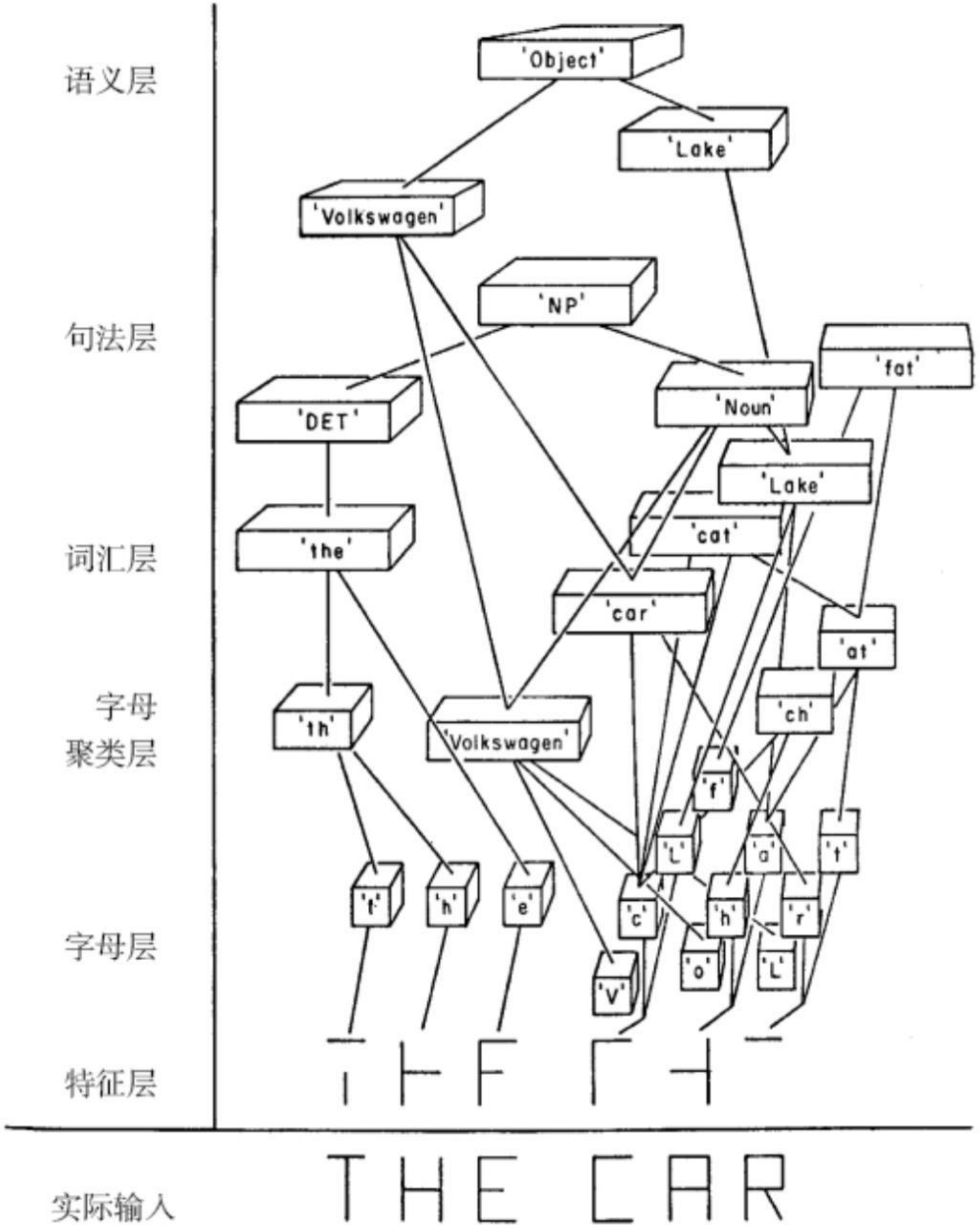


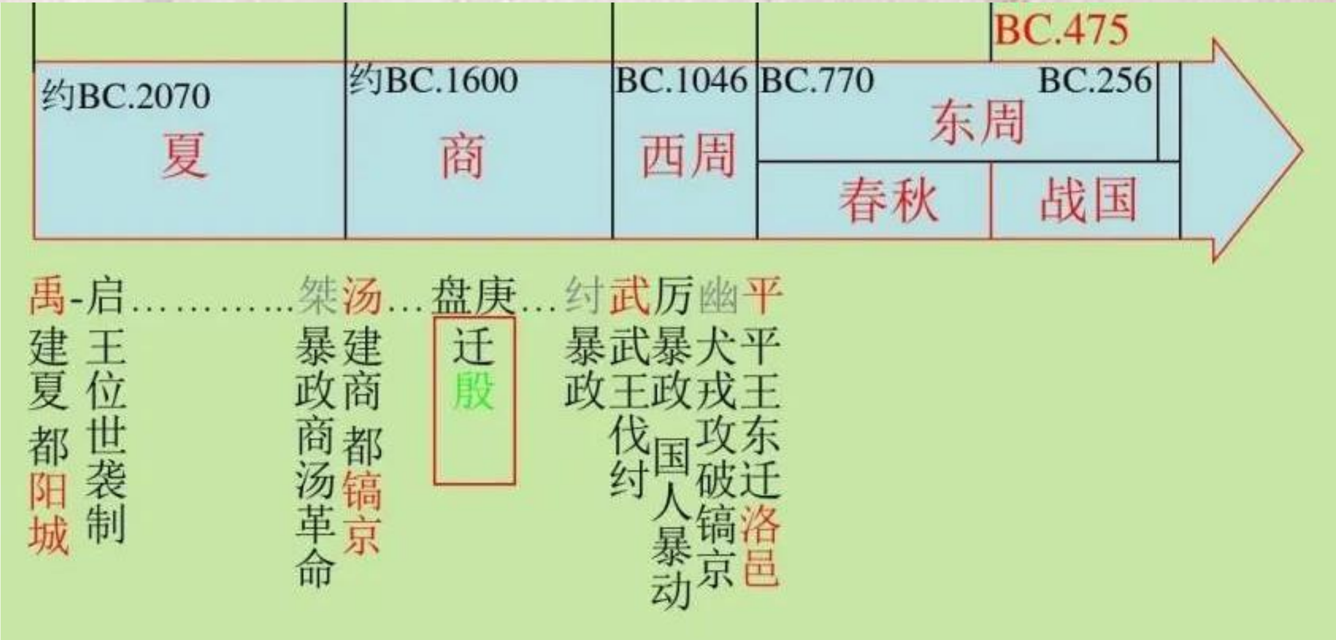
C	B	D=0	D=1
0	0	0.9	0.1
0	1	0.3	0.7
1	0	0.2	0.8
1	1	0.1	0.9

•smoking表示吸烟，其概率用P(S)表示，
•lung Cancer表示的肺癌，一个人在吸烟的情况下得肺癌的概率用P(C|S)表示，
•X-ray表示需要照医学上的X光，肺癌可能会导致需要照X光，吸烟也有可能会导致需要照X光（所以smoking也是X-ray的一个因），所以，因吸烟且得肺癌而需要照X光的概率用P(X|C,S)表示。
•Bronchitis表示支气管炎，一个人在吸烟的情况下得支气管炎的概率用P(B|S)，
•dyspnoea表示呼吸困难，支气管炎可能会导致呼吸困难，肺癌也有可能会导致呼吸困难（所以lung Cancer也是dyspnoea的一个因），因吸烟且得了支气管炎导致呼吸困难的概率用P(D|S,B)表示。

lung Cancer简记为C，Bronchitis简记为B，dyspnoea简记为D，且C = 0表示lung Cancer不发生的概率，C = 1表示lung Cancer发生的概率，B等于0（B不发生）或1（B发生）也类似于C，同样的，D=1表示D发生的概率，D=0表示D不发生的概率，便可得到dyspnoea的一张概率表，如上图的最右下角所示。

1、字母层面:识别个体特征，如线条和曲线，判断它们可能构成哪个字母；
2、词汇层面:根据识别的字母和字母组合猜测可能的词；
3、句法层面:根据猜测的词和语法规则进行进一步的推断；
4、语义层面:考虑上下文信息，对整个句子或短语进行解释





中国历史上有确切纪年之始公认只能上溯至司马迁《史记》中所载西周共和元年（前841年）

夏商周断代工程是中国科学院在20世纪90年代发起的一项系统研究中国早期历史年代问题的大型学术工程。

这是将历史学、考古学和天文学等多学科交叉融合，实行多学科联合攻关的大型科学研究项目。它不仅关联到中国历代各种历书的改进和完善，而且影响到历史、天文、地质、地理、环境、文化等多个领域的研究和教学。其研究方法包括古文献学、考古学、古代科技史以及天文学等方法。

第一，精确测定中国古代历史的年代。

第二，恢复古代历法，尤其是恢复夏商周三代的历法。

第三，确定夏商周三代历史的大事记年代，包括各个历代的王位年号以及重大历史事件的年代。

成果

为中国远古三个重要时期——夏、商和周朝设定了明确的年代。例如，夏朝起始于前2070年，商朝起始于前1600年，周朝起始于前1046年。

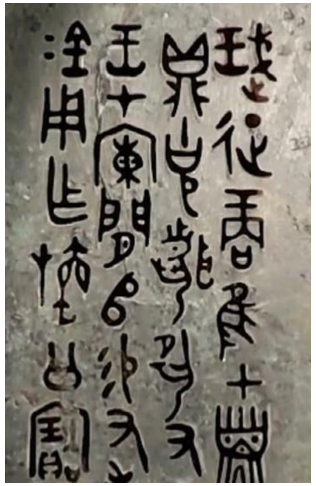
恢复了夏、商、周三代的历法和历史记年方法。明确了夏朝的年号为“干支纪年”，商朝的年号为“天干纪年”，西周的年号为“地支纪年”。这些年号的确定，不仅解决了中国历史的年代问题，也填补了世界早期文明研究的空白。

成果中除了年代确定，还包括对历史大事记的甄别与精确事件年代的推断，如对殷墟甲骨“三日食”的解读等。

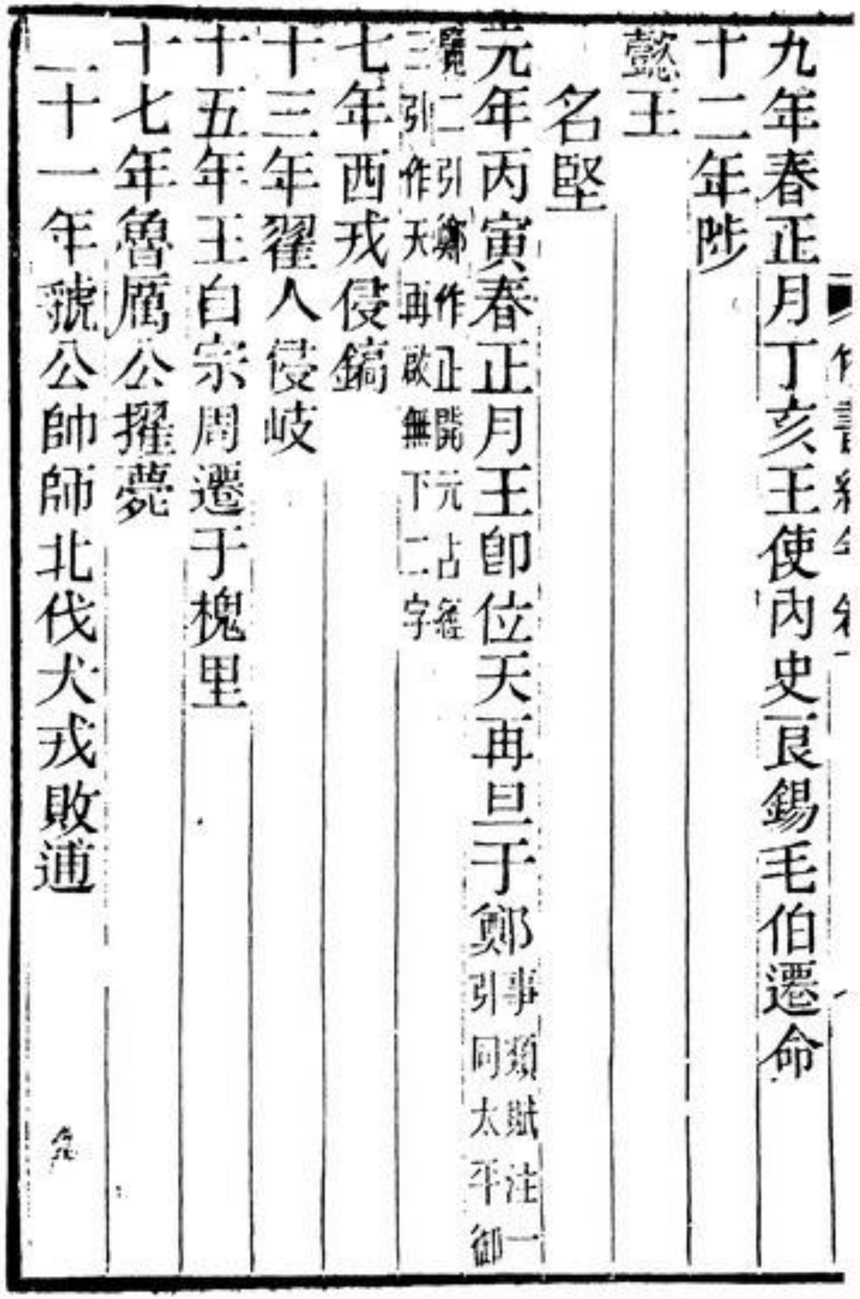
1995年宋健的指导下组织七个部、委主要负责人成立强大的“夏商周断代工程领导小组”。该领导小组成员如下:

国家科委	副主任	邓楠
国家自然科学基金委主任		陈佳洱
中国科学院院长		路甬祥
国家教委	副主任	韦钰
国家文物局	局长	张德勤

邓楠、陈佳洱任正、副组长。



武征商，唯甲子朝，岁鼎，克昏夙有商。辛未，王在阑师，赐右吏利金，用作檀公宝尊彝。


九年春正月丁亥王使內史良錫毛伯遷命

十二年陟

懿王

名堅

元年丙寅春正月王卽位天再旦于鄭事類賦注一引同太平御覽二引鄭作正號元占經三引作天再啟無下二字

七年西戎侵鎬

十三年翟人侵岐

十五年王自宗周遷于槐里

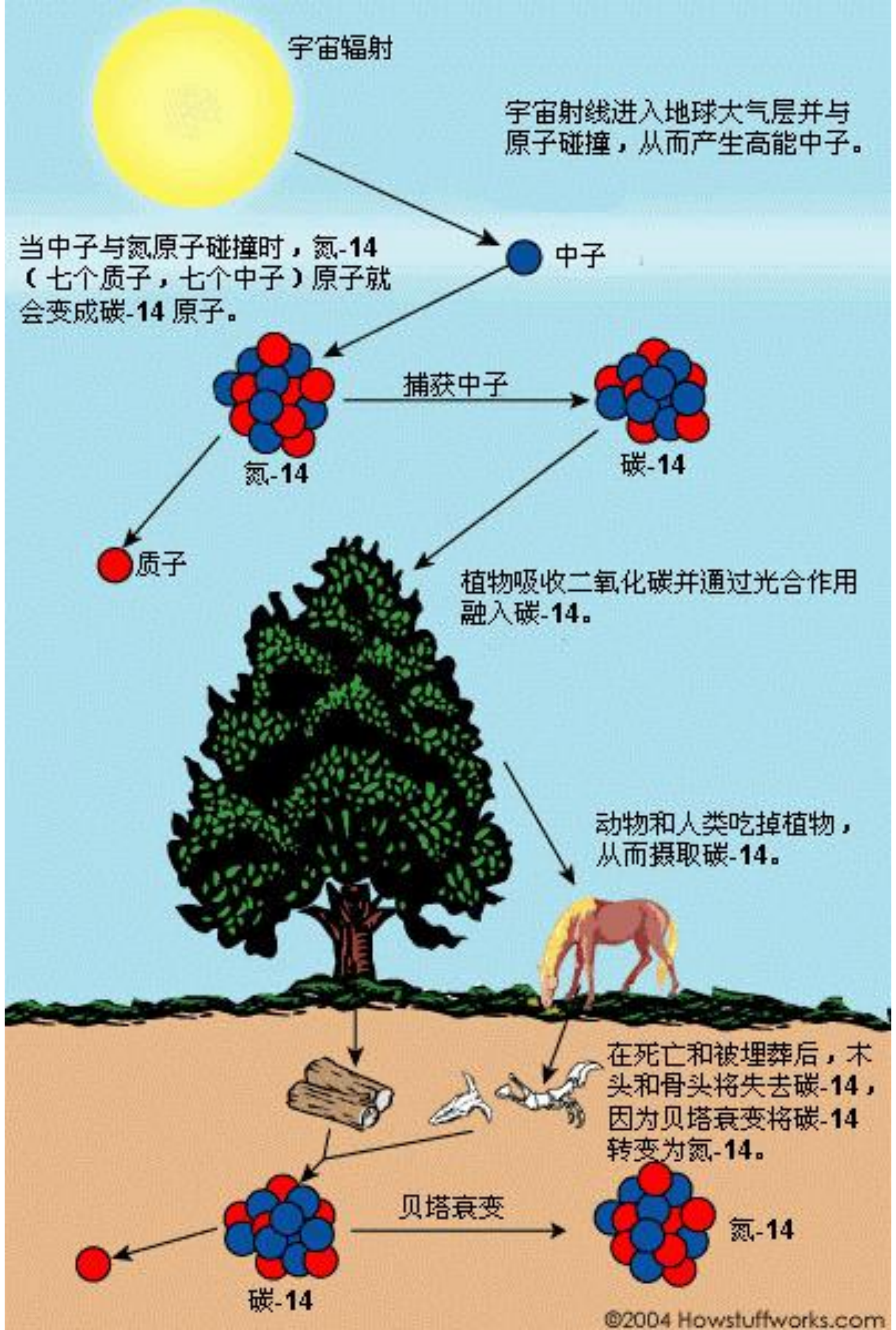
十七年魯厲公擢薨

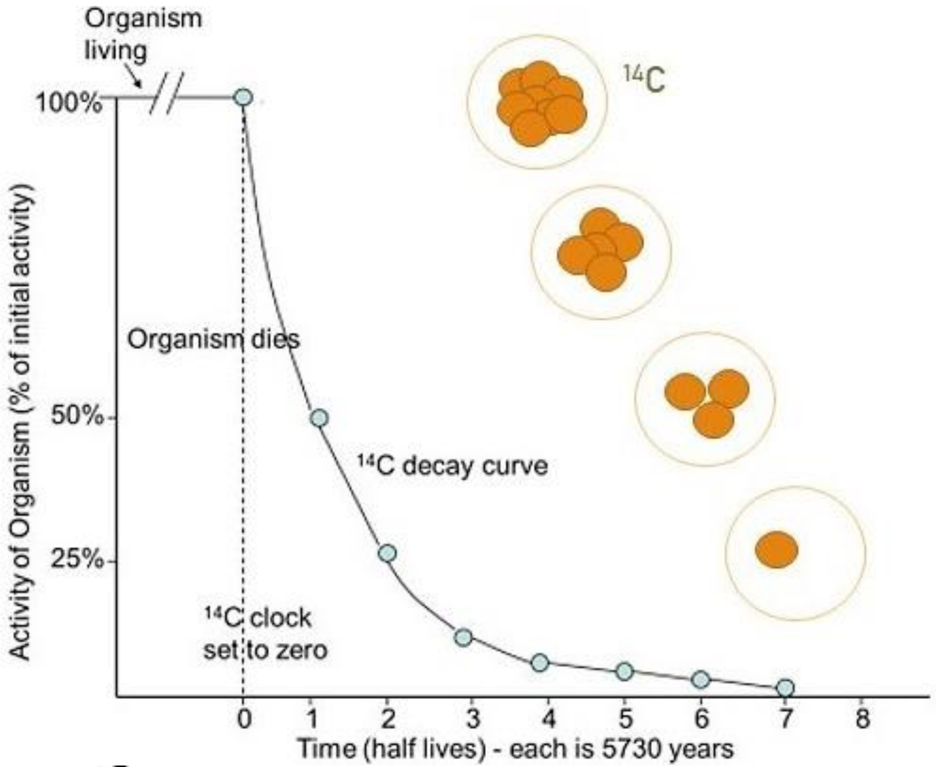
二十一年虢公帥師北伐犬戎敗逋

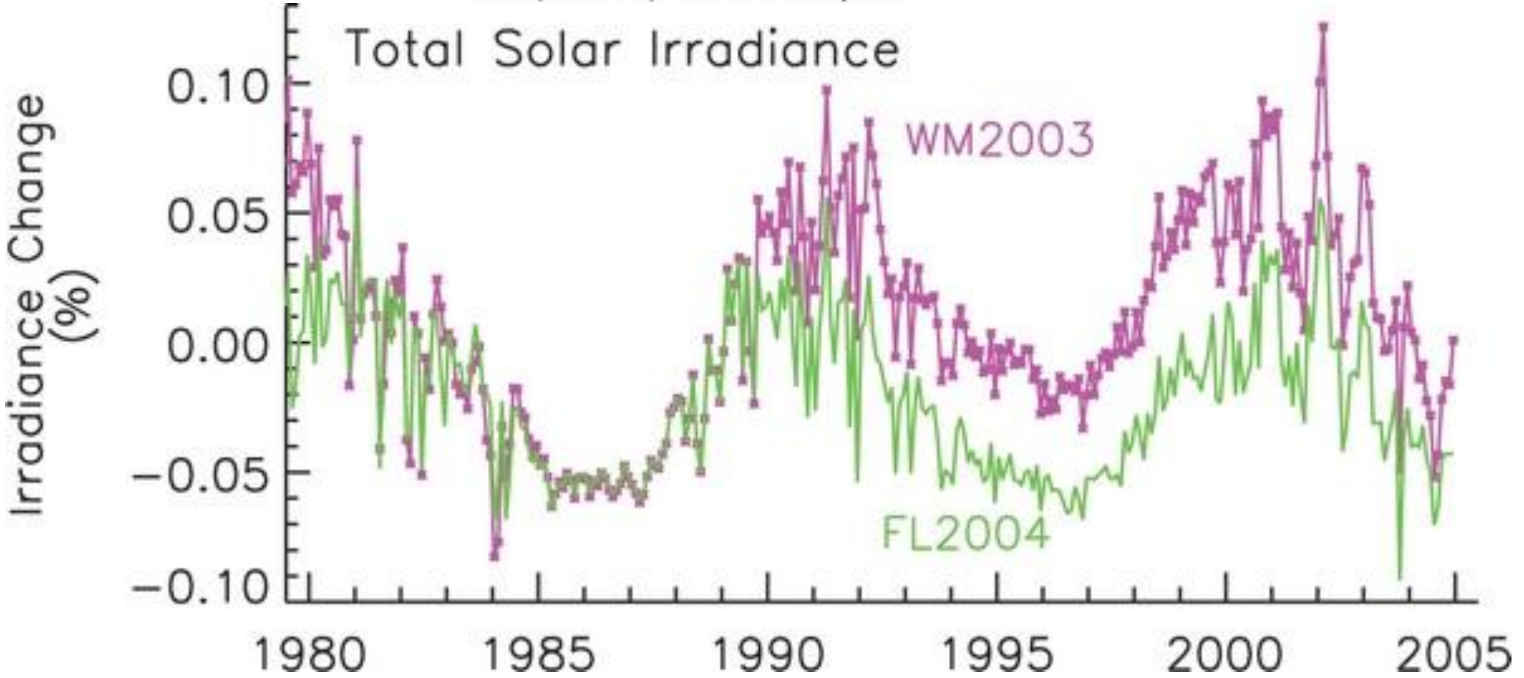
今本《竹书纪年》

葛真参酌西周青铜器铭文推定是次日食发生于前925年或前899年、张培瑜等人经系统性计算日食推定有前925年、前899年、前919年或前903年等几个可能，以及美国学者倪德卫、日本学者平势隆郎推定发生于前903年等说法。

1996年，中国大陆年代学研究“夏商周断代工程”展开，其中为推定西周王年建立了七个支点，而“天再旦于郑”即成为用以确定懿王元年对应西历年份之材料。


宇宙辐射
宇宙射线进入地球大气层并与原子碰撞，从而产生高能中子。
中子
当中子与氮原子碰撞时，氮-14（七个质子，七个中子）原子就会变成碳-14 原子。
捕获中子
氮-14
碳-14
质子
植物吸收二氧化碳并通过光合作用融入碳-14。
动物和人类吃掉植物，从而摄取碳-14。
在死亡和被埋葬后，木头和骨头将失去碳-14，因为贝塔衰变将碳-14转变为氮-14。
贝塔衰变
碳-14
氮-14
©2004 Howstuffworks.com


Organism living
100%
50%
25%
Activity of Organism (% of initial activity)
Organism dies
^{14}C decay curve
^{14}C clock set to zero
^{14}C
0 1 2 3 4 5 6 7 8
Time (half lives) - each is 5730 years

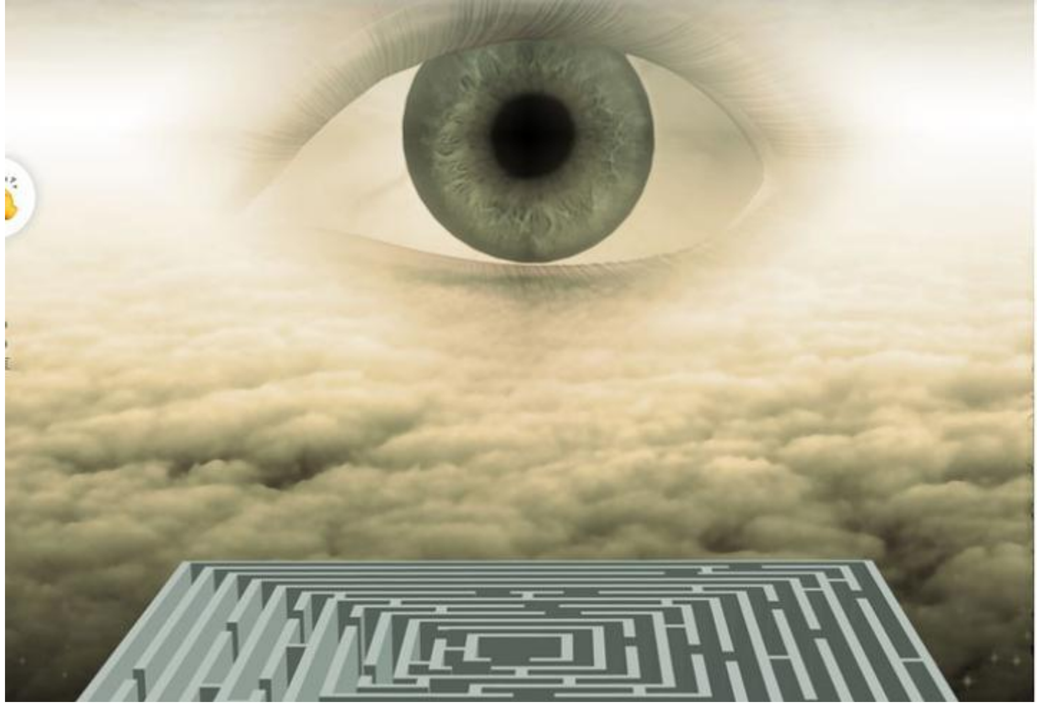

Total Solar Irradiance
Irradiance Change (%)
0.10
0.05
0.00
−0.05
−0.10
WM2003
FL2004
1980 1985 1990 1995 2000 2005

夏商周断代工程中在测年技术研究与改进减小测年误差的基础上，与考古学密切结合，通过系列样品-贝叶斯统计方法的应用与研究，有效地减小了日历年误差，完成河南偃师二里头、郑州商城二里岗、陕西长安马王村H18、安阳殷墟、北京琉璃河西周墓地等系列的年代测定与研究，建立起夏商周考古-碳14年代框架，为三代年表的建立提供了依据。

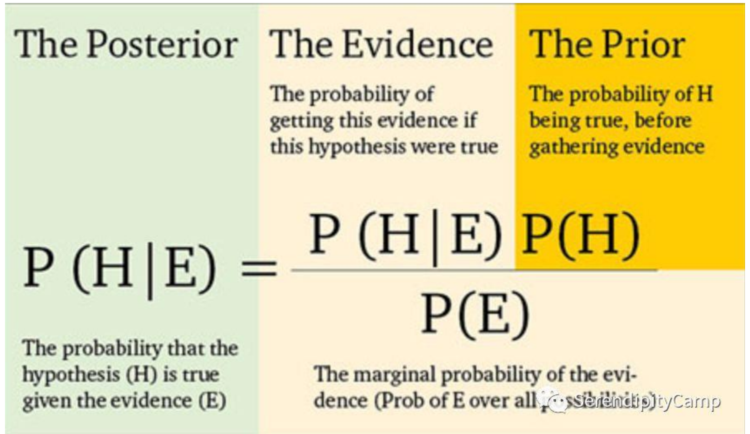
夏商周断代工程也开启了多学科结合应用系列样品-贝叶斯统计方法进行高精度年代研究的先河。

三种关于概率的定义

- 统计定义 $P(A)=\dfrac{\text{事件}A\text{包含的样本数}}{\text{样本空间的总样本数}}=\dfrac{M}{N}$
- 为什么频率等于概率
  - 频率派把需要推断的参数 θ 看做是固定的未知常数，即概率 θ 虽然是未知的，但最起码是确定的一个值，同时，样本 X 是随机的，所以频率派重点研究样本空间，大部分的概率计算都是针对**样本 X 的分布**；
  - 贝叶斯派的观点则截然相反，他们认为参数 θ 是随机变量，而样本 X 是固定的，由于样本是固定的，所以他们重点研究的是**参数 θ 的分布**。
- 贝叶斯信心



$$P(A)=\frac{\text{事件}A\text{已发生的样本数}}{\text{样本空间的可能总样本数}}\approx\frac{M}{N}$$



贝叶斯主义

1. **信念的种子：基础比率的力量**

从一个主观的先验概率开始，贝叶斯定理教会我们如何在不确定性中找到希望的起点。

2. **粗略也是一种智慧：行动在不完美中也能美好**

贝叶斯分析告诉我们，即使在信息不完全或模糊的情况下，也能做出有力的决策。

3. **流动的信念：持续更新的艺术**

贝叶斯思维强调信念不是静态的，而是一个随时间和数据不断更新和适应的动态过程。

4. **简约与全面：奥卡姆剃刀与多维证伪**

贝叶斯分析教我们如何在复杂性和简单性之间找到平衡，同时从多个角度审视问题。

5. **因果的新语言：概率作为解释工具**

通过贝叶斯分析，我们可以用概率作为一种新的工具来理解和解释因果关系。

6. 知识的三重旋律：经验、探索和更新

贝叶斯思维强调知识是基于经验、通过试探获得，并随着新信息而不断更新的。

7. 智慧的进化：不断逼近真相

贝叶斯方法教会我们如何通过不断的自我修正和更新，逐渐接近真相或最优解。

8. 联结的力量：贝叶斯网络与分布式思维

类似于我们大脑的原理，贝叶斯网络展示了如何通过联结和分布式思维来解决复杂问题。

9. 连接定义了事物：联结的权重

在贝叶斯世界中，不仅你和谁连接重要，而且连接的“权重”或质量同样重要。

10. 模型的双面性：在相信与怀疑之间寻找平衡

一个贝叶斯主高手，能够在相信中怀疑，在怀疑中相信，并在一个充满不确定性的世界里，持续前行。

克伦威尔原则（Cromwell's rule）

I beseech you, in the bowels of Christ, think it possible that you may be mistaken.

named by statistician Dennis Lindley:

the use of prior probabilities of 1 (“the event will definitely occur”) or 0 (“the event will definitely not occur”) should be avoided

Arithmetic and logic, alters the first rule of probability, or the convexity rule,

$$0 \le Pr(A) \le 1$$

to

$$0 < Pr(A) < 1$$