神奇的模式概率与“鞅”

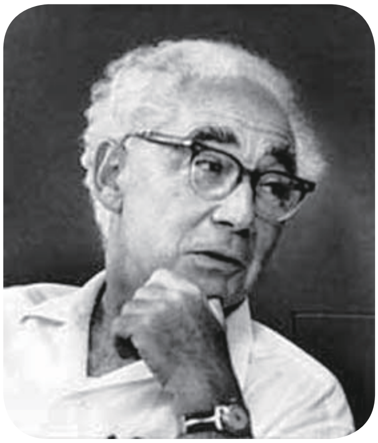
李硕彦

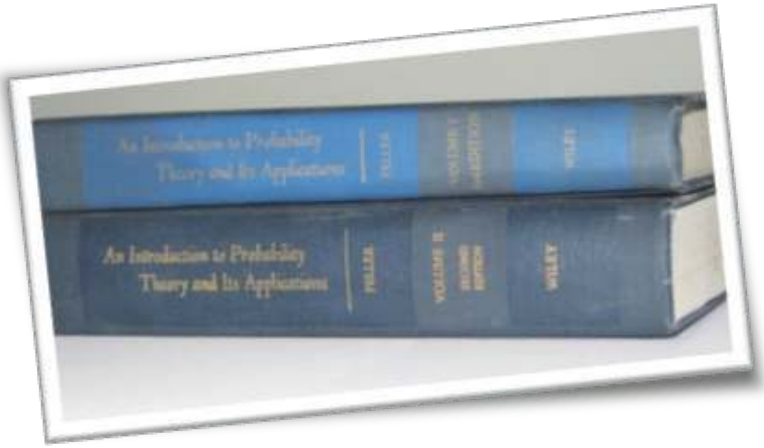
违反直觉的模式概率

William Feller 的经典教科书《概率论及其应用》提到一个试验：重复抛掷一个标准硬币，一直到连续 6 次出现人头。用 H（= Head）代表硬币的人头那面，T（= Tail）代表尾巴，连续出现 6 个 H 所构成的模式就表示为 HHHHHH。抛掷硬币的次数就叫作这个模式的“等待时间”，它是一个随机变量，最小值是 6，最大值无限。Feller 问：这个等待时间的均值有多大？

标准硬币的 H 和 T 出现的概率均等，所有 $2^6 = 64$ 个由 H 和 T 构成的长度为 6 的模式都有相同的机会出现。如果抛掷硬币 64,000,000 次，每一个模式都会出现 1,000,000 次左右。换句话说，平均来讲每掷 64 次就每一个模式都会出现一次。所以我们试着推论：

- 所有 64 个模式的平均等待时间都是 64。





William Feller 生前并没有完成这部上下两册的概率论经典。他死后，学生帮忙作最后的编辑、付印

真的可以这样推论吗？ Feller 不放心，于是将这个问题直接当成马尔科夫链，细心地作了冗长的计算，结果是：

- HHHHHH 模式的平均等待时间不是 64，而是 126。

这否定了前面的推论。Feller 又尝试另一个长度为 6 的模式：

- HHTTHH 模式的平均等待时间不是 64，也不是 126，而是 70。

Feller 称自己这些计算的结果“违反直觉”。

真理和童话愉快地吻合，被冷落一旁的是直觉

当科学真理与直觉相违背的时候，问题自然是在直觉那一方。平均来说，每抛掷硬币 64 次，任何一个长度为 6 的模式都会见到一次，这是正确的。用较为严谨的数学语言来说：

- 每一个长度为 6 的模式的平均“刷新时间”都是 64。

刷新时间指的是从这个模式出现到下一次再次出现的时间，它也是一个随机变量。譬如说，抛掷硬币出现了这个序列：

…HTHHTTHHTHHTTTTHTHTTTTHHHHTTHHTTHHH…

其中三个红色的 H 代表 HHTTHH 模式完成的时候。从第一次完成这个模式之后，又掷了硬币 21 次才出现第二次，这时候，刷新时间这个随机变量就取值为 21。之后又掷了 4 次就出现第三次。这回，刷新时间取值为 4。

HHTTHH 模式的刷新时间的最小值是 4，而等待时间的最小值是 6。两个随机变量明显地不同，而等待时间只可能偏大。只要你的直觉将两者区分开来，结论就应该是：

- HHTTHH 模式的平均等待时间 > 64，其他等长模式的平均等待时间 ≥ 64。

Feller 算出 HHTTHH 的平均等待时间为 70，完全没有矛盾。

计算概率的时候，直觉是最重要的，但也是最易引入陷阱的。底下我们再用一个例子来看一般的直觉和科学真理可以相去多远。还是连续抛掷一个标准硬币，看 THTH 和 HTHH 这两个模式之间“哪一个先出现，也就是说，让这两个模式“赛跑”。从平均等待时间来看，THTH 是 20，而 HTHH 是 18。也就是说，THTH 跑得比较慢，HTHH 比较快。那么，

- 这场赛跑的胜算比 (odds) 是不是应该很接近于 9 比 10 呢？

答案是否定的，其实这是一面倒的竞赛，正确的胜算比是 9 比 5。也就是说，双方赢的概率各自是 9/14 和 5/14。看完本文之后，读者就能很容易地算出这个胜算比，当然也能立刻得出 THTH、HTHH、HHHHHH、HHTTHH 的平均等待时间，而且完全不需要马尔科夫链的冗长计算。

在解释这个一面倒的现象之前，还有一点小小的补充：

- 这个一面倒的胜算比是倒向比较慢的那个模式。平均每 14 场赛跑，THTH 赢 9 次，比较快的模式 HTHH 反倒只赢 5 次。

模式的概率就是这么神奇。有些时候，科学的真理乍听之下好像是童话故事。事实上，用童话故事来解释这个胜算比就再简单不过了：参与赛跑的 THTH 模式以 T 开头，代表乌龟（Tortoise）。另一个参赛者 HTHH 以 H 开头，代表野兔（Hare）。童话故事里，龟兔赛跑通常是乌龟胜，当然

也有说是野兔赢的，两种说法的比例大约是 9 比 5。所以真理和童话愉快地吻合了，被冷落一旁的是直觉。当然，这只是文字游戏，不是科学的解释，用来取悦小孩还可以。

参与赛跑的 THTH 模式以 T 开头，代表乌龟（Tortoise）。对手 HTHH 模式以 H 开头，代表野兔（Hare）。乌龟的平均等待时间是 20，而野兔是 18，所以乌龟跑得比较慢，野兔比较快。但是，绝大多数的时候是乌龟跑赢

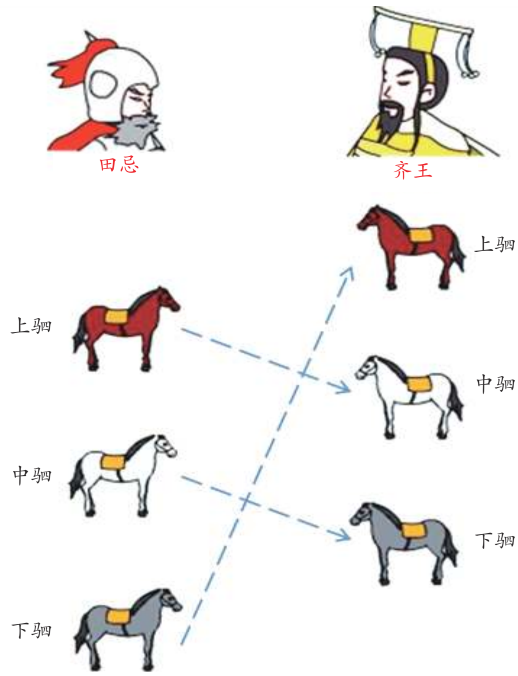
古今中外的田忌赛马

上述受冷落的直觉只是未受过训练的直觉。正确的直觉是什么？底下我们来看一个随机的长序列：

TTHTTHHTTHTHHTTHHHTTTTHTHHTHTTTHTTTTHTHHTHTT
THTHTHTTHTTHTHTTTTHTHHTHTTTHTHHHHTHHTTHTTHTHHT
HHTHHTHTTTHTTTTHTTTTTHTHHTHHTHHTHTTHTHHTHTTT
THTHHTHTTTTHTHHTHHTHTTTHHTHHTHTHTTTHTHTHHTHTHT
THTHHHHTHTHTTTHTTH…

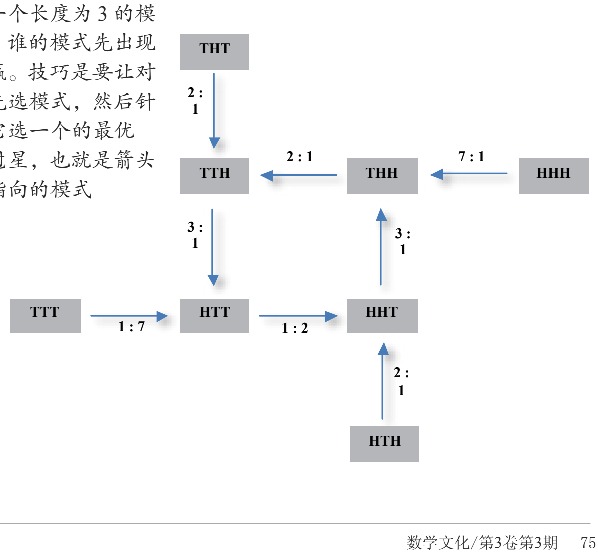
大致上，有十六分之一的字母是绿色的 H，用于表示 THTH 模式完成的时候。有十六分之一是红色的 H，用于表示 HTHH 模式完成的时候。两者的个数大致相等，但是分布的方式不同，有一半的绿 H 其后紧跟着红 H，但是在每一个红 H 之后的 3 个字母绝对不会是绿 H。所以，如果从长序列中随机取一点开始寻找有颜色的 H，第一个碰到的比较可能是绿的。如果用跑马场的语言来说，每次野兔跑赢的时候，都至少要赢一个马身，而乌龟赢的时候有一半是仅赢一个马颈。野兔赢的少，但都是大胜，乌龟赢的多，但常是小胜。合计起来，双方的实力其实相等。

战国时期，齐国的国王总是用赛马来赢公子田忌的钱。有一天，齐王又邀请田忌赛三场马，当时齐王拥有全国第 1、第 3、第 5 快马，而田忌有的是第 2、第 4、第 6。不敢拒绝国王的邀请，但是明摆的又要输三笔钱，田忌实感为难。这时，军事谋略家孙膑是田忌的座上客，他教田忌向齐王建议：加大赌注，然后三赛二胜的一方独赢。齐王欣然同意，等着收取一大笔横财，而且，即使意外地输一场也无妨，等于是买了双保险。比赛当天，田忌祭出孙膑的战略：“以上驷对中驷、中驷对下驷、下驷对上驷”。结果，齐王的上驷先胜一场，然后田忌小胜两场，从此齐王不再当田忌为自动提款机。这就是成语“田忌赛马”的由来。



THTH 常跑赢 HTHH 凭的就是孙膑的小胜多胜战略。1969 年的数学消遣杂志（J. Recreational Math）中，Walter Penney 也提到了掷硬币的赌博：两个人各选一个长度为 3 的模式，谁的模式先出现就赢。技巧是要让对手先选模式，然后针对它选一个的最优的克星。那个年代，美国有一个中学生就将这个游戏理论付诸实践，赢了同学们很多钱。只可惜，他的实践成果太过辉煌，差点被学校开除。

抛掷硬币，两个人各选一个长度为 3 的模式，谁的模式先出现就赢。技巧是要让对手先选模式，然后针对它选一个的最优的尅星，也就是箭头所指向的模式



顺便提一句，这个聪明的学生不是美国华人，更不是孙膑的后人。小胜多胜的直觉可以帮他赢钱，但仍不足以导出概率的公式。下面我们就探讨概率的公式。到目前为止，我们考虑的都是掷标准硬币的随机过程。这是一个“重复实验过程”，简称为“i.i.d. 过程”。这个特殊的 i.i.d. 过程叫作“对称的伯努利过程（symmetric Bernoulli process）”，因为它有下列的两个特性。

- 它是二元的，也就是说，过程中的随机变量只有两个可能的值。
- 它是对称的，也就是说，随机变量的两个值各占一半的概率。

但是，我们底下要计算概率的方式不受限于这两个特殊性，有它们在反而难以突出重点，所以，我们要改掷骰子，一个特殊的骰子。

理想中的公平赌场

将一个普通骰子的 6 面分别标明 a、a、a、b、b、c。这个骰子掷到 a、b、c 的概率就分别是 1/2、1/3、1/6。用数学符号来表达就是：

- $P\{a\} = 1/2$
- $P\{b\} = 1/3$
- $P\{c\} = 1/6$

重复掷这个骰子的随机过程依然是一个 i.i.d. 过程。但它既不是二元的也不是对称的，因为每次掷骰子有多于两种可能的结果，而且，$P\{a\}$、$P\{b\}$、$P\{c\}$ 不相等。

底下用 aba 这个模式为例来描述计算平均等待时间的方法，分为两个步骤。

第一步：计算所谓的“重合系数”ε_1、ε_2、ε_3。

- 拿模式的头 1 个字母与末 1 个字母相比较。如果相同，就令 $\varepsilon_1 = 1$，否则，$\varepsilon_1 = 0$。然后拿模式的头 2 个字母与末 2 个字母相比较。如果相同，就令 $\varepsilon_2 = 1$，否则 $\varepsilon_2 = 0$。再拿模式的头 3 个字母与末 3 个字母相比较。如果相同，就令 $\varepsilon_3 = 1$，否则 $\varepsilon_3 = 0$。
- 就 aba 这个模式而言，$\varepsilon_1 = 1$、$\varepsilon_2 = 0$、$\varepsilon_3 = 1$。

		a	b	a
a	b	a		

	a	b	a
a	b	a	

a	b	a
a	b	a

第二步：用一个繁分式求出答案：

$$\frac{\varepsilon_1+\dfrac{\varepsilon_2+\dfrac{\varepsilon_3}{P\{a\}}}{P\{b\}}}{P\{a\}}=\frac{1+\dfrac{0+\dfrac{1}{1/2}}{1/3}}{1/2}=2+12$$

这个繁分式代表 aba 模式的某种“自相关（autocorrelation）”，可以将它简写为 aba ∗ aba。迟些我们还会将它推广成两个模式之间的“互相关（crosscorrelation）”。

aba 模式的平均等待时间就是 aba ∗ aba。为什么呢？为了要解释这件事，我们需要设计一家特殊的赌场。每个赌客带 \$1 进场，按下列规则压注于 aba 这个模式：

- 骰子第一掷如果是 a，他的 \$1 就变成 $\$2 = \$1 / P\{a\}$ 并且连本带利继续压注；否则变 \$0。
- 骰子第二掷如果是 b，这 \$2 就变成 $\$6 = \$2 / P\{b\}$ 并且连本带利继续压注；否则变 \$0。
- 骰子第三掷如果是 a，这 \$6 就变成 $\$12 = \$6 / P\{a\}$；否则变 \$0。
- 一旦 aba 模式出现，整个赌场的作业就戛然终止。

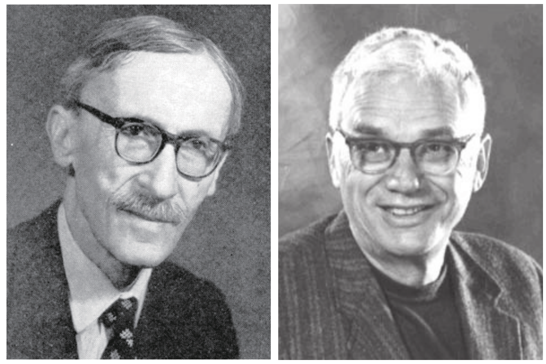
现实世界的赌场都要占赌客的便宜，一般在 5~6%。这里设计的却是个“公平赌场”，它与赌客公平对赌。

赌博团队带来的直觉

假设赌客们都同属一个团队，每掷一次骰子之前，就派一个人进场。那么，不论骰子掷出什么结果，最后整个团队台面上的钱一定是 \$14，包括最后一个队员的 \$2 和倒数第 3 个队员的 \$12。

	队员甲	队员乙	队员丙	队员丁	队员戊	队员己	队员庚
	$1						
掷出 a							
	$2	$1					
掷出 a							
	X	$2	$1				
掷出 b							
		$6	X	$1			
掷出 c							
		X		X	$1		
掷出 a							
					$2	$1	
掷出 b							
					$6	X	$1
掷出 a							
结果					$12		$2

在 7 个队员入场之后，骰子分别掷出 a、a、b、c、a、b、a。每一个人的金额的变化过程如表所列。任何情况下，aba 模式的等待时间这个随机变量，就等于入场的人数，也等于赌博团队带进场的本钱。经由鞅终止定理，就知道它的均值正好是赌博终止时台面上的 $14



左：Paul Levy 将“鞅”这个名词定义成一种随机过程，用来形容公平赌博；右：Joseph Doob 写下了严谨的“鞅终止定理”

因为是个公平赌场，所以团队净赢的金额平均为 $0。也就是说，团队带进场的本钱的均值就等于赌博终止时台面上的 $14 = $2 + $12。请注意：团队的本钱、进场的人数、aba 模式的等待时间都是同一个随机变量。结论是：

aba 模式的平均等待时间
= 2 + 12
= 上述的繁分式
= aba ∗ aba

更进一步说，繁分式里面的重合系数 $\varepsilon_1=\varepsilon_3=1$ 意指倒数第 1 和第 3 个队员是赢家，最后分别有 $2 和 $12。$\varepsilon_2=0$ 意指倒数第 2 个队员最终是输家。

以上“赌博团队”的计算技巧，建立于一个直觉之上：

- 在公平赌场净赢的金额平均为 0。

这个直觉，在概率论里面叫做“鞅终止定理（Martingale Stopping Theorem）”。鞅（martingale）是中古时期法国的一种竞赛。18 世纪时，经由赌博而联系到概率。20 世纪早期，Paul Levy 将这个名词定义成一种随机过程，用来形容公平赌博。鞅的观念和鞅终止定理在数学上用途很广。鞅终止定理虽属直观，但是数学上需要一些零碎的条件来保证“收敛性”，将这些充分条件简洁地陈述出来也是重要的工作。后来，Joseph Doob 写下了严谨的鞅终止定理。

最一般性的公式

接下来，我们要考虑一个很宽泛的问题。重复任何一个实验，实验的可能结果的个数不受限制。概率的分布也不受限制，可以是非对称的。任给 n 个模式 A_1、A_2、⋯、A_n，让这 n 个模式赛跑，它们胜出的概率分别用 p_1、p_2、⋯、p_n 来代示。各个模式的长度不受限制，而且可以各自不同。重复实验直至有一个胜者跑出，以随机变量 N 代表实验的次数，它的均值就写作 EN。我们自然地问：

- $EN=?\ \ p_1=?\ \ p_2=?\ \cdots\ p_n=?$

这里可以假设参赛的任何一个模式不会是另一模式的一截，否则可以剔除后者而不影响结果。

这看似宽泛而艰难的问题，却有一个极简单的答案。首先，定义模式与模式之间的一个算子（operator），以“ ∗ ”这个符号表示，用来描述两个模式之间的某种互相关或者单一模式的自相关。譬如说掷骰子产生的 aba 模式的自相关就写成 aba ∗ aba = 14。又譬如说，第一个模式是掷骰子产生的 bcaca，第二个是 acab，两者之间的互相关就定义为以下的繁分式：

$$\text{bcaca}*\text{acab}=\cfrac{\varepsilon_1+\cfrac{\varepsilon_2+\cfrac{\varepsilon_3+\cfrac{\varepsilon_4}{\mathrm{P}\{b\}}}{\mathrm{P}\{a\}}}{\mathrm{P}\{c\}}}{\mathrm{P}\{a\}}=\cfrac{1+\cfrac{0+\cfrac{1+\cfrac{0}{1/3}}{1/2}}{1/6}}{1/2}=2+24$$

其中重合系数 $\varepsilon_1=1$，因为 bcaca 末 1 个字母与 acab 的头 1 个字母相吻合。$\varepsilon_2=0$，因为 bcaca 末 2 个字母与 acab 的头 2 个字母不吻合。其余类推。

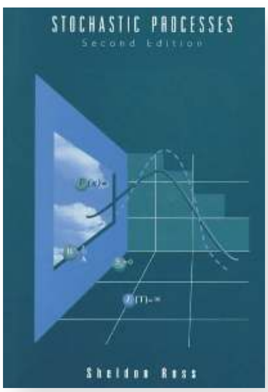
有了 ∗ 这个算子，就可以导出 n+1 个联立一次方程式来解出 EN、p_1、p_2、⋯、p_n 这 n+1 个未知数，写成矩阵形式，就是：

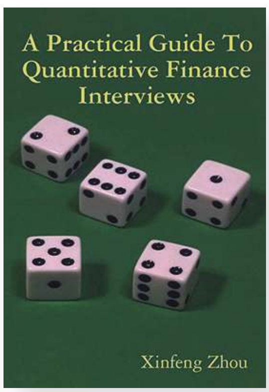
$$\begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ -1 & A_1*A_1 & A_2*A_1 & \cdots & A_n*A_1 \\ -1 & A_1*A_2 & A_2*A_2 & \cdots & A_n*A_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -1 & A_1*A_n & A_2*A_n & \cdots & A_n*A_n \end{pmatrix}\begin{pmatrix} EN \\ p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}=\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

这个方程式的证明过程不是很难。当 $n = 1$，这个方程式就算出模式 A_1 的平均等待时间为 $A_1 * A_1$。当 $n = 2$，这个方程式就算出模式 A_1 对 A_2 的胜算比，同时也算出 A_1 与 A_2 之间的胜者的平均等待时间。

愈是一般性的数学理论，愈是简洁清楚

本文中的数学内容，包括上述的的龟兔赛跑、公平赌场、赌博团队的计算技巧、矩阵方程式，皆取材自概率年鉴（The Annals of Probability）1980 年刊登的一篇拙作，大家简称它为“模式鞅论文（Martingale of patterns paper）”。各类文献也从该文节录类似的内容，包括 Sheldon Ross 的教科书：“Stochastic Processes”、Xinfeng Zhou 的参考书：“A Practical Guide To Quantitative Finance Interviews”，以及“数学与工程的对话”第 5 讲：Martingale of Patterns。模式鞅在数学与工程之间架起一座小桥，它可以用于基因工程、无线 Ad Hoc 网络、财经工程、信息安全等领域，也可以简化“随机行走”的计算，读者可参阅上述诸文献，皆包括于右列参考资料。





除了上述计算 HHHHHH 和 HHTTHH 的平均等待时间，Feller 也考虑过二元但非对称的概率。假设 P{H} = p，P{T} = q =1 – p，让两个模式赛跑，一边是连续 m 个 H，另一边是连续 n 个 T。Feller 花了功夫算出 m 个 H 跑赢的概率是：

$$\frac{p^{m-1} - p^{m-1}q^{n}}{p^{m-1} + q^{n-1} - p^{m-1}q^{n-1}}$$

另外，1974 年《科学美国人》杂志的“数学游戏”专栏发表了 John Conway 的结果：在抛掷标准硬币的时候，也就是二元且对称的情况下，拿 A、B 两个模式之间的重合系数串成一个二元整数 binary($\varepsilon_k \cdots \varepsilon_2 \varepsilon_1$)，以 $L(A, B)$ 代表。如此，$L(A, A)$ 就会是 A 模式平均等待时间的一半，同时 A、B 两个模式赛跑的胜算比就是：

$$[L(A, A) – L(A, B)] : [L(B, B) – L(B, A)]$$

据说用了长达十数页的证明。

上述矩阵方程式是最一般性的公式，非常简洁而证明也直观，相比之下，各种极其特殊的情况给出的公式看起来反而不是那么透明。数学里头，愈是一般性的理论往往愈是简洁清楚，关键是要从新的角度来看问题。最有趣的新角度往往来自原本不相关的数学观念。

不止数学本身如此，数学在工程上的应用也常是如此。直觉上，能用于工程的数学总该包含一些计量的工具，譬如坐标、函数、向量、微分、积分等等。可是愈来愈多不需要数字的数学也和工程对起话来。愈是反直觉的搭档，其效果往往就愈神奇。

参考资料

1. J. L. Doob, Stochastic Processes, Wiley, New York, 1953.
2. M. Gardner, “Mathematical Games,” Scientific American, vol. 10, page 120-125, 1974.
3. W. Feller, An Introduction to Probability Theory and Its Applications, Vol. 1 and 2, Wiley, New York, 1966.
4. S.-Y. R. Li, ““A Martingale Approach to the Study of Occurrence of Sequence Patterns in Repeated Experiments,” The Annals of Probability, Vol. 8, No. 6, 1980.
5. S.-Y. R. Li, “数学与工程的对话，第 5 讲：Martingale of Patterns,” http://www.ie.cuhk.edu.hk/people/A_Dialogue_between_Math_and_Engineering1_016.zip, 2009.
6. W. Penney, “Problem: Penney-ante,” Journal of Recreational Mathematics, Vol. 2, Page 241, 1969.
7. S. Ross, Stochastic processes, 2nd Edition, John Wiley & Sons, New York, 1996.
8. X. Zhou, A Practical Guide To Quantitative Finance Interviews, Morrisville, 2008.

作者简介：李硕彦，英文名简称 Bob Li，香港中文大学信息工程学系讲座教授，台大数学学士、加州伯克利大学数学博士，网络编码理论创建人之一。因对博弈论的理论贡献，曾被电影《21 点》原型团队邀请参与其冒险赌博行为。