The Ace of Hearts Method

1. History. At the ICM in Helsinki in 1978, Alf van der Poorten idly watched John Conway lose repeatedly at backgammon. When play was terminated, the great JHC owed his opponent 370 Finnish Marks. John pulled a 1000 markka note from his pocket, but no one present could make the necessary change. Another onlooker observed, "This calls for the ace of hearts method." Someone produced a coin, and flipped it. John handed the 1000 markka note to his opponent. The coin was flipped again, and the opponent handed the bill back. The coin was flipped a third time, John pocketed the 1000 markka note, and they all departed happily. Alf was left to ponder what it was he had just witnessed.

2. The problem. Devise a simple procedure, based on a coin that comes up heads with probability 1/2, that produces a positive outcome with probability exactly α, for any preassigned α, $0 \le \alpha \le 1$.

3. The procedure. Make a sequence of double-or-nothing bets. Suppose that A owes α to B where $0 \le \alpha < 1$, and but the amount actually paid will either be 0 or 1. If $\alpha \ge 1/2$, then payment is made immediately, with the result that B owes $1-\alpha$ to A; note that $0 \le 1-\alpha \le 1/2$. The person holding the bill calls the flip of a coin. If he calls correctly, then he owes nothing. If he calls incorrectly, then his debt is doubled. The cycle is repeated until the person holding the money calls the toss correctly.

In the historical case recalled above, Conway evidently lost the first bet, which meant that he owed 740 forints. Since this is more than half of 1000, he handed over the bill. His opponent then owed him 260 forints. The opponent lost the next bet, with the result that he owed John 520 forints. He handed the bill back; John then owed 480 forints. John won the next bet, which settled the issue.

4. Analysis. Your friendly professional probabilist will tell you that a double-or-nothing bet is fair; hence the debt is paid with probability α. We amateurs, however, feel the need for something a little more detailed. In the discussion that follows, suppose that $0.b_1\, b_2\, b_3 \ldots$ is the binary expansion of α, which is to say that $\alpha = b_1/2 + b_2/4 + b_3/8 + \cdots$ with each $b_i = 0$ or 1. Suppose that the toss was incorrectly called on the first $k-1$ tosses, and that the money has exchanged hands, if necessary, so that the participants are ready for the k^{th} toss. We claim:

If $b_k = 0$, then A holds the money, and owes

$$2^{k-1}\alpha - \left[2^{k-1}\alpha\right] = \left\{2^{k-1}\alpha\right\} = \frac{b_k}{2} + \frac{b_{k+1}}{2^2} + \cdots .$$

If $b_k = 1$, then B holds the money, and owes

$$\left[2^{k-1}\alpha\right] + 1 - 2^{k-1}\alpha = 1 - \left\{2^{k-1}\alpha\right\} = \frac{1-b_k}{2} + \frac{1-b_{k+1}}{2^2} + \cdots .$$

Here $[u]$ denotes the integer part of u, which is to say that $[u]$ is the largest integer note exceeding u. In other words, $[u]$ is the unique integer such that $[u] \le u < [u] + 1$. The *fractional part* of u is $\{u\} = u - [u]$.

To prove the claim we argue by induction on k. Clearly the claim is correct when $k = 1$. Suppose the claim is correct for k. If $b_k = 0$, and A calls the toss incorrectly, then he owes twice as much, namely the amount

$$\frac{b_{k+1}}{2} + \frac{b_{k+2}}{2^2} + \cdots .$$

This number lies in the interval $[0, 1/2)$ or in $[1/2, 1)$ according as $b_{k+1} = 0$ or 1. Thus if $b_{k+1} = 0$, then A retains the money, and is ready for round $k + 1$. If $b_{k+1} = 1$, then A hands the bill to B, and B owes

$$1 - \Big(\frac{b_{k+1}}{2} + \frac{b_{k+2}}{2^2} + \cdots\Big) = \frac{1 - b_{k+1}}{2} + \frac{1 - b_{k+2}}{2^2} + \cdots ,$$

which is in accordance with the claim. Now suppose that $b_k = 1$, and that B calls the toss incorrectly. Then B owes twice as much, which is

$$\frac{1 - b_{k+1}}{2} + \frac{1 - b_{k+2}}{2^2} + \cdots .$$

This lies in $[0, 1/2)$ or in $[1/2, 1)$ according as $b_{k=1} = 1$ or $b_{k=1} = 0$. Thus if $b_{k+1} = 1$, then we are ready for round $k + 1$, while if $b_{k+1} = 0$, then B hands the bill back to A, and A owes B the amount

$$1 - \Big(\frac{1 - b_{k+1}}{2} + \frac{1 - b_{k+2}}{2^2} + \cdots\Big) = \frac{b_{k+1}}{2} + \frac{b_{k+2}}{2^2} + \cdots .$$

Thus the claim is established for $k + 1$, and the induction is complete.

5. Probability. On the basis of the above insights, it is now an easy exercise to determine the probability that B ends up with the money. Let E denote this event. Our sample space consists of the outcomes of the bets, each one of which may be won (W) or lost (L); we continue until a bet is won. Thus the space is $W,\ LW,\ LLW,\ \ldots,\ L^{k-1}W, \ldots,$ and finally the unlikely event L^∞. Since the sample space is partitioned into these various cases, we see that

$$\begin{aligned} P(E) &= P(E \cap W) + P(E \cap LW) + P(E \cap LLW) + \cdots \\ &= P(E|W)P(W) + P(E|LW)P(LW) + P(E|LLW)P(LLW) + \cdots . \end{aligned}$$

As $P\big(L^{k-1}W\big) = 2^{-k}$, the above is

$$= \frac{P(E|W)}{2} + \frac{P(E|LW)}{2^2} + \frac{P(E|LLW)}{2^3} + \cdots .$$

In the preceding paragraph we found that B wins the money, say on round k, precisely when $b_k = 1$. That is, $P\big(E|L^{k-1}W\big) = b_k$. Thus the above sum is exactly

$$\frac{b_1}{2} + \frac{b_2}{2^2} + \frac{b_3}{2^3} + \cdots = \alpha\,.$$

6. Efficiency. Let X denote the number of tosses needed to decide the issue. Then X is a geometric random variable with parameter $p = 1/2$, so the expected value of X is $E[X] = 2$. It seems remarkable that any desired threshold α can be fairly measured in so few steps.

7. Alternative procedures. A die with 100 numbered faces and the property that it comes up on each face with probability 1/100 would seem to be hard to construct. In any case, while it would deal with cents of a dollar, it would not accurately measure $\alpha = 1/3$, much less $\alpha = 1/\sqrt{2}$. One could ask a 'random' number generator to produce a number β, uniformly distributed in $[0, 1]$, and then A pays B if $0 \le \beta \le \alpha$. In the absence of a random number generator, one could flip a coin to determine the number β through its binary expansion, say $\beta = 0.d_1\, d_2\, \ldots = d_1/2 + d_2/2^2 + \cdots$. One would continue until one can distinguish which is the larger of α and β. Suppose $b_i = d_i$ for $1 \le i < k$. If $b_k = 0$ and $d_k = 1$, then $\alpha < \beta$, and A keeps the money. If $b_k = 1$ and $d_k = 0$, then $\beta < \alpha$, and A pays B. This, of course, is equivalent to the ace of hearts method, but lacks the immediacy and charm.

8. Final question. What on earth does any of this have to do with the ace of hearts?

Robin's Airplane

An airplane has n seats, and is fully booked with n passengers. Each passenger is assigned to one particular seat, no two to the same seat. The passengers enter the plane one at a time. The first passenger who enters, picks a seat at random, and sits in it. After that, each passenger on entering sits in his assigned seat if it is free, and otherwise chooses a seat randomly from among the remaining empty seats. Let A denote the event that the last passenger sits in the seat to which he was assigned.

1. Show that $P(A) = 1$ when $n = 1$.

2. Show that $P(A) = 1/2$ when $n = 2$.

3. From now on, suppose that $n > 1$, and let B denote the event that the seat assigned to the first passenger is taken before the seat assigned to the last passenger. Show that $P(B) = 1/2$.

4. Show that if the seat assigned to the first passenger is taken before the seat assigned to the last passenger, then the last passenger sits in his assigned seat. (That is, $B \subseteq A$.)

5. Show that if the last passenger sits in his assigned seat, then the seat assigned to the first passenger was taken earlier. (That is, $A \subseteq B$.)

6. Conclude that $P(A) = 1/2$ for all $n > 1$.

Andrew's Special Dice

Suppose we have three dice, called A, B, C, whose faces bear the following numbers:

A: 1, 1, 6, 6, 8, 8
B: 3, 3, 5, 5, 7, 7
C: 2, 2, 4, 4, 9, 9

We assume that when one of these dice is tossed, each face comes up with probability 1/6. Suppose these dice are tossed. Let s_A denote the value shown by die A, s_B the value shown by die B, and s_C the value shown by C.

1. Find $P(s_A > s_B)$.

2. Find $P(s_B > s_C)$.

3. Find $P(s_C > s_A)$.

4. Find $P(s_A > s_B > s_C)$.

5. Find $P(s_A > s_C > s_B)$.

6. Find $P(s_B > s_A > s_C)$.

7. Find $P(s_B > s_C > s_A)$.

8. Find $P(s_C > s_A > s_B)$.

9. Find $P(s_C > s_B > s_A)$.

10. Answers in 4–9 should sum to 1. Do they?

11. Three of the answers in 4–9 should sum to the quantity in 1. Which three?

12. Three of the answers in 4–9 should sum to the quantity in 2. Which three?

13. Three of the answers in 4–9 should sum to the quantity in 3. Which three?

Moments of binomial random variables

Let X be a binomial random variable with parameters n and p. Our object is to determine moments $E\big[X^r\big]$ in terms of n and p. We first show that if r is a positive integer, then

$$E[X(X-1)\cdots(X-r+1)] = p^r n(n-1)\cdots(n-r+1)\,. \tag{1}$$

To show this, we first note that if $n < r$, then both sides above are 0. Hence we may assume that $n \ge r$. For such n, the left hand side above is

$$= \sum_{k=0}^{n} k(k-1)\cdots(k-r+1)\binom{n}{k}p^k(1-p)^{n-k}\,.$$

If $k < r$, then the summand above is 0, so we may restrict k to the interval $r \le k \le n$. Thus the above is

$$= \sum_{k=r}^{n} k(k-1)\cdots(k-r+1)\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=r}^{n} k(k-1)\cdots(k-r+1)\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

$$= \sum_{k=r}^{n} \frac{n!}{(k-r)!(n-k)!}p^k(1-p)^{n-k} = p^r n(n-1)\cdots(n-r+1)\sum_{k=r}^{n}\frac{(n-r)!}{(k-r)!(n-k)!}p^{k-r}(1-p)^{n-k}$$

On setting $k - r = \ell$, we find that the last sum on the right above is

$$\sum_{\ell=0}^{n-r}\frac{(n-r)!}{\ell!(n-r-\ell)!}p^\ell(1-p)^{n-r-\ell} = \sum_{\ell=0}^{n-r}\binom{n-r}{\ell}p^\ell(1-p)^{n-r-\ell} = \big(p+(1-p)\big)^{n-r} = 1$$

by the binomial theorem. Thus we have (1).

Next we note that if f and g are real-valued functions and X is a discrete random variable, then

$$\begin{aligned} E[f(X)+g(X)] &= \sum_i \big(f(x_i)+g(x_i)\big)p(x_i) = \sum_i f(x_i)p(x_i) + \sum_i g(x_i)p(x_i) \\ &= E[f(X)] + E[g(X)]\,. \end{aligned} \tag{2}$$

This is a special case of a more general result: $E[X+Y] = E[X] + E[Y]$ for arbitrary random variables.

By taking $r = 1$ in (1), we see that $E[X] = np$. By taking $r = 2$, we find that $E[X(X-1)] = p^2 n(n-1)$. By (2) it follows that

$$E\big[X^2\big] = E[X(X-1)+X] = E[X(X-1)] + E[X] = p^2n(n-1) + pn = n^2p^2 + np(1-p)\,.$$

Hence

$$\operatorname{Var}(X) = E\big[X^2\big] - E[X]^2 = n^2p^2 + np(1-p) - (np)^2 = np(1-p)\,.$$

For $r = 3$ we find that $E[X(X-1)(X-2)] = p^3n(n-1)(n-2)$ and $X^3 = X(X-1)(X-2) + 3X(X-1) + X$, so

$$E\big[X^3\big] = p^3n(n-1)(n-2) + 3p^2n(n-1) + pn\,.$$

For general r the product $x(x-1)\cdots(x-r+1)$ can be expanded, so there are integers $\left[{r \atop j}\right]$, known as *Stirling numbers of the first kind* such that

$$x(x-1)\cdots(x-r+1) = \sum_{j=1}^{r} \left[{r \atop j}\right] x^j\,.$$

The numbers $\left[{r \atop j}\right]$ are sometimes denoted $s(r,j)$. They can be generated by the Pascal-like recursion

$$\left[{r \atop j}\right] = \left[{r-1 \atop j-1}\right] - (r-1)\left[{r-1 \atop j}\right].$$

In the reverse direction there exist numbers $\left\{{r \atop j}\right\}$, known as *Stirling numbers of the second kind* such that

$$x^r = \sum_{j=1}^{r} \left\{{r \atop j}\right\} x(x-1)\cdots(x-j+1)\,.$$

The $\left\{{r \atop j}\right\}$ are sometimes denoted $S(r,j)$. They can be generated by the Pascal-like recursion

$$\left\{{r \atop j}\right\} = j\left\{{r-1 \atop j}\right\} + \left\{{r-1 \atop j-1}\right\}.$$

Stirling numbers arise in combinatorics: $(-1)^{r-j}\left[{r \atop j}\right]$ is the number of permutations of r objects that have exactly j cycles. The number $\left\{{r \atop j}\right\}$ is the number of ways of partitioning a set of r objects into exactly j nonempty subsets.

$n\backslash k$	1	2	3	4	5	6
1	1					
2	−1	1				
3	2	−3	1			
4	−6	11	−6	1		
5	24	−50	35	−10	1	
6	−120	274	−225	85	−15	1

TABLE 1. Stirling numbers $\left[{n \atop k}\right]$ of the first kind.

$n\backslash k$	1	2	3	4	5	6
1	1					
2	1	1				
3	1	3	1			
4	1	7	6	1		
5	1	15	25	10	1	
6	1	31	90	65	15	1

TABLE 2. Stirling numbers $\left\{{n \atop k}\right\}$ of the second kind.

Non-attacking Rooks: An Exercise in Conditioning

Suppose that for $k = 1, 2, \ldots, 8$ we are to place k rooks randomly in k different locations on an 8×8 chess board. What is that probability that they are mutually non-attacking? Without conditioning, we would argue that there are $\binom{64}{k}$ ways of placing the rooks. There are $\binom{8}{k}$ ways of choosing the k rows that they are to fall in. In the first row chosen, there are 8 possible columns where the rook can be placed. In the second chosen row there are 7 columns available, and so on, so the probability is

$$\frac{\binom{8}{k} 8 \cdot 7 \cdots (8-k+1)}{\binom{64}{k}} = \frac{8!^2 (64-k)!}{(8-k)!^2 64!} .$$

Suppose we argue by conditioning. Let N_i denote the event that the i^{th} rook is not attacked by any of the first $i-1$ rooks. Given that none of the first $i-1$ rooks is attacking any other, that is, given $N_1 N_2 \cdots N_{i-1}$, there are $9-i$ rows that the i^{th} rook can lie in, and the same number of columns, giving $(9-i)^2$ acceptable locations out of the $65-i$ empty cells. Thus

$$P(N_i | N_1 N_2 \cdots N_{i-1}) = \frac{(9-i)^2}{65-i} .$$

Consequently, by the multiplication rule of conditioning,

$$\begin{aligned} P(N_1 N_2 \cdots N_k) &= P(N_1) P(N_2|N_1) P(N_3|N_1 N_2) \cdots P(N_k | N_1 \cdots N_{k-1}) \\ &= \frac{8^2}{64} \cdot \frac{7^2}{63} \cdot \frac{6^2}{62} \cdots \frac{(9-k)^2}{65-k} \\ &= \frac{8!^2}{(8-k)!^2} \cdot \frac{(64-k)!}{64!} . \end{aligned}$$

Thus

$$\begin{aligned} P(N_1) &= 1 = 1.00000000 \\ P(N_1 N_2) &= \frac{7}{9} = 0.77777777 \\ P(N_1 N_2 N_3) &= \frac{14}{31} = 0.4516129032 \\ P(N_1 N_2 N_3 N_4) &= \frac{350}{1891} = 0.1850872554 \\ P(N_1 N_2 N_3 N_4 N_5) &= \frac{280}{5673} = 0.04935660145 \\ P(N_1 N_2 N_3 N_4 N_5 N_6) &= \frac{840}{111569} = 0.007528973102 \\ P(N_1 N_2 N_3 N_4 N_5 N_6 N_7) &= \frac{1680}{3235501} = 0.0005192395243 \\ P(N_1 N_2 N_3 N_4 N_5 N_6 N_7 N_8) &= \frac{560}{61474519} = 0.000009109465338 \end{aligned}$$

Simpson's Paradox

To exemplify the phenomenon under discussion, we begin by considering the batting of two baseball players. Suppose that Jackie is at bat 100 times, during which he makes 30 hits for a proportion 0.30. On the other hand, Joe is also at bat 100 times but only collects 27 hits, for a proportion 0.27. Of course a batter's performance may depend on whether the pitcher throws left-handed or right-handed, so we consider how Jackie and Joe fare in these two cases:

	AGAINST LEFTIES			AGAINST RIGHTIES		
	At Bats	Hits	Ave	At Bats	Hits	Ave
Jackie	25	5	.20	75	25	.33
Joe	80	20	.25	20	7	.35

Thus Joe performs better against both left-handed and right-handed pitching, even though Jackie had the better overall performance. Although this might seem surprising at first, the explanation is easy to find. Both batters performed better against right-handed pitching, but Jackie was fortunate to have most of his batting against such pitchers, while Joe faced far more left-handed pitchers. The figures we have used here are fictitious, but Simpson's paradox does arise frequently this way in baseball statistics.

To consider Simpson's Paradox more abstractly, suppose that A and B are two events, and that $F_1, F_2, \ldots, F_n$ is a partitioning of our probability space into pairwise disjoint events. If $P(A|F_i) > P(B|F_i)$ for all i, then it follows that $P(A) > P(B)$, since by conditioning we see that

$$P(A) = \sum_{i=1}^{n} P(A|F_i)P(F_i) \geq \sum_{i=1}^{n} P(B|F_i)P(F_i) = P(B).$$

But this is not Simpson's Paradox. In Simpson's Paradox we are given that $P(X|AF_i) > P(X|BF_i)$, but these inequalities do not imply that $P(X|A) > P(X|B)$.

Our first example of Simpson's Paradox seemed quite reasonable, but now we consider a second hypothetical example that is a littler harder to live with. Suppose we are treating a life-threatening disease. We have to choose between an old method of treatment and a new one. To assess whether the new treatment is better than the old, we take 80 patients, and treat 40 of them in the new way, and 40 of them in the old. We find that among the patients treated in the new way, 20 are cured and 20 die, while among the patients treated in the old way, 24 are cured and 16 die.

	cured	died
new	20	20
old	24	16

Here the old cure-rate, 60%, was better than the new, 50%. But now we make a finer analysis of the same data by examining separately male and female patients. Suppose that when this is done, the following numbers emerge:

	MALES			FEMALES	
	cured	died		cured	died
new	8	2	new	12	18
old	21	9	old	3	7

(OVER)

Among male patients the new cure rate, 80%, is better than the old cure rate, 70%. Also, among female patients, the new cure rate, 40%, is better than the old, 30%. Thus a man would prefer the new treatment, and so would a woman, but if patients in general are treated in the new way then fewer will survive.

The instances of Simpson's Paradox we have considered thus far have been hypothetical, but the phenomenon has been noted many times in real life data. We mention a few examples.

- Comparison of TB deaths in 1910 in New York City versus Richmond, Virginia reveal that the mortality was lower in New York City. However, the mortality among whites was higher in New York City, and the mortality among blacks was also higher in New York City.

- From January 1979 to February 1979, the renewal rate of subsriptions to *American History Illustrated* rose from 51.2% to 64.1%. However, when the renewals were broken down into disjoint categories (gift, previous renewal, subscription service, catalog order), every category showed a lower renewal rate.

- The US Federal Income Tax rates between 1974 and 1978 decreased in every income category, but rose overall.

Many learned papers have been written in an effort to explain Simpson's Paradox (for a start, see the March 1976 issue of *Scientific American*, or C. R. Blyth, On Simpson's Paradox and the Sure-Thing Principle, *Jour. Amer. Stat. Assoc.* **67** (1972), 364–381). Perhaps the simplest useful observation is that The Law of Total Probability expresses a probability $P(X|A)$ as a weighted average of conditional probabilities $P(X|AF_i)$,

$$P(X|A) = \sum_i P(X|AF_i)P(F_i|A).$$

This weighted average may lie anywhere between the least or the largest of the conditional probabilities, depending on the weights $P(F_i|A)$. A similar formula applies to $P(X|B)$, but a disparity may arise if the weights $P(F_i|A)$ are large for those i for which $P(X|AF_i)$ is small, while the weights $P(F_i|B)$ emphasize those i for which $P(X|BF_i)$ is large.

Simpson's Paradox is also known as the Yule–Simpson Principle, as the Reversal Paradox, and as the Amalgamation Paradox. The Simpson in question is not the Thomas Simpson of London in the 1750's but rather one Edward H. Simpson who wrote about the phenomenon in 1951. However, it had been written about earlier, for example by statisticians Karl Pearson (1899) and Udny Yule (1903). The term "Simpson's Paradox" seems to have been coined by Blyth (1972).