INTRODUCTION TO PROBABILITY THEORY

[1] General Principles – Joint & conditional probability

[2] BAYES' THEOREM

[3] Permutations & Combinations

[4] Binomial, Poisson, & Gaussian Distributions

* [5] Probability Paradoxes & Scams

BASIC IDEAS

If an experiment is performed N times with $N_s \leq N$ "successes" we define the probability for success P as

$$P = \frac{N_s}{N} \qquad 0 \leq P \leq 1 \qquad (1)$$

It is assumed that $P \to$ constant as $N \to \infty$.

Fundamental Laws of Probability:

We wish to analyze the joint probabilities of having 2 things happen: For example you allow your children (assumed to be infinite in number!) to each take 2 jelly beans from a jar containing jelly beans of different colors. We can classify the various possibilities as follows:

A= picks RED Jelly bean
B= " BLACK " "

Let n_1 = number of times that A happens but not B
n_2 = " " " " B " " " A
n_3 = " " " both A and B happen
n_4 = " " " neither A nor B happens

In evaluating joint probabilities it is important that the enumeration of categories is both EXHAUSTIVE and EXCLUSIVE

EXHAUSTIVE ⇒ every possible outcome is accounted for

EXCLUSIVE ⇒ each outcome can be uniquely assigned to a category

To see how this works in the present case we enumerate all possibilities where X denotes a case where a color other than red or black is picked:

$$\begin{array}{llll} AA \to n_1 & \quad & XB \to n_2 & \\ AB \to n_3 & & BX \to n_2 & \\ BA \to n_3 & & XA \to n_1 & \\ BB \to n_2 & & AX \to n_1 & \\ & & XX \to n_4 & \end{array} \qquad (2)$$

Since this exhausts all possibilities we have: $n_1 + n_2 + n_3 + n_4 = n$

The probability P(A) of getting at least one RED jelly bean is

$$P(A) = \frac{\text{number of times A happens}}{n} = \frac{n_1 + n_3}{n} \qquad (3)$$

Similarly:

$$P(B) = \frac{n_2 + n_3}{n} \qquad (4)$$

Next we ask for the probability of A or B or both. Evidently this is the probability of getting at least one RED or BLACK. This is:

$$P(A+B) = \frac{n_1 + n_2 + n_3}{n} \qquad (5)$$

here + means "or"

The probability of both A and B occurring ≡ JOINT PROBABILITY distribution and is denoted by $P(AB)$

multiplication denotes "and"

Clearly in the present case we have $P(AB) = \dfrac{n_3}{n}$ (6)

The probability that A occurs given that B occurs is defined as the CONDITIONAL PROBABILITY ≡ $P(A|B)$. In the present example the number of times that B occurs is

n_2 (alone) + n_3 (in combination with A)

Hence: $$P(A|B) = \frac{n_3}{n_2 + n_3} = \frac{\text{B alone}}{\text{B total}} \qquad (7)$$

Similarly: $$P(B|A) = \frac{n_3}{n_1 + n_3} = \frac{\text{A alone}}{\text{A total.}} \qquad (8)$$

Note that $P(A|B) \neq P(B|A)$ in general.

We can infer from the results in (3)-(8) a set of general rules which apply to other cases as well. [We will not prove these rules rigorously here.]

$$\boxed{\begin{aligned} &P(A+B) = P(A) + P(B) - P(AB) && (9) \\ &P(AB) = P(B)\,P(A|B) = P(A)\,P(B|A) && (10) \end{aligned}}$$

We can use the previous results to verify the rules in (9), (10):

$$P(A+B) = \frac{n_1+n_2+n_3}{n} \;;\quad P(A) = \frac{n_1+n_3}{n} \;;\quad P(B) = \frac{n_2+n_3}{n}$$

$$P(AB) = \frac{n_3}{n} \;;\quad P(A|B) = \frac{n_3}{n_2+n_3} \;;\quad P(B|A) = \frac{n_3}{n_1+n_3} \qquad (11)$$

$$\therefore\; P(A+B) = \frac{n_1+n_2+n_3}{n} \stackrel{?}{=} \frac{n_1+n_3}{n} + \frac{n_2+\cancel{n_3}}{n} - \frac{\cancel{n_3}}{n} = \frac{n_1+n_2+n_3}{n} \checkmark \qquad (12)$$

<u>Comment:</u> We see from this example that the contribution $P(AB)$ which is subtracted out in (9) has the effect of eliminating the <u>"double counting" problem</u> which arises because the same events are counted in both $P(A)$ & $P(B)$. (We return to this later).

Consider next Eq. (10) for $P(AB)$. It is sometimes easier to remember this equation in the form:

$$P(AB) = P(A|B)\,P(B) = P(B|A)\,P(A) \qquad (13)$$

first A happens

then, given that A has happened, B happens

In any case, in our example:

$$P(AB) = \frac{n_3}{n} \stackrel{?}{=} \underbrace{P(A|B)}_{\frac{n_3}{n_2+n_3}} \cdot \underbrace{P(B)}_{\frac{n_2+n_3}{n}} = \frac{n_3}{n} \checkmark \qquad (14)$$

Also:

$$P(AB) = \frac{n_3}{n} \stackrel{?}{=} \underbrace{P(B|A)}_{\frac{n_3}{n_1+n_3}} \cdot \underbrace{P(A)}_{\frac{n_1+n_3}{n}} = \frac{n_3}{n} \checkmark \qquad (15)$$

APPLICATIONS: As usual with word problems, most of the difficulty is associated with identifying the variables in the problem with the quantities appearing in the equations. Here we give some examples:

[1] We are given 2 decks of cards, and 1 card is to be drawn from each. What is the probability that at least one card will be an ace?

Solution: As in the case of the jelly beans we form an exclusive and exhaustive list as follows:

$$
\begin{aligned}
n_1 &= \text{number of times a card from deck \#1 is an ace, but from deck 2 is not an ace.}\\
n_2 &= \text{" " " " " " \#2 " " " " " 1 " "}\\
n_3 &= \text{" " " both decks give an ace}\\
n_4 &= \text{" " " neither deck gives an ace}
\end{aligned}
\qquad (1)
$$

$$
\begin{aligned}
\text{Define } P(A_1) &= \text{probability of getting an ace from deck \#1} = \frac{4}{52} = \frac{1}{13}\\
P(A_2) &= \text{" " " " " " deck \#2} = \frac{1}{13}
\end{aligned}
\qquad (2)
$$

$P(A_1)$ is the total probability that deck # 1 gives an ace, irrespective of what happens in deck # 2 and hence $P(A_1) = 1/13$. Combining Eqs. (1) & (2) we find

$$P(A_1) = \frac{n_1+n_3}{n} = \frac{1}{13} \quad ; \quad P(A_2) = \frac{n_2+n_3}{n} = \frac{1}{13} \qquad (3)$$

From the previous discussion we want $P(A_1+A_2)$ which is the probability of an ace in deck # 1 or an ace in deck # 2. Then,

decks are independent

$$P(A_1+A_2) = P(A_1) + P(A_2) - P(A_1A_2) = \frac{1}{13} + \frac{1}{13} - \frac{1}{13} \times \frac{1}{13} = \frac{25}{169} \; ⊛ \qquad (4)$$

⊛ Note that we can see here the need for subtracting $P(A_1A_2)$: If we did not do this subtraction, then if we had > 13 decks, $P(A_1+A_2)$ would eventually exceed unity.

[2] What is the probability of drawing 2 hearts when 2 cards are drawn successively from a deck without replacement?

<u>Solution</u>: The first step is to recognize that we want to use the formula

$$\underbrace{P(2H)}_{\text{2 hearts}} = P(H|H)\,P(H) \qquad (1)$$

$P(H)$: probability of getting a heart.

$P(H|H)$: Conditional probability of getting a heart as the second card once a heart is drawn.

To simplify the discussion we use subscripts to indicate what happens on each draw:

$$\boxed{P(2H) = \underbrace{P(H_2|H_1)}_{12/51}\ \underbrace{P(H_1)}_{1/4} = \frac{1}{17}} \qquad (2)$$

<u>Additional Concepts</u>: Given the definitions of $P(AB)$ and $P(A|B)$ we can introduce the following concepts:

$$\boxed{P(AB) = 0 \Leftrightarrow A \text{ and } B \text{ are \underline{mutually exclusive}}} \qquad (3)$$

When combined with the relation $P(A+B) = P(A) + P(B) - P(AB)$ this implies

$$P(A+B) = P(A) + P(B) - P(AB) \longrightarrow P(A) + P(B) \qquad (4)$$

<u>example</u>: What is the probability of getting $A=1$ <u>or</u> $B=2$ on a single throw of a die? Since $P(A)$ and $P(B)$ are mutually exclusive the result is

$$P(A+B) = P(A) + P(B) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \qquad (5)$$

If $\boxed{P(AB) = P(A)\,P(B) \Longleftrightarrow A \text{ and } B \text{ are \underline{statistically independent}}}$ (6)

It then follows that $P(AB) \underset{\text{in general}}{=} P(B|A)\,P(A) \underset{\text{from (6)}}{=} P(A)P(B) \Rightarrow \boxed{P(B) = P(B|A)}$ (7)

The result in Eq. (7) makes sense since <u>P(B) should be independent of A if B and A are statistically independent.</u>

BAYES' THEOREM This is a fundamental result in probability theory, both practically and philosophically.

We begin with $P(AB) = P(B|A)\,P(A) = P(A|B)\,P(B)$ (1)

Hence
$$P(B|A) = \frac{P(B)}{P(A)}\,P(A|B) \qquad (2)$$

Similarly:
$$P(C|A) = \frac{P(C)}{P(A)}\,P(A|C) \qquad (3)$$

Dividing (2) by (3) $\Rightarrow$
$$\boxed{\frac{P(B|A)}{P(C|A)} = \frac{P(B)}{P(C)}\;\frac{P(A|B)}{P(A|C)}} \qquad (4)$$

BAYES' THEOREM

Example: Let there be 3 drawers A, B, C in each of which there are 2 coins, $g \equiv$ gold $s =$ silver, as shown

gg	gs	ss
A	B	C

We do not know initially which is A, B, C. Let D denote the event of drawing the first coin from one box, with that coin being gold. What is the probability that the other coin is also gold?

Solution: We know immediately that $P(C|D) = 0$, trivially.

Also: $P(A|D) + P(B|D) + P(C|D) = \cancel{0}1$ (5)

Hence to find the individual probabilities all we have to do is to solve for the ratio $P(A|D)/P(B|D)$. From BAYES' THEOREM (4) this ratio is given by

$$\frac{P(A|D)}{P(B|D)} = \frac{P(A)}{P(B)}\cdot\frac{P(D|A)}{P(D|B)} = \frac{(1/3)}{(1/3)}\cdot\frac{(1)}{(1/2)} = 2 \qquad (6)$$

Combining (5) & (6) $\Rightarrow$
$$\boxed{\begin{matrix} P(A|D) + P(B|D) + P(C|D) = 1 \\ (2/3) + (1/3) + 0 \end{matrix}} \qquad (7)$$

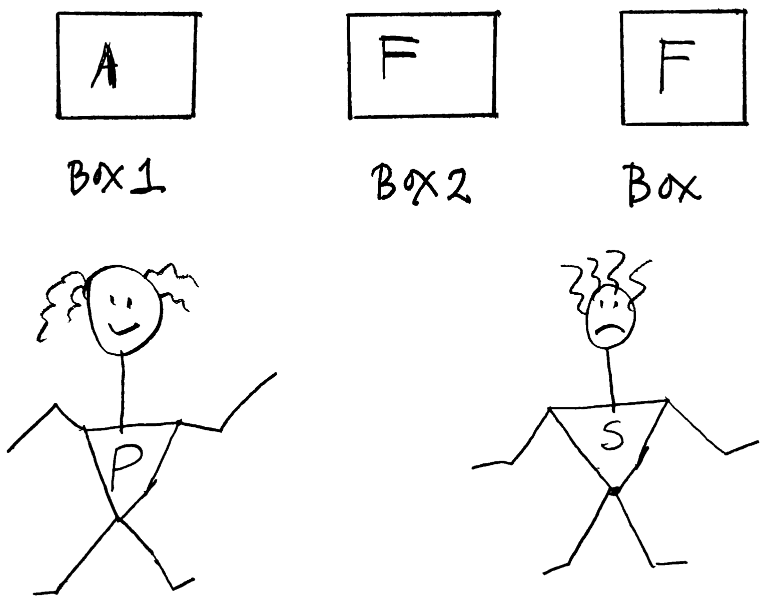
The probability that the 2ND coin is gold is the same as the probability of being in drawer A (given that event D has happened). So finally

$$\boxed{P(\text{2ND coin is gold}) = P(A|D) = 2/3} \quad (8)$$

<u>BAYESIAN STATISTICS</u>: This is a school of statistics that deals with the question of calculating probabilities given some <u>a priori</u> knowledge. In the preceding example, the <u>a priori</u> knowledge is the event D.

THE "MONTY HALL" PROBLEM: TO SWITCH OR NOT TO SWITCH

This can be viewed as an example of the use of Bayesian Statistics. We will present it this way: I (the professor) pick a student and tell him/her that they can have whatever grade they pick from inside a closed box:



The game is played as follows: The student picks a box (which will determine the grade). But before the box is opened the professor opens one of the (other 2) boxes that contains the F, and offers the student the opportunity to switch his/her choice before the box is opened. The question is: Should the student switch his/her choice?

NAIVE ANALYSIS: After the box with the grade of F is opened there remain 2 boxes: one with an **A** and the other with an F. Whichever one of these 2 boxes he chooses, it would appear that either way he/she has a 50% chance of getting the A.

BUT THIS IS WRONG!!

RESOLVING THE MONTY HALL PROBLEM

P = PROF S = STUDENT 3/4/04

STEP 1: P CAN HIDE THE "A" IN

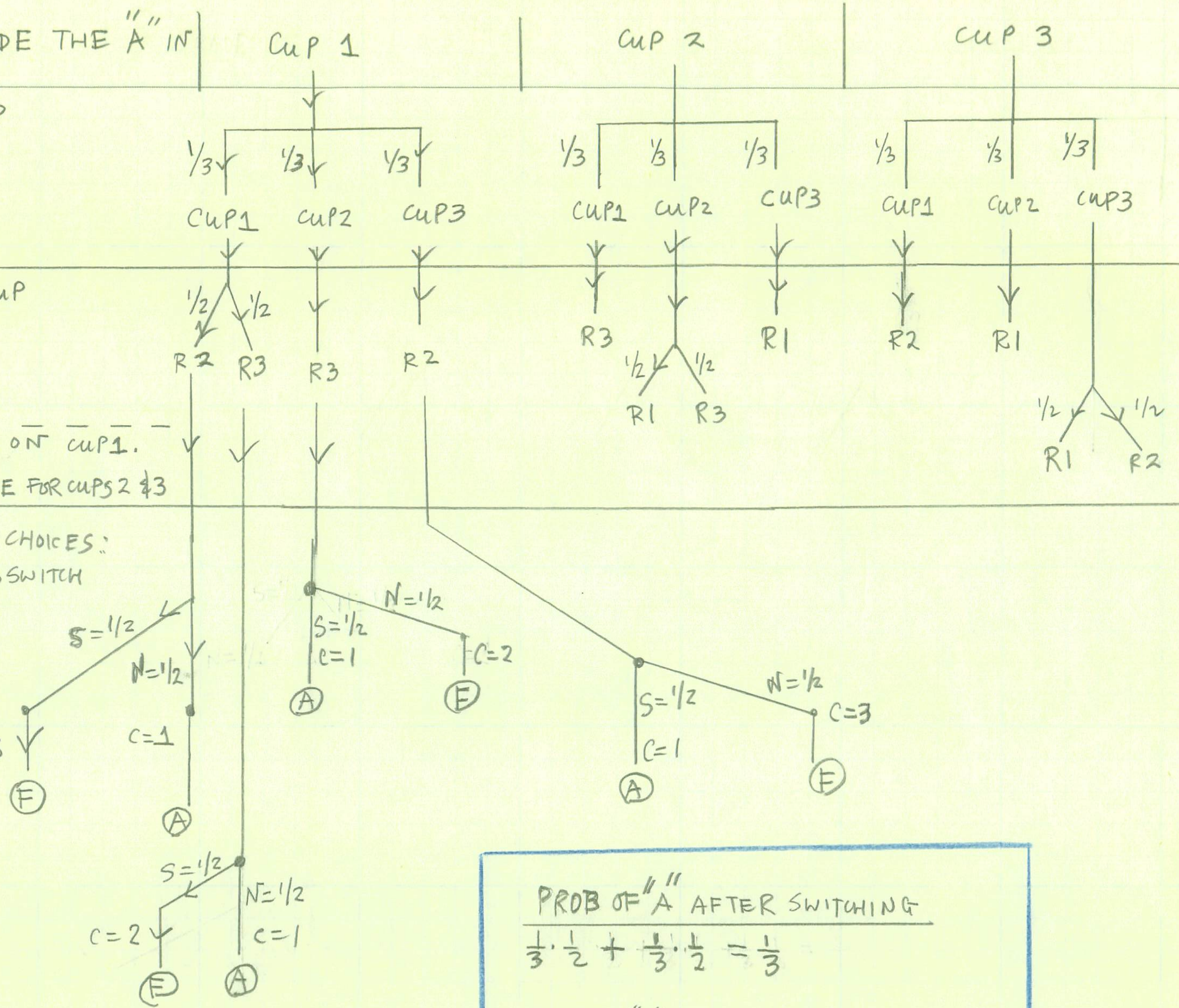
STEP 2: S PICKS A CUP

STEP 3: P REMOVES (R) A CUP WITH AN "F"

R2 ≡ REMOVE CUP2 etc.

FROM HERE ON FOCUS ON CUP1.
THE RESULTS ARE THE SAME FOR CUPS 2 & 3

STEP 4: STUDENT HAS 2 CHOICES:
S = SWITCH; N = NO SWITCH
C = CHOICE



PROB OF "A" AFTER SWITCHING

$\frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot\frac{1}{2} = \frac{1}{3}$

PROB OF "A" WITHOUT SWITCHING

$\frac{1}{3}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{3}\cdot\frac{1}{2}\cdot\frac{1}{2} = 1/6$

PHYSICS 601

SPRING 2004

HOMEWORK # "MONTY HALL"

25 points (optional)

State and resolve the "Monty Hall" problem by a method other than the tree-diagram method used in class. For example, use Bayes' theorem directly.

I will personally grade all the solutions. You can use any reference you choose (web, books, etc.) but you must clearly state your source.

Springer
New York
Berlin
Heidelberg
Hong Kong
London
Milan
Paris
Tokyo

Mark P. Silverman*
mark.silverman@trincoll.edu

A Universe of Atoms, An Atom in the Universe

With 74 Illustrations

*Trinity College
Hartford, Conn.
Ph: (860) 297-2298
fax: (860) 987-6239

2002

Springer

CHAPTER 8

Computers, Coins, and Quanta: Unexpected Outcomes of Random Events

8.1. The Suggestive Power of Fun

Many years ago, I participated in an international conference devoted to improving the teaching of science at all levels of instruction. Although I now recall little of the numerous talks and heated discussions that the conference engendered, there was one event that I have not forgotten. In his introductory remarks, an invited speaker, noted for his compendious study of the life of Isaac Newton, starkly announced that not once, in all the years that Newton engaged in his physical researches, had he (Newton) ever had any "fun." According to the speaker, the pursuit of scientific knowledge for Newton was a solemn and sacred undertaking which the word fun grotesquely trivialized. Moreover, the speaker continued somewhat scornfully, this is precisely how it *should* be; science is too serious a matter to be pursued—*or taught*—with the idea of fun in mind . . . and the sooner teachers grasped this point, the sooner they would be able to teach science more effectively.

I was stunned. I am not a historian, although I have read enough books about Newton to agree that "fun-loving" is not exactly the adjective to apply to a reclusive genius with tendencies toward paranoia. On the other hand, as a scientist—one of very few at the conference in question—I have also read Newton's own writings. It is impossible to read Newton's *Opticks*, for example, and not sense the enormous personal satisfaction and pleasure that its author must have experienced in reflecting upon the deep philosophical problems posed by the behavior of light and in designing and executing simple, yet incisive, experiments to help unravel these mysteries of natural philosophy. Perhaps fun may not be the appropriate word, but any conception of science that ignores the intellectual delight of satisfying one's curiosity, overcoming challenges, and making discoveries has missed a seminal attraction of science both in Newton's time and our own. Indeed, it is precisely this sense of exhilaration and fulfillment in the pursuit of understanding how the world—or a tiny part of it—works that a teacher must communicate to students if they are to appreciate science as something more than a collection of facts and formulas.

Newton was fascinated by the physical behavior of much of what he encountered around him: how objects moved when pushed, how objects fell when released, how objects cooled when heated, how fluids flowed, and what happened to light when it passed through or around various things, to cite but a few of Newton's preoccupations. In the motions and transformations of familiar physical objects, Newton found far-reaching principles waiting to be revealed.

Science has evolved over the past three centuries in ways that Newton could never have imagined, and the objects familiar to many a physicist today now comprise those that can be seen only by powerful microscopes or with satellite-based telescopes or by means of some other kind of expensive apparatus usually requiring the financial support of one government agency or another. In some ways, that is rather unfortunate, although seemingly necessary if the boundaries of scientific knowledge are to expand, for it tends to breed an attitude among at least some scientists and science editors not unlike the attitude of the historian above. The remark of one anonymous wag in the audience of a quantum mechanics conference I spoke at long ago captured this frame of mind precisely. Paraphrasing physicist John A. Wheeler's cryptic assertion that "a phenomenon is not a phenomenon until it is a *measured* phenomenon,"[1] the wag blurted out, "a phenomenon is not a phenomenon until it is a *funded* phenomenon!" Scientists who have ever tried to publish in a premier research journal without having a funding agency to acknowledge as evidence that the submitted work was "serious" science (and *not* fun) will understand the import of the wag's observation.

I have been doing scientific research for over forty years. Much of this research, as recounted in this book and other volumes noted in the Preface, is "serious" science, i.e., part of a carefully planned research agenda. However, a significant fraction of my work was not part of any research plan at all, but undertaken on a whim, for amusement, or out of surprise at some unexpected turn of events. These adventitious projects were often the ones that I enjoyed most and from which I always learned something new and interesting. I cannot believe that a true scientist, including even Newton, does not have fun.

This two-sided nature of scientific motivation—serious and playful—is aptly expressed in Harvey Lemon's vignette of the Nobel Prize-winning American physicist, A. A. Michelson,[2] who, like Newton, was a pioneer in the investigation of light:

> When asked by practical men of affairs for reasons which would justify the investment of large sums of money in researches in pure science, he was quite able to grasp their point of view and cite cogent reasons and examples whereby industry and humanity could be seen to have direct benefits from such work. But his own motive he expressed time and again to his associates in five short words, "It is such good fun."

In this chapter, I discuss a project that started as a computer game, but evolved—unexpectedly—into tests of what is perhaps the most fundamental characteristic of the quantum world: the intrinsically unpredictable occurrence of individual quantum events.

8.2. To Switch or Not to Switch—*That* Is the Question

I never heard of the so-called "Monty Hall" problem until a few years ago when I first saw mention of it in a review[3] of a newly published book of mathematical oddities. Even then, having (by choice) no television in my house, the association of the name with the host of a TV game show ("Let's Make A Deal") meant nothing to me. The problem is easy enough to state, but its solution is counterintuitive in the extreme. Indeed, I have read that, when first brought to the American public's attention by a columnist for a popular magazine,[4] it had driven even professional mathematicians to distraction.[5]

There are three closed boxes. Inside one of them is a valuable cash prize and inside each of the others is a banana. The player picks a box, but before its content is revealed, the game master (who is aware of what is inside each box) opens one that he knows contains a banana. Now, the game master offers the player the following option: The player may keep his or her original choice or (for a small fee in one version of the game) choose the other unopened box. What should the player do?

The nearly universal reply—and indeed the reply given by everyone to whom I personally posed this problem—was that it cannot matter which of the two options is selected. With but two choices remaining, there is a 50% chance of winning in either case. (It would, therefore, be ridiculous to *pay* to switch, respondents said.) This, however, is not the case. Players *double* their chance of winning if they switch. Think about that a while, before continuing.

How can one possibly double his chance of winning by choosing the other of only two boxes? The argument is actually quite simple. Assuming that there is an equal likelihood for any one of the three boxes to contain the prize, a player will have a chance of 1/3 of winning if he selects a box and keeps it. This means that there is a probability of 2/3 of not getting the prize on the first selection. However, if the player switches, then 2/3 becomes the probability of winning, for, under the prevailing circumstances, the unopened box to which the player switches *must* contain the prize if the originally chosen box does not. Thus, the odds of winning are twice as great if the player switches.

The preceding reasoning (as well as other more formal arguments) generally elicited a storm of protest from the ordinarily placid students, colleagues, neighbors, and friends on whom I tried the problem. Probability is a measure of present knowledge they all said; once the game master opens a box, the odds of winning jump from 1/3 to 1/2 whether or not the player switches. The fallacy of thinking this way, however, lies in ignoring the *order* in which events transpire, for this order defines the conditions which determine the probability of winning. The probability P_{switch} of winning by switching is a product of two probabilities: (a) the probability $P(A)$ that the player first picks a box with a banana (event A) and (b) the probability $P(B|A)$ that the player next picks the box with a prize (event B) given that event A has occurred[6]:

$$P_{switch} = P(A)P(B|A). \quad (8.1)$$

Under the rules of the game, the probability of initially selecting a banana is $P(A) = 2/3$ and the conditional probability of selecting the prize *after* the game master has revealed one of the banana-containing boxes is $P(B|A) = 1/1$ (i.e., 100%). Hence, $P_{switch} = 2/3$. If the game master were to have revealed the content of one of the boxes *before* the player made a first choice, then the probability of winning would have been the same whether the player kept that choice or switched. Order matters.

Same for Simpson, etc.

However, suppose—as one dissatisfied colleague argued—that the player simply flipped an unbiased coin to determine the strategy: heads (H) he keeps, tails (T) he switches. Clearly, in this case there must be a 50% chance of winning the prize either way. That observation, in fact, is true, but it does not conflict with the previous conclusion that the player is better off *choosing* to switch. The "coin-toss" strategy, which underlies the intuitive but misguided reasoning of most players, is again compounded of two distinct sets of probabilities. If P_{toss} is the probability of winning when a coin toss determines strategy, P_{keep} and P_{switch} are the original probabilities of winning by keeping or switching one's initial choice, and P_H and P_T are the probabilities (both 50%) of a fair coin landing H or T, then

$$\begin{aligned} P_{toss} &= \frac{1}{2} = P_H P_{keep} + P_T P_{switch} \\ &= \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{2}\right)P_{switch}. \end{aligned} \quad (8.2)$$

From Eq. (8.2), it again follows that P_{switch} must be 2/3 and, therefore, twice P_{keep} if the overall probability of winning by the outcome of a random process is to be 1/2.

To convince both myself and others that, however unexpected, switching really doubles the odds of winning, I asked my son Chris, a high-school junior at the time, to program the game on a computer, using a random-number algorithm to distribute the prize among the boxes. In the first version of our program, created with the HyperTalk language for the Macintosh, a player picks a box, and the computer, again using the random-number generator, opens one of the two remaining boxes. If the opened box contained the prize, then obviously the player lost—but this event was not included in the dataset from which statistics were compiled, for there had been no option of switching. In a second version of the program, the computer played the entire game itself, executing many rounds of prize distribution and box selection with the opening of a prize-containing box automatically excluded.

The results of 20,000 games—10,000 each for the strategies of keeping or switching—are summarized in Figure 8.1. The fraction of times each box was assigned a prize was very close to 1/3, as was also the fraction of times each box was selected by the "player." The strategy of keeping the original choice resulted in winning the prize in 3359/10,000 = 33.59% of the games. By switching, however, the fraction of wins jumped to a smashing 6639/10,000 = 66.39%.

What more can I say?

8.3. On the Run: How Random Is Random?

Actually, there *is* more to say. It was while programming and playing the game that we noticed that the computer seemed to behave rather oddly at times. Although, on average, each box was assigned the prize in one-third of the total trials, in detail the computer occasionally assigned the prize to the *same* box three or four or more times in succession. Was there a defect in the program? Could it be that the internal random-number generator was not generating random numbers? Or were these outcomes to be expected even in the case of a perfectly random selection process? Thus, began my interest in the matter of "runs."

Random events occur without any assignable cause. Emphasis here is on "assignable," for random occurrences do not represent a suspension of the laws of physics; rather, in the absence of sufficient knowledge of initial conditions, one cannot predict their outcome individually. Consider one of the classic examples of a random process: coin tossing. Certainly, the coin is subject to Newton's laws; however,