UNDERESTIMATING THE PROBABILITY OF COINCIDENCE

MA Muller
MSc PhD
University of Stellenbosch

SUMMARY

Coincidences are more common than most people might expect. It is quite possible that different pieces of evidence that seem to point in the same direction do so coincidentally. We come to the best possible conclusion about (say) the probability of guilt only after careful analysis of the combination of probabilities of the respective pieces of evidence has been performed in conformance with the principles of probability theory. Several methods are available for the evaluation and handling of such contingencies. Depending on the way a particular situation presents itself, Bayes's theorem in one of its equivalent guises is often used. The danger in avoiding this type of reasoning is that incorrect conclusions may be drawn, believing that events are somehow beyond coincidence. When it happens in a court of law it may be extremely prejudicial to the defendant. Coincidences are best understood within the context of probability theory.

1   Introduction

The Dutch writer Maarten 't Hart gave us a glimpse of life in a bygone age. He wrote the following about an ancestor of Ludwig van Beethoven:

"Op 5 augustus 1595 werd te Kampenhout de 50-jarige Josyne van Beethoven gearresteerd. Dorpelingen beschuldigden haar ervan dat zij met de duivel een verbond had gesloten. Vier keer was, juist op het moment dat zij langsliep, een paard plotseling gestorven. Ze werd overgebracht naar de gevangenis van Brussel. Daar liepen nieuwe beschuldigingen binnen. Toen ze langs een stoeterij stapte, vertelde een nieuwe getuige, had een paard spontaan bloed gepiest en was aan een koliek gestorven. Een boer wist te vertellen dat de melk van één zijner koeien reeds zuur uit de uier was gekomen toen Josyne voorbijkwam.

Josyne van Beethoven had net zo'n karakter als haar beroemde nazaat. Onverschrokken ontkende ze alle beschuldigingen. Je had toen nog geen Pieter Baan centrum, maar wel de Pijn Bank. Daar werd ze net zo lang gemarteld tot ze bekende dat ze een heks was. Eind september werd zij op de Grote Markt van Brussel levend verbrand. Aan de vooravond van haar terechtstelling slikte ze vergeefs potscherven in de hoop daaraan te overlijden." [1]

---

[1] M 't Hart "Een hedendaags heksenproces" *NRC Handelsblad* (07-10-2006), *www. vorige.nrc.nl/ nieuwsthema/luciadeb/article1730533.ece/Een_hedendaags_ heksenproces* (accessed 17-01-2014). See appendix (5) for a translation of this passage.

These events occurred at a time when people were more frequently exposed to farm animals than we are today. Animal-drawn vehicles was the only mode of transport. Sooner or later an animal would die, possibly in the vicinity of passers-by who merely happened to be there. If some person was present at more than one such incident the populace would not easily have accepted it as coincidence. Embedded in the medieval world-view, concurrence of such events would have been interpreted as a manifestation of worst evil in this person.

Strangely enough, people still seem to have difficulty trying to make sense of coincidence. Ray Hill writes

"How good is our intuition regarding coincidences? The phrase 'lightning doesn't strike twice' does seem to reflect our perceptions in that almost everyone underestimates the true chances of coincidence."[2]

In a previous paper it was shown that matching DNA profiles do not necessarily imply that a perpetrator is identified.[3] It may well happen that two DNA profiles coincidentally match for no particular reason at all. In general, different pieces of evidence pointing in the same direction may be coincidental.[4] In part 2 it is shown that in certain situations where the appearance of coincidences might least be expected, such coincidences are in fact highly probable.

We come to the best possible conclusion about (say) the probability of guilt only after careful analysis of the combination of probabilities of the respective pieces of evidence has been performed in conformance with the principles of probability theory. The danger in avoiding this type of reasoning is that incorrect conclusions may be drawn, believing that events are somehow beyond coincidence.[5] Coincidences are best understood within the context of probability theory.

Bayes's theorem tells us how likely it is that a particular preconception is correct in light of the evidence we observe. It is an important tool in handling uncertainty by means of probability theory. Fenton and Neil assert

"Bayes's theorem is a basic rule, akin to any other proven maths theorem, for updating the probability of a hypothesis, given evidence. Probabilities are either combined by this rule, or they are combined wrongly."[6]

The following analogy illustrates why using Bayes's theorem is the best way of handling uncertainties:

"Saying that the expert should not use this 'mathematical formula' … is like saying that if one is just estimating by eye the area of a field, one is not allowed to multiply estimates of its width and length together. Clearly it is the correct procedure: there is no uncertainty in the relation between length, width and area, only in their values. … The

---

[2] R Hill "Multiple Sudden Infant Deaths – Coincidence or Beyond Coincidence?" (2004) 18 *Paediatric and Perinatal Epidemiology* 320 322.

[3] MA Muller "Handling Uncertainty in a Court of Law" (2012) 23 3 *Stellenbosch Law Rev* 599 599-609.

[4] See Examples 1, 2 and 6.

[5] R Hill "Reflections on the cot death cases" (2005) 2 *Significance* 13 14

[6] N Fenton & M Neil "On limiting the use of Bayes in presenting forensic evidence" (Draft) (2012), *www.eecs.qmul.ac.uk/~norman/papers/likelihood_ratio.pdf* (accessed 17-01-2014)

fact that variables cannot be precisely expressed does not affect the validity of the relationships described by the formula." [7]

There are equivalent versions of Bayes's theorem. Depending on the way a particular situation presents itself, the most convenient version of Bayes's theorem is used. In this paper Bayes's theorem is formulated in terms of the *probability* of an event.[8] It serves as an alternative to the equivalent version of Bayes's theorem which is formulated in terms of the *odds* of an event. Notation and terminology are explained in the appendix to this paper, as well as in a previous paper.[9] Since randomness is an important aspect of many considerations involving uncertainty we recall some properties of this often counter-intuitive notion.

2 Randomness

The word "random" is sometimes used in discussions about legal matters. It may be that randomness is an essential precondition to a particular proposition under consideration. Properties of randomness are best illustrated by considering the well-known random process of coin flipping.

When an unbiased coin is repeatedly flipped and the outcome of each trial [10] is written down, one after the other, a sequence is obtained consisting of *H*'s ("heads") and *T*'s ("tails") in random order. At any stage in the process of coin flipping it was unknown and unknowable what the outcome of subsequent trials would be.[11]

2 1   Deviations

We typically expect the proportion of *H*'s in any finite segment of our sequence to be around 50%, and should the proportion of *H*'s in a certain segment deviate significantly from 50%, we anticipate that in a forthcoming segment a correction will occur in the opposite direction. Many people seem to believe that deviations will be neutralized, and that things will "even out".[12] This is not so.

The process of coin flipping has neither memory nor feelings of remorse about results of past trials. At any given point in the abovementioned sequence, outcomes of subsequent flips of the coin proceed as if it had never been flipped before. Previous deviations do not somehow become self-corrected. For example, a losing streak in gambling does not imply that good luck is "due", or that losses will be compensated. In this sense randomness is not as fair as the gambler

---

[7] B Robertson, GA Vignaux & CEH Berger "Extending the Confusion about Bayes" (2011) 74 *Mod L Rev* 444 452, 455.

[8] See appendix (1).

[9] Muller (2012) *Stellenbosch Law Rev* 608-609. Basic aspects of conditional probabilities and background material concerning Bayes's theorem were presented in the cited paper. These particulars are not repeated in the current paper.

[10] The word *trial* is used as in probability theory. In this case it refers to a single flip of the coin.

[11] DJ Bennett, *Randomness* (1998) 165.

[12] A Tversky & D Kahneman "Evidential Impact of Base Rate" in D Kahneman, P Slovic & A Tversky (eds) *Judgment under Uncertainty: Heuristics and Biases* (1982) 153-160.

expects it to be. Randomness seems to behave as if it has an agenda of its own. Deviations merely become diluted as the number of trials increase.[13]

On the other hand, people may be tempted to act on the consequences of randomness as though observed outcomes herald some emerging tendency. This conviction is also invalid.

The following example illustrates an effect of the belief that deviations will eventually tend to cancel each other. The average mark of all students who sat for a test is 55%. Suppose a group of 30 of these students is randomly selected and we are told that some student in this selected group attained 85% in his test. Given this information, the expected average mark of this group of 30 students is 56%, although many people would anticipate 55%.[14]

2 2   Sometimes most probable events are improbable

Consider the flipping of an unbiased coin a large number of times. Since the appearance of *H* and *T* are equally probable, we expect to observe "more or less" the same number of appearances of *H* and *T*. For instance, if the coin is flipped 2000 times we expect that *H* will appear more or less 1000 times, and that *T* will appear more or less 1000 times. However, the probability of the event that *exactly* 1000 appearances of each of *H* and *T* occur, is a mere 1.8%.[15] In this example the most probable outcome is itself very improbable.

2 3   Clusters

In any soccer match the probability that at least two people on the field share the same birthday is above 50%.[16] If an unbiased coin is flipped 100 times, then the probability that somewhere in the sequence of outcomes the same side of the coin will be observed in a run of five or more consecutive trials is a remarkable 97.1%.[17]

Randomness does not imply that outcomes will be "more or less alternating" or "evenly distributed". Fraudsters might tend to avoid long runs of similar (or identical) outcomes because more often than not they would be under the mistaken impression that such a situation is improbable and that their fraud might therefore be detected. It is a remarkable consequence of randomness that the appearance of sizeable clusters of similar outcomes is highly probable.[18] The following diagrams illustrate this aspect of randomness in the context of points distributed over a flat surface:

---

[13] Tversky & Kahneman *Judgment under Uncertainty: Heuristics and Biases* 153-160.

[14] A Tversky & D Kahneman "Belief in the Law of Small Numbers" (1971) 76(2) *Psychological Bulletin* 105 105-106.

[15] By the binomial distribution, $\binom{2000}{1000}(0.5)^{1000}(1-0.5)^{1000} = 0.018$. See appendix (4).

[16] By "birthday" we mean the anniversary of the day on which someone was born; H Tijms *Understanding Probability* (2004) 78-79.

[17] GC Berresford "Runs in coin tossing" (2002) 33 *The College Math Journal* 391 391-393.

[18] A website which demonstrates this by means of repeated coin flipping is *www.shodor.org/interactivate-java/activities/Coin/* (accessed 17-01-2014).
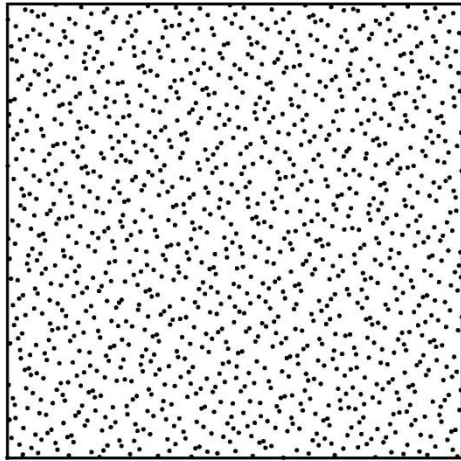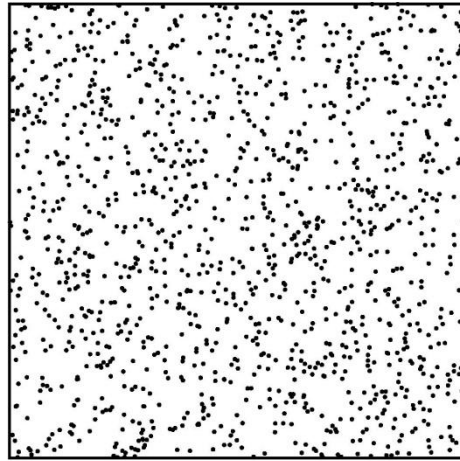
| Fig. 1 | Fig. 2 |

The above squares each contain the same number of dots. The dots in fig. 2 are randomly distributed, but not so in fig. 1. In the case of randomly distributed dots the occurrence of bundles and streaks of dots, as well as thinly populated regions, are very probable.

In 2005 Apple introduced an iPod with the attribute that it could be set to play music tracks in random order. Soon people complained that this particular feature of the new iPod did not seem to work — under this option several tracks by the same artist would sometimes come up one immediately after the other. This is indeed a likely consequence of randomness. Eventually Apple's CEO, Steve Jobs surrendered to consumer demands and reprogrammed the iPod. He would give them what they wanted. "We're making it less random to make it feel more random," was his famous comment.[19]

3   Attempts at identifying the most likely person

Example 1:   A bank-robbery is attempted during a busy hour of the day. An employee raises the alarm and all exits from the bank shut down automatically. Two masked robbers find themselves trapped in the bank. Unseen they dispose of their masks and join the group of trapped customers. The police now have a total of 46 potential suspects, all claiming to be innocent. One person of the group is randomly selected for a lie detector test. The outcome of this person's test is positive. It is known that if someone is guilty, then the lie detector test indicates guilt with probability 90%. It is also known that if someone is innocent, the test indicates guilt with probability 5%. What is the probability that this person is indeed one of the robbers?

There are two possibilities.   On the one hand we have the possibility that the selected person is a robber (which is unlikely, because only 2/46 (= 4.3%) of the trapped people are robbers), combined with the event that the outcome of his test is positive (which is likely, because in cases where the subject is guilty the test shows guilt with probability 90%). On the other hand there is the possibility that the selected person is innocent (which is likely, because 44/46 (= 95.7%) of the trapped people are innocent), combined with the event that his lie detector test is

[19] J Maslin "His Heart Belongs to (Adorable) iPod" *The New York Times* (19-10-2006), *www.nytimes.com/2006/10/19/books/19masl.html?_r=0* (accessed 17-01-2014).

positive (which is unlikely, because in cases where the subject is innocent the test shows guilt with probability of only 5%).

We resolve the above dichotomy by means of Bayes's theorem which is formulated in terms of probabilities of events.[20] In this manner we avoid the prosecutor's fallacy and arrive at the best possible conclusion.

Let A and B be the following events:

A : The result of this person's lie detector test is positive.
B : This person is one of the robbers.

From the above and the information in appendix (2) we obtain the following probabilities:

$$P(B) = 2/46, \quad P(\sim B) = 44/46, \quad P(A|B) = 90/100 \quad \text{and} \quad P(A|\sim B) = 5/100.$$

Given that the result of this person's lie detector test is positive, the above figures are substituted into Bayes's equation. Then the conditional probability that he is indeed a robber is

$$P(B|A) \;=\; \frac{P(B)\,P(A|B)}{P(B)\,P(A|B) \,+\, P(\sim B)\,P(A|\sim B)}$$

$$=\; \frac{(2/46)\,(90/100)}{(2/46)\,(90/100) \,+\, (44/46)\,(5/100)}$$

$$=\; 0.45 \quad (= \; 45\%).$$

Clearly this person cannot be found guilty on charges of attempted bank-robbery on grounds of the above evidence. Notwithstanding the event that he was trapped in the bank, and the event that his lie detector test was positive, the available evidence suggests that these events were probably coincidental. In some countries the results of lie detector tests are not admissible as evidence in court. Example 1 serves as motivation for that point of view.

Example 2: For unknown reasons Sally Clark lost two babies within a period of fourteen months.[21] Initially it was thought that these were "cot deaths", also known as "sudden infant death syndrome" or "SIDS". At her murder trial and subsequent first appeal she was found guilty of murdering her children. The manner in which statistics were presented in court brought disquiet to the ranks of the Royal Statistical Society and it lead to a public statement by the RSS expressing concern about the handling of statistics. After her second appeal Sally Clark was acquitted of all charges.

---

[20] See appendix (2).
[21] *R v Sally Clark* [2000] EWCA Crim 54, [2003] Crim 1020; Muller (2012) *Stellenbosch Law Rev* 605-606.

Helen Joyce and Barry Lewis used Bayes's theorem in estimating the probability of Sally Clark's innocence, given the evidence.[22] Assume that there are the two possibilities: either the babies died of natural causes, or they were murdered.

Let *A*, *B* and *C* be the following events:

A : Two children of the same family died.
B : Both children died of natural causes.
C : Both children were murdered.

We calculate the conditional probability $P(B|A)$, the probability that both children died of natural causes (*ie* that Sally did not murder her children), given the event that both children died. By Bayes's theorem,

$$P(B|A) \;=\; \frac{P(B)\,P(A|B)}{P(B)\,P(A|B) \;+\; P(C)\,P(A|C)} \;.$$

We need only determine the numerical value of the different entries on the right hand side of the equation. Based on research into the prevalence of infant deaths Joyce and Lewis estimated that *P(B)*, the absolute probability that two children from the same family in England and Wales might die of natural causes,[23] would be about one in 130,000. Furthermore, $P(A|B)$ is the conditional probability that two children died, given the fact that they died of natural causes, therefore $P(A|B) = 1$, a tautology. For the same reason it follows that $P(A|C) = 1$.

The absolute probability *P(C)*, *ie* the probability that two children from a randomly chosen family are murdered by their mother posed some problems. But Joyce came forward with a reliable figure for *P(C)*. She refers to information from the British Home Office that each year less than 30 children are murdered by their mothers. Since some 650,000 children are born in England and Wales each year, and since double murders in the same family are more scarce than single murders, she arrived at a figure less than 30 in 650,000. If we divide this figure by a factor 10 and thus obtain 3 in 650,000, the probability of double murders in the same family might still be moderately overestimated.

Inserting these figures into Bayes's equation we obtain the following estimate of the probability that Sally Clark did not murder her children:

$$P(B|A) \;=\; \frac{(1/130{,}000)(1)}{(1/130{,}000)(1) \;+\; (3/650{,}000)(1)}$$

$$= 0.625 \quad (= 62.5\%)$$

ie the probability of her innocence is five in eight. Every year some 650,000 babies are born in England and Wales and we may expect that 650,000/130,000 = 5 families would suffer such a double tragedy. A double cot death would therefore occur in England and Wales on average as often as once every 10 weeks. In

[22] H Joyce "Beyond Reasonable Doubt" *Plus Magazine* (1-9-2002) *www.plus.maths.org/ issue21/ features/clark* (accessed 17-01-2014); B Lewis "Taking Perspective" (2003) 87 *Math Gazette* 418 422-425.
[23] Joyce *Plus Magazine* (1-9-2002).

2001 *The Daily Telegraph* carried a report that Sally received letters of support from 11 families who each lost two babies as a consequence of SIDS.[24] It is clear from the above that the probability of the coincidence of two babies from the same family dying of SIDS is higher than generally anticipated.

Methods of probability theory are not only applicable in matters relating to criminal cases. Perhaps criminal cases only tend to be more newsworthy. The following example illustrates a particular type of situation that was first presented by Martin Gardner. [25]

Example 3: A businessman buys a house in a town. He has a specific legal problem concerning his business and wishes to obtain sound legal assistance from an expert in his new neighbourhood. He seems to remember that he once heard that the best lawyer in the country for this particular kind of problem has his practice in this town, but he does not know who this person is. There are only three lawyers (Anderson, Brown and Clark) in separate law firms in the town, and the expert could be anyone of them. Not knowing what to do, he randomly selects any one of them, say Anderson. He decides to make an appointment with him after his return from holiday. The businessman's new neighbour is a recently retired judge who invites him over to his house one evening. Eventually the businessman mentions the problem he has finding the expert lawyer. The neighbour knows exactly who he is looking for, but for ethical reasons feels that he should not disclose the expert's name to the businessman. But he would like to help his neighbour as much as he can. He tells him that he will neither comment on his choice of Anderson, nor name the expert. But he is prepared to inform him that Clark is definitely not the person he is seeking. With this information, should the businessman persist with his initial choice of Anderson, or should he switch to Brown? We show that choosing Brown at this point would be to the businessman's advantage. Remarkably, the probability that Brown is the expert is in fact double the probability that Anderson is the expert.

Let *A, B* and *C* be the following events:

A : Anderson is the expert.
B : Brown is the expert.
C : Clark is the expert.

Initially any one of Anderson, Brown or Clark could with equal probability have been the expert, so $P(A) = 1/3$, $P(B) = 1/3$ and $P(C) = 1/3$. Let X and Y be the following events:

X : The neighbour names Brown as a person who is not the expert.
Y : The neighbour names Clark as a person who is not the expert.

Suppose Anderson is indeed the expert. Since the neighbour does not comment on Anderson, he could with equal probability have named anyone of

---

[24] J Rozenberg "Solicitor tackles system from the inside" *The Daily Telegraph* (01-05-2001), *www.telegraph.co.uk/news/uknews/1328668/Solicitor-tackles-system-from-the-inside.html* (accessed 17-01-2014)

[25] M Gardner "Problems involving questions of probability and ambiguity" (Apr 1959) 201 *Scientific American* 174 180-182.

Brown or Clark as a person who is not the expert. The neighbour made a random choice between Brown and Clark, and then informed the businessman that Clark is not the expert. Hence $P(X|A) = 1/2$ and $P(Y|A) = 1/2$.

Suppose Brown is indeed the expert. Since the neighbour does not comment on Anderson, he could only have named Clark as a person who is not the expert. Therefore $P(Y|B) = 1$.

Substituting the above values into Bayes's equation, the probability that Anderson is the expert, given the fact that the neighbour named Clark as a person who is not the expert, is

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

$$= \frac{(1/3)(1/2)}{(1/3)(1/2) + (1/3)(1)}$$

$$= 1/3 \quad (= 33.3\%).$$

Similarly, the probability that Brown is the expert, given the fact that the neighbour named Clark as a person who is not the expert, is

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

$$= \frac{(1/3)(1)}{(1/3)(1/2) + (1/3)(1)}$$

$$= 2/3 \quad (= 66.6\%).$$

So by switching his choice to Brown the probability that he identifies the expert is then double the probability of success associated with his initial choice when he randomly chose Anderson.

The argument is independent of the businessman's original choice. No matter which lawyer the businessman initially chose, his neighbour would have (correctly) named one of the other two lawyers as a person who is not the expert in the manner described above. It is greatly to the businessman's advantage to change his original choice by subsequently choosing the particular lawyer which was neither initially selected, nor was mentioned by the neighbour.

There are variants of Example 3. Suppose that in a situation similar to that of Example 3 there are *four* lawyers in town, and one of them is the unknown expert. The businessman randomly chooses any one of them. The learned neighbour does not comment on the businessman's choice but helps him by (correctly) naming *two* non-experts amongst the remaining lawyers. If the businessman then switches his choice to the last remaining lawyer it is again to his advantage. By arguments akin to those presented in Example 3, the probability of identifying the expert in this way is treble the probability of success associated with the businessman's initial choice.[26]

---

[26] R Flohr "Bayesiaanse evenredigheid" (2012) 87(5) *Euclides* 201 201-204.

At first, the conclusion of the above may seem strange by our habitual way of thinking, but truth does bring about some intriguing consequences.

4 The binomial distribution

There are particular ways by means of which probabilities of certain events may be established. The binomial distribution[27] is a tool that is used in a variety of situations.

Example 4: A multiple choice test consists of seven questions of equal weight. Each question has five options from which a candidate should select the correct answer. A candidate passes the test if he attains at least 50% on aggregate. A certain candidate did not study the subject material and is utterly clueless. Nonetheless, he completes the questionnaire by randomly marking options. We calculate the probability that he will pass the test.

In order to pass the test, the candidate should make no more than three mistakes. By the binomial distribution, the probability that all seven answers are correct, is $\binom{7}{7}(0.2)^7(0.8)^0$. Similarly, the probablity that six questions are correctly answered is $\binom{7}{6}(0.2)^6(0.8)^1$. The probablity of five correct answers is $\binom{7}{5}(0.2)^5(0.8)^2$, and the probablity of four correct answers is $\binom{7}{4}(0.2)^4(0.8)^3$. Therefore the probability that he passes the test is[28]

$$\binom{7}{7}(0.2)^7(0.8)^0 + \binom{7}{6}(0.2)^6(0.8)^1 + \binom{7}{5}(0.2)^5(0.8)^2 + \binom{7}{4}(0.2)^4(0.8)^3$$

$$= \ 0.00001 + 0.0003 + 0.0043 \ + 0.0286$$

$$= \ 0.033 \quad (= \ 3.3\%).$$

In Example 4 the probability is very low that the candidate would coincidentally select four or more correct options in completing his multiple choice test and thus pass the test. The probability of coincidentally selecting two or three correct answers by chance is much higher, but in the context of Example 4 such coincidences would be of no consequence since the candidate would then fail his test. This is an example where most people readily accept the truth of the above conclusions. People are generally familiar with tests and exams.

Example 5: An accused is found guilty by a panel of five judges if a majority of judges vote "guilty". Each judge comes to his own decision independently of his colleagues. For each judge the probability that he (or she) comes to the correct conclusion is 80%. Suppose the defendant is indeed guilty.
(a) What is the probability that the panel will come to the conclusion that the defendant is guilty?

---

[27] See appendix (4).

[28] By appendix (3) the numerical values of the respective binomial coefficients are $\binom{7}{7} = 1$, $\binom{7}{6} = 7$, $\binom{7}{5} = 21$ and $\binom{7}{4} = 35$.

(b) If exactly three of the judges come to the same conclusion, what is the probability that the panel comes to the conclusion that the defendant is guilty? [29]
     Let A and B be the following events:

A: The panel finds the defendant guilty.
B: Three judges come to the same conclusion.

(a) By the binomial distribution, the probability that only three judges vote "guilty" is $\binom{5}{3}(0.8)^3(0.2)^2$. The probability that four judges vote "guilty" is $\binom{5}{4}(0.8)^4(0.2)^1$, and the probability of five judges voting "guilty" is $\binom{5}{5}(0.8)^5(0.2)^0$. So the probability that the panel arrives at the correct conclusion that the defendant is guilty, is [30]

$$P(A) \ = \ \binom{5}{3}(0.8)^3(0.2)^2 \ + \ \binom{5}{4}(0.8)^4(0.2)^1 \ + \ \binom{5}{5}(0.8)^5(0.2)^0$$

$$= \ 0.2048 \ + \ 0.4096 \ + \ 0.32768$$

$$= \ 0.942 \ \ (= \ 94.2\%).$$

(b) Using the expression in appendix (2) we calculate $P(A|B)$, the conditional probability that the panel comes to the conclusion that the defendant is guilty, given that three judges vote the same:

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

$$= \ \frac{P(A \text{ and } B)}{P(B \text{ and } A) + P(B \text{ and } (\sim A))}$$

$$= \ \frac{\binom{5}{3}(0.8)^3(0.2)^2}{\binom{5}{3}(0.8)^3(0.2)^2 \ + \ \binom{5}{2}(0.8)^2(0.2)^3}$$

$$= \ 0.8 \ \ (= \ 80\%).$$

     If the word "guilty" in Example 5 were everywhere replaced by the word "innocent" then the corresponding conclusions about innocence would follow in exactly the same way.

     Example 6: A large retail company has 500 stores, and 60 female workers are employed at each store. The prevalence of breast cancer among female store workers is known to be 2%. We determine the number of stores in which five or more cases of breast cancer would occur.
     By the binomial distribution, the probability that there is no case of breast cancer in a randomly chosen store is $\binom{60}{0}(0.02)^0(0.98)^{60}$. Similarly, for any

---

[29] S Ross *A First Course in Probability* (2010) 184.
[30] As before, $\binom{5}{2} = 10$, $\binom{5}{3} = 10$, $\binom{5}{4} = 5$ and $\binom{5}{5} = 1$.

randomly chosen store, the probability of only one case of breast cancer is $\binom{60}{1}(0.02)^1(0.98)^{59}$; the probability of two cases is $\binom{60}{2}(0.02)^2(0.98)^{58}$; the probability of three cases is $\binom{60}{3}(0.02)^3(0.98)^{57}$; and the probability of four cases is $\binom{60}{4}(0.02)^4(0.98)^{56}$.

So the probability that there are five or more cases of breast cancer in a randomly chosen store is [31]

$$1 - \binom{60}{0}(0.02)^0(0.98)^{60} - \binom{60}{1}(0.02)^1(0.98)^{59} - \binom{60}{2}(0.02)^2(0.98)^{58}$$
$$- \binom{60}{3}(0.02)^3(0.98)^{57} - \binom{60}{4}(0.02)^4(0.98)^{56}$$

$$= 1 - 0.2976 - 0.3644 - 0.2194 - 0.0865 - 0.0252$$

$$= 0.0069.$$

Since $500 \times 0.2976 = 148.8$, there are some 148 stores in which there are no cases of breast cancer. Also, since $500 \times 0.0069 = 3.4$, we discover that there are three stores in which five or more female employees develop breast cancer. People may feel the need to assign blame or accountability. Yet there is no particular reason for the respective clusters of breast cancer cases. There are no toxic substances around, no construction errors were made when the stores were built, no negligence nor malice. In fact, the occurrence of these clusters of breast cancer in some stores is mere coincidence, similar to the outcomes of repeated coin flipping (see part 2 3).

5. Conclusion

In some situations most people might not regard the appearance of coincidences to be particularly strange. Example 4 could be an illustration of this. However, humans are generally known having difficulty getting their minds around aspects of uncertainty. They make errors of logic and their approach is often wrong. For instance, using the same tools as in Example 4, Example 6 shows that coincidences of substantial magnitude are implicit in larger populations — for no apparent reason at all. People do not usually expect this to be true. Unless we do the proper calculation we will not be certain of the truth of what our intuition tells us.

In situations similar to those in Examples 1 and 2 we rely on the application of another tool at our disposal. In these cases Bayes's theorem enables us to calculate the probability of guilt (or innocence) of a suspect, given the evidence. This gives an indication of the probability that the relevant events are coincidental. Since Bayes's theorem is such a powerful tool in this work, we include Examples 3 and 5, the former of which has a remarkable (yet true) conclusion.

Appendix: notation and terminology

---

[31] Again, $\binom{60}{0} = 1$, $\binom{60}{1} = 60$, $\binom{60}{2} = 1770$, $\binom{60}{3} = 34{,}220$ and $\binom{60}{4} = 487{,}635$.

(1) As was mentioned in the introduction of this paper we use the logic of probability theory, and therefore our terminology is that of probability theory.[32] For example, in our context the nouns *probability*, *odds*, *likelihood* and *possibility* are not synonyms.

The *probability P(A)* of an event *A* is a measure of the strength of one's conviction of the truth that the event *A* occurs.[33] It is always a number between 0 and 1, or equivalently, a percentage between 0% and 100%. For example, if *A* is the event that heads appear when an unbiased coin is flipped, then we write *P(A)* = 50%, or we write *P(A)* = 1/2. If for some event *B* we have *P(B)* = 0% we say the event *B* is *impossible*, and if *P(B)* = 100% we say the event *B* is *certain*.

For any event *A*, the *complementary event* ~*A* of *A* is the event when *A* does *not* occur. For example, if *A* is the event that heads appear when an unbiased coin is flipped, then ~*A* is the event that tails appear when an unbiased coin is flipped. For any event *A* we always have *P(A)* + *P(~A)* = 1 (also written *P(A)* + *P(~A)* = 100% in terms of percentages).

(2) If *A* and *B* are any events then "*A and B*" denotes the event when both events *A* and *B* occur. Suppose *P(A)* > 0. Then the *conditional probability* of *B*, given *A* (*ie* the probability of the event *B*, given that event *A* occurs) is denoted by *P(B|A)*. The second entry in the bracket is the condition under which the probability of the first entry is given.

Equivalent versions of *Bayes's theorem* which are formulated in terms of probabilities[34] state

$$P(B|A) \; = \; \frac{P(B \text{ and } A)}{P(A)} \; = \; \frac{P(B \text{ and } A)}{P(A \text{ and } B) + P(A \text{ and } (\sim B))} \; ,$$

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

$$P(B|A) \; = \; \frac{P(B)\,P(A|B)}{P(B)\,P(A|B) + P(\sim B)\,P(A|\sim B)} \; .$$

The latter two expressions establish the relationship between the conditional probabilities *P(B|A)* and *P(A|B)*. In general *P(B|A)* and *P(A|B)* may have numerical values which are widely different.[35]

(3) For any positive integer *n* the number *n factorial* (written *n*!) is defined to be the product of the first *n* positive integers, *ie*

$$n! \; = \; 1 \times 2 \times 3 \times 4 \times \dots \times n.$$

We also define 0! = 1. For example, there are 5! = 1 × 2 × 3 × 4 × 5 = 120 ways of arranging five different books in sequence on a bookshelf.

---

[32] Ross *A First Course in Probability* 1-116
[33] 22-57
[34] 65-71
[35] Muller (2012) *Stellenbosch Law Rev* 600.

If k and n are integers with $0 \leq k \leq n$, then the *binomial coefficient* $\binom{n}{k}$ is defined to be the following number:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \;.$$

$\binom{n}{k}$ represents the number of ways a set of k objects can be chosen from a set of n objects.

(4) If an experiment is repeated n times, and the probability of each successful outcome is known to be p (where $0 \leq p \leq 1$), then the probability of exactly k successes is

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

This expression is called the *binomial distribution.*

(5) Translation of the Dutch quotation in the introduction:

> On August 5, 1595 Josyne van Beethoven (age 50) was arrested in the town of Kampenhout. Townsfolk accused her of having entered into a pact with the devil. It was alleged that on four occations a horse suddenly died the moment she walked by. She was brought to prison in Brussels, where further accusations were levelled against her. A new witness claimed that a horse urinated blood and died of colic as she walked past a farm. Another farmer said that milk from one of his cows came sour from the udder just the moment Josyne walked by.
>
> But Josyne van Beethoven had a character similar to that of her illustrious descendant. Boldly she denied all allegations. In those days they did not have a Pieter Baan centre, but the torturer's rack was at hand. She was tortured until she confessed that she was a witch. In September she was burnt at the stake in the Grote Markt in Brussels. In vain she attempted suicide by swallowing potsherds on the eve of her execution.