Parabola Volume 52, Issue 1 (2016)

If you don't find anything, is there really nothing there?

Dr Bernard Kachoyan[1]

Let us suppose that you are looking for some objects in a particular area. Now also suppose that you find nothing after searching the area. Is there really nothing there? Or, more generally, if you find N objects, then how many objects are there left to be found? Equivalently, how many were actually there to be found in the first place? The objects might be physical objects such as people in a search and rescue operation, or animals, schools of fish, ore pockets or oil deposits. Or they might be "virtual" things like bugs in computer code, faults in manufactured products, etc.

If the sensors you use, whether some electronic means or even your own eyes, are perfectly capable and will find an object with 100% probability, then the answer is clearly trivial. You have found all that were there to be found. But what happens if your sensors are less than perfect, either inherently so, or perhaps because the objects may be concealed or the environment difficult to search?

Moreover, if you know before the search (mathematicians call this *a priori*) that there are M objects there and you find N, then you also know that there are $M - N$ left after the search (called *a posteriori*). But what if the number of present objects is uncertain and can be expressed as a probability distribution on the number of objects thought to be present. Intuitively, if we find N of these, then there should be a resultant (*a posteriori*) distribution of the number of objects left.

We will see how these two distributions can be connected, together with the probability of finding each object, through what is known as Bayes' Theorem.

Thomas Bayes was an eighteenth century English statistician, as well as being a Presbyterian minister. While his contemporaries were considering the more traditional probabilities, for example the probability of pulling out a black ball from an urn containing a certain number of white and black balls, he was considering what is sometimes called the inverse problem: given the colour of the drawn ball or balls, what can we infer about the (unknown) number of white and black balls in the urn?

Before I present Bayes' Theorem, we need a reminder about the topic of conditional probability. This is defined as the probability of an event B happening given that another event A has also happened. This probability is expressed as $P(B|A)$. It is easy to show that $P(B|A) = P(A \cap B)/P(A)$, where $P(A \cap B)$ is the probability that both A and B are true. As a simple example, suppose that we are told that a dice has just

[1]Dr Bernard Kachoyan is an Adjunct Associate Professor in the School of Mathematics and Statistics at UNSW Australia.

been rolled and that the rolled number is even. We are then asked to determine the probability that the number is 2. The formula for conditional probability tells us that

$$P(\text{It is a 2}\,|\text{It is even}) = \frac{P(\text{It is a 2 and it is even})}{P(\text{It is even})} = \frac{1/6}{3/6} = \frac{1}{3}\,,$$

which is obviously correct since there are only three possibilities (namely, 2, 4 and 6).

Bayes' deceptively simple theorem states that $P(A|B)P(B) = P(B|A)P(A)$. It is more usually expressed as follows.

Bayes' Theorem

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

This theorem directly relates the probability that B is true given the observation A to the probability that you *would have observed* A if B were actually true, and this can be calculated directly from an estimate of whether B is true.

As a simple example of the use of this theorem, suppose that you want to estimate the probability that it is sunny outside (event B) from an observation that a person has come in the door wearing sunglasses (event A). In order to do this you need the probability that someone would be wearing sunglasses if it were sunny and the similar probability if it were not sunny (these do not necessarily add to 1). To complete the analysis, you also need an initial estimate of the probability that it is sunny (sometimes controversially called the prior *belief* that it is sunny), based on say a weather report or what the weather was like some time ago, or even a best guess.

Just for the purposes of this example, suppose that we guess that it is generally sunny half of the time, so $P(B : \text{sunny}) = 0.5$. We then further assume, estimate or otherwise that $P(A : \text{sunglasses}|B : \text{sunny}) = 0.8$; in other words, there is an 80% chance that a person will wear sunglasses if it is sunny. Similarly, we assume that $P(\text{sunglasses}|\text{not sunny}) = 0.3$. This means that the overall probability that someone is wearing sunglasses, $P(A)$, is

$$0.8 \times 0.5 + 0.3 \times 0.5 = 0.55\,.$$

From Bayes' Theorem, we can then estimate that

$$P(B\!:\!\text{sunny}|A\!:\!\text{sunglasses}) = \frac{P(A\!:\!\text{sunglasses}|B\!:\!\text{sunny})P(\text{sunny})}{P(\text{sunglasses})} = \frac{0.8\!\times\!0.5}{0.55} = 0.73.$$

We see that if the person is wearing sunglasses, then there is roughly three-quarters chance that it is sunny outside.

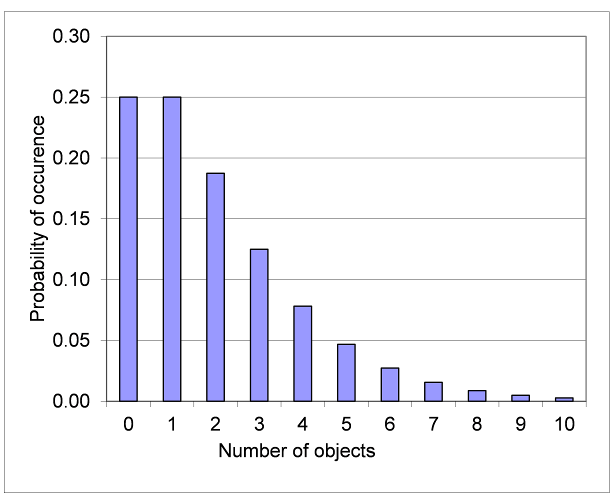
The key concept to gain from this example is that from $P(B : \text{sunny}) = 0.5$, the *prior belief* in the truth of hypothesis B, we can *after* the observation $A : \text{sunglasses}$ estimate a new value of the belief in hypothesis B, expressed as the probability 0.73. The observation of sunglasses has increased our belief in the hypothesis that it is sunny.

One of the advantages of this approach is that it can be used sequentially so that if another observation is made of someone wearing sunglasses, then we can reuse the formulas with the new "belief" value of 0.73 instead of 0.5. Similarly, if we observe someone without sunglasses, then we can use the formula again, with commensurate "no sunglasses" probability, to get a new (lower) estimate of sunniness. (Try it!)

It is hopefully clear that if we can use the formula for individual probabilities, then we can do so for probability *distributions*, simply by considering each option separately. Suppose that, instead of the simple sunny/not sunny result in the above example, we had three outcome options: sunny, partly cloudy or overcast. The *a priori* probabilities could be 0.5, 0.4 and 0.1, with commensurate probabilities of wearing sunglasses in each type of weather. An observation of whether sunglasses were worn could then be used to update each of the *a priori* probabilities. I will not go through the calculations, but I hope that I have convinced you that it is doable.

Let us now return to our original problem. Suppose that you have some idea, before you start the search, of the probability of the number of objects present. This probability might be based on word-of-mouth rumour, output from a mathematical model or even just a gut feeling. For the purposes of this example, let us suppose that this distribution is as shown in Figure 1.

Figure 1: Probability distribution of the initial estimate of the number of objects present.

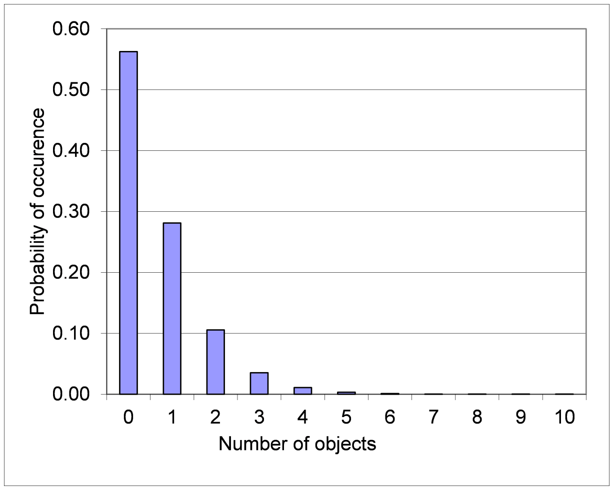


I have chosen to use a so-called *Katz distribution*, a generalization of the simple binomial distribution, as it turns out to make the maths (relatively) easier. A particularly useful property of these distributions is that

$$\frac{P(n+1)}{P(n)} = \frac{\alpha + \beta n}{n+1},$$

where $P(n)$ is the probability that the number n occurs, and α and β are related to the

Figure 2: Probability distribution of the number of objects left after a single search where none are found, using a sensor with probability of detection 0.5.



mean and standard deviation (SD) of the distribution as follows:

$$\text{mean} = \frac{\alpha}{1-\beta}\,, \qquad \text{SD} = \frac{\sqrt{\alpha}}{1-\beta}\,.$$

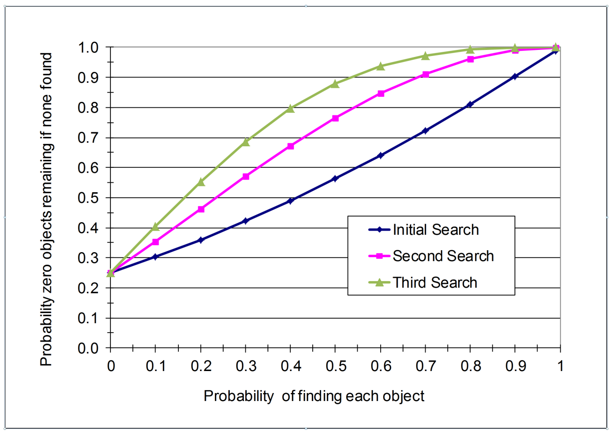
In Figure 1, we have chosen α and β so that the mean and SD of the expected number of objects are both 2.

Now suppose that we search for these objects with a sensor for which the probability of detecting each of these objects is p. Suppose also that each detection is independent and that we find N objects. What is our new (*a posteriori*) distribution for the number of objects that are left, i.e. that were there but we did not find? Using Bayes' Theorem as discussed above, it can be shown that if the *a priori* distribution is the Katz distribution with parameters α and β, the *a posteriori* distribution of how many are left is also a Katz distribution (I told you the maths would be "easier"), with new parameters $\alpha = (1-p)(\alpha + \beta N)$ and $\beta = (1-p)\beta$. It is important to note that subtracting N objects from the area does not necessarily reduce the mean number of objects expected to remain (you can compare the new mean to the old mean to demonstrate this). This may be a bit counter-intuitive but it arises because finding objects not only removes them from the sample but also provides confirmatory evidence that there are objects there to be found in the first place.

I will now illustrate with the question I originally posed in the title: What if you look and do not find anything; is nothing actually there? This now simply relates to looking at the $N = 0$ case. For a sensor with detection probability 0.5, the new distribution looks like Figure 2.

Compared to Figure 1, the probability of there being multiple objects present has gone down while the probability of there being at most 1 has increased. From Figure 2, the answer to the question "What is the chance that nothing is there, given that we did not find anything?" is 0.56. Figure 3 shows how this answer varies with the probability of detection.

Figure 3: Probability that no objects remain after a search which has found none, plotted as a function of the detection probability of the search sensor. Graphs are also shown for multiple searches of the same region, each which fails to find anything.



If $p = 0$, then we have not added any information and the chance is the same as that we started with in Figure 1, i.e., 0.25. At the other extreme, if our sensor is 100% effective, then we are sure there was nothing there to be found. The behaviour in between is of the most interest and is as shown. I have already noted that Bayes' Theorem can be used recursively, so I have also plotted how the probability increases if you search the area multiple times without finding anything. It might be interesting to note that if your sensor is less than about 40% effective, then there is still a reasonable chance (say more than 20%) that something is there even if you have searched the area three times. If you turn this statement around, then you can use this to avoid having to directly specify the sensor's performance. For instance, as long as you are confident that the sensor is more than 60% effective, then you are fairly sure, depending on your risk tolerance of course, that there is nothing to be found after three searches.

If the use of the distribution in Figure 1 seems bit artificial, then recall that the underlying methodology applies no matter what the initial distribution is, even a made-up one, such as the one where you are 60% sure there is one object there but think that there is a 20% chance that there is nothing there, a 10% chance that there are two there and a 10% chance that there are six there. It is just that the calculations become ones that you have to do directly and numerically without the nice properties that I mentioned.

As a final word, I note that there are a number of ways to approach this problem and I have obviously simplified the description of the real world quite a lot. Nevertheless, this is a good example of both the use of Bayes' Theorem and how to mathematically handle uncertainty in real-world problems.

Acknowledgement

I am drawing heavily on the work of Alan Washburn at the Naval Postgraduate School, Monterey, CA, USA, who pioneered these ideas in the areas of mine detection; see for example [Katz distributions and minefield clearance, *Military Operations Research* **11** (2006), 63–74].