Effect of Evidence on Probability

Alistair Turnbull

20th February 2004

1 Introduction

Having seen thirty white swans and one black swan, what can you say about the fraction of swans that are black? Certainly it is non-zero, and it is implausible that it is more than (say) one in five, but how implausible, and what is the most plausible value? Is it one in thirty-one? Stepping back a bit, if I had to make a decision based on the fraction of black swans, should I use a slightly larger value to be safe, because larger values, though implausible, are numerous and still possible? What is the expected value, and what is the error?

Questions analogous to the above are common in a broad variety of disciplines including law, economics, all branches of science, engineering, and literary analysis. That alone is reason to know a standard technique for answering them. They also raise an interesting philosophical point: the fraction of swans that are black is fixed (at any one time) and unique, so does it make sense to talk about its probability distribution? Surely the mode and mean are equal to the true value, and the error is zero? Or prehaps we are averaging not over events within the real world but over our belief: over the variety of possible worlds, given the evidence? No: surely the variety of *probable* worlds? But that's just the same question again. It's something to worry about on dark nights.

2 Method

The standard answer is to use Bayesian inference. Start by describing the population of possible worlds between which you wish to distinguish. Assign to them a *prior* probability distribution that reflects your personal prejudice. You have to do this, but as you amass evidence it becomes decreasingly important what you choose. Work out for each world the probability that you would observe the evidence that you do in fact observe. Finally, do some algebra to calculate the *posterior* probability distribution, telling you which possible worlds are most likely.

2.1 Scope

Bayesian inference is a very general technique, which will work no matter what structure you choose for the population of possible worlds. Here I will study a special case which includes the question of the fraction of swans that are black. I will use Bayesian inference to distinguish between worlds that differ only in the value of a single variable.

My model for the evidence is that every sample (each swan seen) is independent (the swans are all mixed up) and has a binary result (every swan is either black or white) governed by the single variable (the fraction of swans that are black). My prior probability distribution, chosen mainly for algebraic convenience, is that every value of the variable is equally probable. To be more precise, I will initially believe it equally plausible that the fraction of black swans is between 0% and 1% as between 1% and 2% as between 2% and 3%, and so on, and to any resolution you like.

2.2 The only integral

Several integrations are required in the course of solving the problem, but they are all essentially the same. They are also all quite hard, and they tend to distract and confuse, which is not good, because the big picture is itself quite demanding. This section therefore does the integral once and for all as an abstract exercise. On first reading you might want to jump from the first line of algebra in this section straight to the last, as an act of faith.

The integral in question is the following, in which a and b are non-negative integer parameters:

$$I_{ab} = \frac{(a+b)!}{a!b!} \int_0^1 p^a (1-p)^b \mathrm{d}p$$

I will do the integral using induction. First consider the case in which $a = 0$:

$$\begin{aligned} I_{0b} &= \frac{b!}{b!} \int_0^1 (1-p)^b \mathrm{d}p \\ &= \frac{-1}{b+1} \left[(1-p)^{b+1} \right]_0^1 \\ &= \frac{1}{b+1} \end{aligned}$$

If we are not so fortunate as to have $a = 0$, we must reduce a one step at a time:

$$\begin{aligned}
I_{ab} &= \frac{(a+b)!}{a!b!}\int_0^1 p^a(1-p)^b \mathrm{d}p \\
&= \frac{-1}{b+1}\frac{(a+b)!}{a!b!}\left[p^a(1-p)^{b+1}\right]_0^1 + \frac{a}{b+1}\frac{(a+b)!}{a!b!}\int_0^1 p^{(a-1)}(1-p)^{(b+1)}\mathrm{d}p \\
&= 0 + \frac{(a+b)!}{(a-1)!(b+1)!}\int_0^1 p^{(a-1)}(1-p)^{(b+1)}\mathrm{d}p \\
&= I_{(a-1)(b+1)}
\end{aligned}$$

Thus we may conclude for all non-negative integers a and b that:

$$I_{ab} = \frac{1}{a+b+1}$$

2.3 Probability of evidence given model

Consider a world in which one swan in thirty-one is black. What is the chance that of the first thirty-one swans I see exactly one is black? That is the subject of this section. I will use the notation $P(1, 30|\frac{1}{31})$ for the answer. More generally, I will write $P(a, b|p)$ for the probability of seeing a black swans and b white swans in a world where the fraction of black swans is p.

The distribution is a binomial distribution. We can immediately write a formula for $P(a, b|p)$ in terms of a, b and p using our knowledge of binomial distributions. There are $\frac{(a+b)!}{a!b!}$ ways of seeing a black swans and b white ones, and each way has a probability of $p^a(1-p)^b$. Therefore:

$$P(a, b|p) = \frac{(a+b)!}{a!b!}p^a(1-p)^b$$

2.4 Probability of evidence

What is the probability of seeing one black swan out of thirty-one when we don't know the fraction of black swans? I will use the notation $P(1, 30)$ for the answer. More generally, I will write $P(a, b)$ for the probability of seeing a black swans and b white swans given only personal prejudices about the fraction of black swans in the world.

To work out the answer we must average over our uncertainty about the world. I will use the notation $P(p)$ for the probability density that we live in a world in which the fraction of black swans is p. To be more precise, $P(p)\mathrm{d}p$ is the probability that the fraction of black swans lies in a band centred on p of infinitesimal width $\mathrm{d}p$. As I explained in the introduction, my personal prejudice, easily swayed for convenience, is that $P(p) = 1$ for all $0 \leq p \leq 1$.

The calculation is based on two standard formulae:

$$\begin{aligned} P(X\&Y) &= P(X|Y)P(Y) \\ P(X) &= \sum_Y P(X\&Y) \end{aligned}$$

We can apply the first formula fairly directly:

$$\begin{aligned} P(a,b\&p)\mathrm{d}p &= P(a,b|p)P(p)\mathrm{d}p \\ \therefore P(a,b\&p) &= P(a,b|p)P(p) \\ &= \frac{(a+b)!}{a!b!}p^a(1-p)^b \end{aligned}$$

Because p is a continuous variable, the sum in the second formula becomes an integral:

$$\begin{aligned} P(a,b) &= \int_0^1 P(a,b\&p)\mathrm{d}p \\ &= \frac{(a+b)!}{a!b!}\int_0^1 p^a(1-p)^b\mathrm{d}p \\ &= I_{ab} \\ &= \frac{1}{a+b+1} \end{aligned}$$

2.5 Probability of model given evidence

Now for the original question: what is the probability density that we live in a world in which the fraction of black swans is p, given (our prior prejudice and) that we have seen one black swan and thirty white swans? I will write $P(\frac{1}{31}|1,30)$ for the answer. More generally and precisely, $P(p|a,b)\mathrm{d}p$ is the probability, given that we have seen a black swans and b white swans, that the fraction of black swans lies in a band centred on p of width dp.

It is here that we use Bayesian inference. The calculation is based on the first of the two formulae used in the previous section. It should come as no surprise that we need the same formula again, since it is the definition of conditional probability. This time we use it backwards:

$$P(X|Y) = \frac{P(X\&Y)}{P(Y)}$$

We also use it with the roles of X and Y interchanged: this time X is a model and Y is the evidence.

$$\begin{aligned}
P(p|a,b)\mathrm{d}p &= \frac{P(p\&a,b)\mathrm{d}p}{P(a,b)} \\
\therefore P(p|a,b) &= \frac{P(a,b\&p)}{P(a,b)} \\
&= \frac{\frac{(a+b)!}{a!b!}p^a(1-p)^b}{\frac{1}{a+b+1}} \\
&= \frac{(a+b+1)!}{a!b!}p^a(1-p)^b
\end{aligned}$$

And that is the final answer. It would be wise to check that the probabilities add to one:

$$\begin{aligned}
\int_0^1 P(p|a,b)\mathrm{d}p &= \frac{(a+b+1)!}{a!b!}\int_0^1 p^a(1-p)^b\mathrm{d}p \\
&= (a+b+1)I_{ab} \\
&= \frac{a+b+1}{a+b+1} \\
&= 1
\end{aligned}$$

3 Results

Armed with a formula that allows us to quantify the effect of evidence on belief:

$$P(p|a,b) = \frac{(a+b+1)!}{a!b!}p^a(1-p)^b$$

we can now calculate some of the statistics requested in the opening paragraph. Specifically, I will calculate the mode (the most plausible value of p), and the mean and standard deviation (the statistics needed to make decisions that recognise the uncertainty).

3.1 Mode

The most plausible value for the fraction of black swans is the value with the highest probability density, given the evidence. In other words, it is the value of p which maximises $P(p|a,b)$. Because $P(p|a,b)$ is a smooth curve, its maximum will be a point where its derivative is zero, or possibly at $p = 0$ or $p = 1$. Setting the derivative to zero and solving for p, we obtain several solutions one of which is obviously the maximum.

First let's calculate the derivative:

$$\begin{aligned}\frac{\mathrm{d}P(p|a,b)}{\mathrm{d}p} &= \frac{(a+b+1)!}{a!b!}\frac{\mathrm{d}p^a(1-p)^b}{\mathrm{d}p} \\ &= \frac{(a+b+1)!}{a!b!}\left(ap^{(a-1)}(1-p)^b - bp^a(1-p)^{(b-1)}\right)\end{aligned}$$

Here we should pause to consider the cases in which a and b are small:

- If both are zero, then the derivative is everywhere zero, because the probability density is everywhere one. That's my prior probability distribution, as you'd expect, since we have not yet seen any swans. There is no unique maximum.
- If $a = 0$ and $b > 0$ then the derivative is non-zero everywhere between $p = 0$ and $p = 1$, so the maximum must be at $p = 0$ or $p = 1$. Returning to the original formula, we find $P(p|0,b) = (b+1)(1-p)^b$, which is one at $p = 0$ and zero at $p = 1$, so the maximum is at $p = 0$. Having seen no black swans, it is most plausible to believe that all swans are white.
- A similar argument applies if $a > 0$ and $b = 0$.

Let us then proceed, assuming $a > 0$ and $b > 0$:

$$\begin{aligned}0 &= \frac{\mathrm{d}P(p|a,b)}{\mathrm{d}p} \\ &= \frac{(a+b+1)!}{a!b!}\left(p^{(a-1)}(1-p)^{(b-1)}\right)(a(1-p)-bp)\end{aligned}$$

The initial constant is never zero. The middle bracket can be zero only if $p = 0$ or $p = 1$, but the probability is zero at both these points, so they are minima. The maximum must therefore be where the third bracket is zero:

$$\begin{aligned}0 &= a(1-p)-bp \\ \therefore p &= \frac{a}{a+b}\end{aligned}$$

Observe that this formula also gives the right answer when $a = 0$ or $b = 0$ and gives an undefined result when both $a = 0$ and $b = 0$. In effect it neatly summarises the earlier case analysis, and we can quote it as our final answer for all cases. It also pleasingly in agreement with intuition: after seeing one black swan and thirty white swans, the most plausible hypothesis is that one swan in every thirty-one is black.

3.2 Mean

The expected (or mean) fraction of black swans is obviously different from the most plausible value in cases with $a = 0$ and $b > 0$, *i.e.* after seeing only white swans. The modal value is $p = 0$, but there is obviously still a chance that we might later see a black swan hiding behind a bush somewhere. We would certainly be foolish to bet our worldly wealth on the impossibility of a black swan having seen just one swan which happened to be white.

In general, the mean of a function f that depends on an uncertain variable X is defined by the following standard formula:

$$\bar{f} = \sum_X f(X)P(X)$$

In our case the uncertain variable is p, and we want to find the expected fraction of black swans, so the function f is also just p. As usual, the sum becomes an integral because p is a continuous variable:

$$\begin{aligned}
\bar{p} &= \int_0^1 pP(p|a,b)\mathrm{d}p \\
&= \frac{(a+b+1)!}{a!b!}\int_0^1 p^{(a+1)}(1-p)^b\mathrm{d}p \\
&= (a+1)I_{(a+1)b} \\
&= \frac{a+1}{a+b+2}
\end{aligned}$$

This formula also has a pleasing intuition: to calculate the mean we should use the same formula as to calculate the mode, but to be safe we should include an extra swan of each colour in addition to the ones we have actually seen. Although we think it is most plausible that the fraction of black swans is $\frac{1}{31}$, we would be wise to bet that it is $\frac{2}{33}$.

3.3 Standard deviation

The standard deviation of a function f is a standard way of quantifying the uncertainty of its expected value. It is defined to be the square-root of its variance, which in turn is calculated using a standard formula:

$$\begin{aligned}
\mathrm{var}(f) &= \overline{(f-\bar{f})^2} \\
&= \overline{(f)^2} - (\bar{f})^2
\end{aligned}$$

In our case, the function f is again just p:

$$\begin{aligned}
\mathrm{var}(p) &= \int_0^1 p^2 P(p|a,b)\mathrm{d}p - (\bar{p})^2 \\
&= \frac{(a+b+1)!}{a!b!}\int_0^1 p^2 p^a (1-p)^b \mathrm{d}p - \left(\frac{a+1}{a+b+2}\right)^2 \\
&= \frac{(a+1)(a+2)}{a+b+2} I_{(a+2)b} - \left(\frac{a+1}{a+b+2}\right)^2 \\
&= \frac{(a+1)(a+2)}{(a+b+2)(a+b+3)} - \frac{(a+1)^2}{(a+b+2)^2} \\
&= \frac{(a+1)\left((a+2)(a+b+2) - (a+1)(a+b+3)\right)}{(a+b+2)^2(a+b+3)} \\
&= \frac{(a+1)(b+1)}{(a+b+2)^2(a+b+3)} \\
&= \frac{\bar{p}(1-\bar{p})}{a+b+3}
\end{aligned}$$

Thus the variance is largest if $\bar{p}$ is middling (about $\frac{1}{2}$) and decreases in inverse proportion to the amount of evidence available. The standard deviation therefore decreases in inverse proportion to the square-root of the amount of evidence available. After seeing one black and thirty white swans, the standard deviation in our estimate that the fraction of black swans is $\frac{2}{33}$ is $\sqrt{\frac{2\times 31}{33\times 33\times 34}}$ which is roughly 0.04, or about 1.3 swans in 33.

4 Conclusion

Bayesian inference is tractable and powerful enough to answer all the questions in the opening paragraph, and many similar and more complex questions too. However, Bayes asks you to produce a prior probability distribution, and regarding late-night worries about philosophy he therefore merely passes the buck (sensible chap).