LAPLACE'S RULE OF SUCCESSION

Jonathan Weisberg

The Rule of Succession gives a simple formula for "enumerative induction": reasoning from observed instances to unobserved ones. If you've observed 8 ravens and they've all been black, how certain should you be the next raven you see will also be black? According to the Rule of Succession, 90%. In general, the probability is $(k+1)/(n+2)$ that the next observation will be positive, given k positive observations out of n total.

When does the Rule of Succession apply, and why is it $(k+1)/(n+2)$? Laplace first derived a special case of the rule in 1774, using certain assumptions. The same assumptions also allow us to derive the general rule, and following the derivation through answers both questions.

As motivation, imagine we're drawing randomly, with replacement, from an urn of marbles some proportion p of which are black. Strictly speaking, p must be a rational number in this setup. But formally, we'll suppose p can be any real number in the unit interval.

If we have no idea what p is, it's natural to start with a uniform prior over its possible values. Formally, p is a random variable with a uniform density on the $[0,1]$ interval. Each draw induces another random variable,

$$X_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ draw is black,} \\ 0 & \text{otherwise.} \end{cases}$$

We'll define one last random variable S_n, which counts the black draws:

$$S_n = X_1 + \ldots + X_n.$$

Laplace's assumptions are then as follows.

1. Each X_i has the same chance p of being 1.
2. That chance is independent of whatever values the other X_j's take.
3. The prior distribution over p is uniform: $f(p) = 1$ for $0 \leq p \leq 1$.

Given these assumptions, the Rule of Succession follows:

$$P(X_{n+1} = 1 \mid S_n = k) = \frac{k+1}{n+2}.$$

We'll start by deriving this result for the special case where all observations are positive, so that $k = n$.

1 LAPLACE'S SPECIAL CASE

When $k = n$, the Rule of Succession says:

$$P(X_{n+1} = 1 \mid S_n = n) = \frac{n+1}{n+2}.$$

To derive this result, we start with the Law of Total Probability.

$$\begin{aligned} P(X_{n+1} = 1 \mid S_n = n) &= \int_0^1 P(X_{n+1} = 1 \mid S_n = n, p) f(p \mid S_n = n)\, \mathrm{d}p \\ &= \int_0^1 P(X_{n+1} = 1 \mid p) f(p \mid S_n = n)\, \mathrm{d}p \\ &= \int_0^1 p\, f(p \mid S_n = n)\, \mathrm{d}p. \end{aligned} \tag{1}$$

To finish the calculation, we need to compute $f(p \mid S_n = n)$. We need to know how observing n out of n black marbles changes the probability density over p.

For this we turn to Bayes' theorem.

$$\begin{aligned} f(p \mid S_n = n) &= \frac{f(p) P(S_n = n \mid p)}{P(S_n = n)} \\ &= \frac{P(S_n = n \mid p)}{P(S_n = n)} \\ &= \frac{p^n}{P(S_n = n)} \\ &= c p^n. \end{aligned}$$

Here c is an as-yet unknown constant: the inverse of $P(S_n = n)$, whatever that is. To find c, first observe by calculus that:

$$\int_0^1 c p^n \, \mathrm{d}p = \left(\frac{c p^{n+1}}{n+1} \right) \Big|_0^1 = \frac{c}{n+1}.$$

Then observe that this quantity must equal 1, since we've integrated $f(p \mid S_n = n)$, a probability density. Thus $c = n + 1$, and hence

$$f(p \mid S_n = n) = (n+1)p^n.$$

Returning now to finish our original calculation in Equation (1):

$$\begin{aligned} P(X_{n+1} = 1 \mid S_n = n) &= \int_0^1 p\, f(p \mid S_n = n)\, \mathrm{d}p \\ &= \int_0^1 p\,(n+1)p^n\, \mathrm{d}p \\ &= (n+1) \int_0^1 p^{n+1}\, \mathrm{d}p \\ &= (n+1) \left(\frac{p^{n+2}}{n+2} \right) \Bigg|_0^1 \\ &= \frac{n+1}{n+2}. \end{aligned}$$

This is the Rule of Succession when $k = n$, as desired.

2 THE GENERAL CASE

The proof of the general case starts similarly. We first apply the Law of Total Probability to obtain

$$P(X_{n+1} = 1 \mid S_n = k) = \int_0^1 p\, f(p \mid S_n = k)\, \mathrm{d}p. \tag{2}$$

Then we use Bayes' theorem to compute $f(p \mid S_n = k)$.

$$\begin{aligned} f(p \mid S_n = k) &= \frac{P(S_n = k \mid p)}{P(S_n = k)} \\ &= \frac{\binom{n}{k} p^k (1-p)^{n-k}}{P(S_n = k)}. \end{aligned} \tag{3}$$

Note that we used the formula for a binomial probability here to calculate the numerator $P(S_n = k \mid p)$.

Computing the denominator $P(S_n = k)$ requires a different approach from the special case. We start with the Law of Total Probability:

$$P(S_n = k) = \int_0^1 P(S_n = k \mid p) f(p)\, \mathrm{d}p$$

$$
\begin{aligned}
&= \int_0^1 P(S_n = k \mid p)\,\mathrm{d}p \\
&= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k}\,\mathrm{d}p \\
&= \binom{n}{k} \int_0^1 p^k (1-p)^{n-k}\,\mathrm{d}p.
\end{aligned}
$$

This leaves us facing an instance of a famous function, the "beta function," which is defined:

$$B(a,b) = \int_0^1 x^a (1-x)^b\,\mathrm{d}x.$$

In our case a and b are natural numbers, so $B(a,b)$ has an elegant formula, which we use now and prove later:

$$B(a,b) = \frac{a!b!}{(a+b+1)!}.$$

For us, $a = k$ and $b = n - k$, so we have

$$P(S_n = k) = \binom{n}{k} B(k, n-k) = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!}.$$

Substituting back into our calculation of $f(p \mid S_n = k)$ in Equation (3):

$$
\begin{aligned}
f(p \mid S_n = k) &= \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k} B(k, n-k)} \\
&= \frac{(n+1)!}{k!(n-k)!} p^k (1-p)^{n-k}.
\end{aligned}
$$

Then we finish our original calculation from Equation (2):

$$
\begin{aligned}
P(X_{n+1} = 1 \mid S_n = k) &= \int_0^1 p \frac{(n+1)!}{k!(n-k)!} p^k (1-p)^{n-k}\,\mathrm{d}p \\
&= \frac{(n+1)!}{k!(n-k)!} \int_0^1 p^{k+1} (1-p)^{n-k}\,\mathrm{d}p \\
&= \frac{(n+1)!}{k!(n-k)!} B(k+1, n-k) \\
&= \frac{(n+1)!}{k!(n-k)!} \frac{(k+1)!(n-k)!}{(k+1+n-k+1)!} \\
&= \frac{k+1}{n+2}.
\end{aligned}
$$

This is the Rule of Succession, as desired.

3 THE BETA FUNCTION

Finally, let's derive the formula we used for the beta function:

$$\int_0^1 x^a (1-x)^b \, dx = \frac{a!b!}{(a+b+1)!},$$

where a and b are natural numbers. We proceed in two steps: integration by parts, then a proof by induction.

Notice first that when $b = 0$ our integral simplifies and is straightforward:

$$\int_0^1 x^a \, dx = \frac{1}{a+1}.$$

So let's assume $b > 0$ and pursue integration by parts. If we let

$$u = (1-x)^b, \quad dv = x^a \, dx,$$

then

$$du = -b(1-x)^{b-1}, \quad v = \frac{x^{a+1}}{a+1}.$$

So

$$\begin{aligned} \int_0^1 x^a (1-x)^b \, dx &= \left(\frac{x^{a+1}(1-x)^b}{a+1} \right) \Bigg|_0^1 + \frac{b}{a+1} \int_0^1 x^{a+1} (1-x)^{b-1} \, dx \\ &= \frac{b}{a+1} \int_0^1 x^{a+1} (1-x)^{b-1} \, dx. \end{aligned}$$

Now we use this identity in an argument by induction. We already noted that when $b = 0$ we have $B(a, 0) = 1/(a+1)$. This satisfies the general formula

$$B(a, b) = \frac{a!b!}{(a+b+1)!}.$$

By induction on $b > 0$, we find the formula holds in general:

$$\begin{aligned} B(a,b) &= \int_0^1 x^a (1-x)^b \, dx \\ &= \frac{b}{a+1} \int_0^1 x^{a+1} (1-x)^{b-1} \, dx \\ &= \frac{b}{a+1} B(a+1, b-1) \\ &= \frac{b}{a+1} \frac{(a+1)!(b-1)!}{(a+1+b-1+1)!} \\ &= \frac{a!b!}{(a+b+1)!}. \end{aligned}$$

4 ACKNOWLEDGMENTS

Our proof of the special case follows this excellent video by Joe Blitzstein. And our proof of the general case comes from Sheldon Ross' classic textbook, *A First Course in Probability*, Exercise 30 on page 128 of the 7th edition.