

CRASH COURSE IN INDUCTIVE LOGIC

JONATHAN WEISBERG

There are four ways things can turn out with two flips of a coin:

$$HH, HT, TH, TT.$$

If we know nothing about the coin's tendencies, we might assign equal probability to each of these four possible outcomes:

$$Pr(HH) = Pr(HT) = Pr(TH) = Pr(TT) = 1/4.$$

But from another point of view, there are primarily three possibilities. If we ignore order, the possible outcomes are 0 heads, 1 head, or 2 heads. So we might instead assign equal probability to these three outcomes, then divide the middle $1/3$ evenly between HT and TH :

$$Pr(HH) = 1/3 \quad Pr(HT) = Pr(TH) = 1/6 \quad Pr(TT) = 1/3.$$

This two-stage approach may seem awkward at first. But it's actually friendlier from the point of view of inductive reasoning. On the first scheme, a heads on the first toss doesn't increase the probability of another heads. It stays fixed at $1/2$:

$$Pr(HH | H) = \frac{1/4}{1/4 + 1/4} = \frac{1}{2}.$$

Whereas it does increase on the second strategy, from $1/2$ to $2/3$:

$$Pr(HH | H) = \frac{1/3}{1/3 + 1/6} = \frac{2}{3}.$$

The two-stage approach thus learns from experience, where the single-step division is skeptical about induction.

This holds true as we increase the number of flips. If we do three tosses for example, we'll find that $Pr(HHH | HH) = 3/4$ on the two-stage analysis. Whereas this probability stays stubbornly fixed at $1/2$ on the first approach. It won't budge no matter how many heads we observe, so we can't learn anything about the coin's bias this way.

This is the difference between Carnap's famous account of induction, from his 1950 book *Logical Foundations of Probability*, and the account he finds in *Wittgenstein's Tractatus*.¹ Although Carnap had actually been scooped by W. E. Johnson, who worked out a similar analysis about 25 years earlier.

This is a short explainer on some key elements of inductive logic that Johnson and Carnap worked out, and the place of those ideas in the story of inductive logic.

1 STATES & STRUCTURES

Carnap called fine-grained specifications like TH , *state-descriptions*. The coarser grained "1 head" is a *structure-description*. A state-description specifies which flips land heads, and which tails. While a structure-description specifies *how many* land heads and tails, without necessarily saying which.

It needn't be coin flips landing heads or tails, of course. The same ideas apply to any set of objects or events, and any feature they might have or lack.

Suppose we have two objects a and b , each of which might have some property F . Working for a moment as Carnap did, in first-order logic, here is an example of a structure-description:

$$(Fa \ \& \ \sim Fb) \vee (\sim Fa \ \& \ Fb).$$

But this isn't a state-description, since it doesn't specify which object has F . It only says how many objects have F , namely 1. One of the disjuncts alone would be a state-description though:

$$Fa \ \& \ \sim Fb.$$

¹ Carnap also cites Keynes and Peirce as endorsing the Wittgensteinian approach. But thanks to Jonathan Livengood I learned this is actually a misattribution: Keynes mistakenly attributes the view to Peirce, and Carnap seems to have followed Keynes' error.

Carnap's idea was that all structure-descriptions start out with the same probability. And these probabilities are then divided equally among the state-descriptions that make up a structure-description.

For example, if we do three flips, there are four structure-descriptions: 0 heads, 1 head, 2 heads, and 3 heads. Some of these have only one state-description. For example, there's only one way to get 0 heads, namely *TTT*. So

$$Pr(TTT) = 1/4.$$

But others have multiple state-descriptions. There are three ways to get 1 head for example, so we divide $1/4$ between them:

$$Pr(HTT) = Pr(THT) = Pr(TTH) = 1/12.$$

The effect is that more homogeneous sequences start out more probable. There's only one way to get all heads, so the *HH* state-description inherits the full probability of the corresponding "2 heads" structure-description. But a 50-50 split has multiple permutations, each of which inherits only a portion of the same quantum of probability. A heterogeneous sequence of heads and tails thus starts out less probable than a homogeneous one.

That's why the two-stage analysis is induction-friendly. It effectively builds Hume's "uniformity of nature" assumption into the prior probabilities.

2 THE RULE OF SUCCESSION

The two-stage assignment also yields a very simple formula for induction: Laplace's famous Rule of Succession. (Derivation in the [Appendix](#).)

THE RULE OF SUCCESSION Given k heads out of n observed flips, the probability of heads on a subsequent toss is

$$\frac{k+1}{n+2}.$$

Laplace arrived at this rule about 150 years earlier by somewhat different means. But there is a strong similarity.

Laplace supposed that our coin has some fixed, but unknown, chance p of landing heads on each toss. Suppose we regard all possible values $0 \leq p \leq 1$

as equally likely.² If we then update our beliefs about the true value of p using Bayes' theorem, we arrive at the Rule of Succession. (Proving this is a bit involved. Maybe I'll go over it another time.)

Johnson and Carnap's way of assigning prior probabilities is essentially the same idea, just applied in a discrete setting. By treating all structure-descriptions as equiprobable, we make all possible frequencies of heads equiprobable. This is a discrete analogue of treating all possible values of p as equiprobable.

3 THE CONTINUUM OF INDUCTIVE METHODS

Both Johnson and Carnap eventually realized that the two methods of assigning priors we've considered so far are just two points on a larger continuum.

THE λ CONTINUUM Given k heads out of n observed flips, the probability of heads on a subsequent toss is

$$\frac{k + \lambda/2}{n + \lambda},$$

for some λ in the range $0 \leq \lambda \leq \infty$.

What value should λ take here? Notice we get the Rule of Succession if $\lambda = 2$. And we get inductive skepticism if we let λ approach ∞ . For then k and n fall away and the ratio converges to $1/2$, no matter what k and n are.

If we set $\lambda = 0$, we get a formula we haven't discussed yet: k/n . Reichenbach called this the Straight Rule. (In modern statistical parlance it's the "maximum likelihood estimate.")³

The overall pattern is: the higher λ , the more "cautious" our inductive inferences will be. A larger λ means less influence from k and n : the probability of another heads stays closer to the initial value of $1/2$. In the extreme case where $\lambda = \infty$, it stays stuck at exactly $1/2$ forever.

A low value of λ , on the other hand, will make our inferences more ambitious. In the extreme case $\lambda = 0$, we jump immediately to the observed frequency. Our expectation about the next toss is just k/n , the frequency

² More precisely: we regard them as having the same probability *density*, namely 1.

³ For the $\lambda = 0$ case, we need probability axioms that permit conditioning on zero-probability events. For example, $Pr(HH | H) = 1$ so $Pr(HT | H) = 0$. Thus $Pr(HT) = 0$, and $Pr(HTH | HT)$ is undefined on the usual, Kolmogorov axioms.

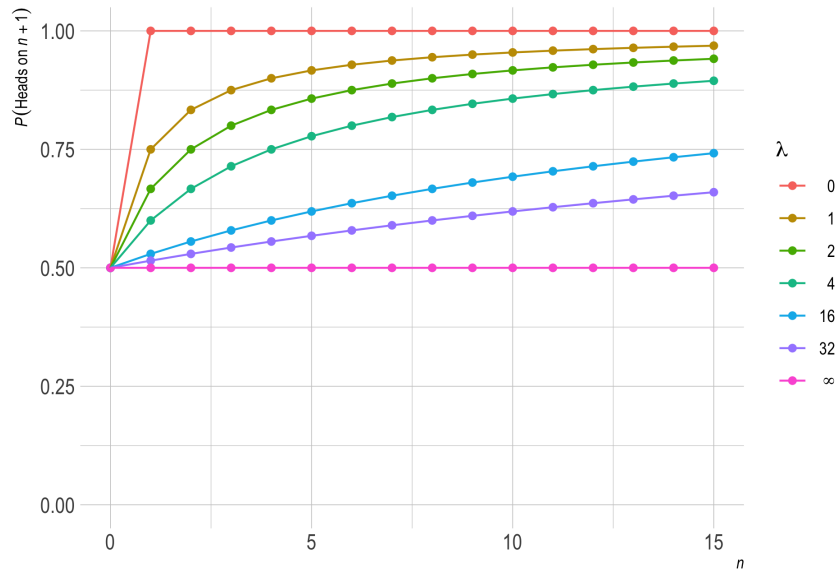


Figure 1: Some λ values with heads on every flip

we've observed so far. If we've observed only one flip and it was heads ($k = n = 1$), we'll be certain of heads on the second toss!⁴

We can illustrate the effect of various settings of λ with a plot. If the coin keeps coming up heads ($k = n$), Figure 1 shows what happens as n increases. Figure 2 shows the case where the coin only lands heads every third time, so that $k \approx n/3$. Notice how lower settings of λ bounce around more here before settling into roughly $1/3$. Higher settings approach $1/3$ more steadily, but they take longer to get there.

4 CARNAP'S PROGRAM

Johnson and Carnap went much further, and others since have gone further still. For example, we can include more than one predicate, we can use relational predicates, and much more.

But philosophers aren't too big on this research program nowadays. Why not?

Choosing λ is one issue. Once we see that it's more than a binary choice, between inductive optimism and skepticism, it's hard to see why we should

⁴ When $n = 0$ we have to stipulate that the probability is $1/2$, the limit as $\lambda \rightarrow 0$.

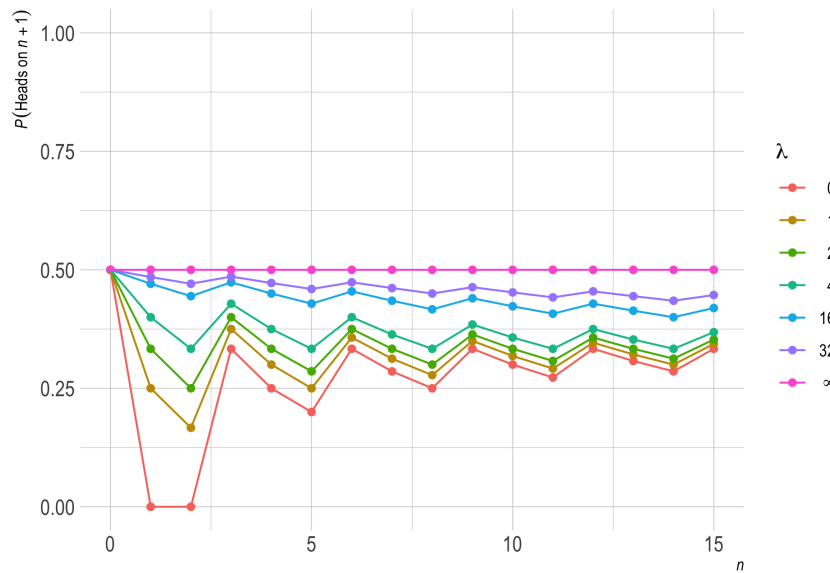


Figure 2: Some λ values with heads on every third flip

plump for any particular value of λ . We could set $\lambda = 2$, or π , or 42 . By what criterion could we make this choice? No clear answer emerged from Carnap's program.

Another issue is Goodman's famous *grue puzzle*. Suppose we trade our coin flips for emeralds. We might replace the heads/tails dichotomy with green/not-green then. But we could instead replace it with grue/not-grue. The prescriptions of Carnap's inductive logic depend on our choice of predicate—on the underlying language to which we apply our chosen value of λ .

So the Johnson/Carnap system doesn't provide us with rules for inductive reasoning, more a framework for formulating such rules. We have to decide which predicates should be projectible by choosing the underlying language. And then we have to decide how projectible they should be by choosing λ . Only then does the framework tell us what conclusions to draw from a given set of observations.

Personally, I still find Carnap's framework useful. It provides a lovely way to express informal ideas more rigorously. In it we can frame questions about induction, skepticism, and prior probabilities with lucidity.

I also like it as a source of toy models. For example, I might test when a given claim about induction holds and when it doesn't, by playing with different incarnations of λ .

The utility of Carnap's framework is thus a lot like that of its deductive cousins. Compare Timothy Williamson's use of modal logic to create **models of Gettier cases**, for example, or his model of **improbable knowledge**.

Even in deductive logic, we only get as much out as we put in. We have to choose our connectives in propositional logic, our accessibility relation in modal logic, etc. But a flexible system like possible-world frames still has its uses, because we can use it to explore philosophical options and their interconnections.

5 FURTHER READINGS

For more on this topic, here are some suggested readings.

- **Inductive Logic**, by Branden Fitelson
- **Carnap on Probability and Induction**, by Sandy Zabell
- **The IEP entry on Confirmation and Induction**, by Franz Huber
- **The SEP entry on Inductive Logic**, by James Hawthorne
- **Pure Inductive Logic**, by Jeffrey Paris and Alena Vencovská

And if you want to see how Johnson and Carnap's two-stage assignment of priors yields the Rule of Succession, check out the Appendix.

6 APPENDIX: DERIVING THE RULE OF SUCCESSION

To derive the Rule of Succession from the two-stage assignment of priors, we need two key formulas.

1. The prior probability of a particular sequence with k heads out of n flips.
2. The prior probability of the same initial sequence, followed by one more heads.

The first quantity is the probability of getting k heads out of n flips, regardless of order, divided by the number of ways to get k heads out of n flips. The

number of ways to get k heads out of n flips is called the **binomial coefficient**. It's written $\binom{n}{k}$, and there's a nice formula for calculating it:

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

Since a sequence of n flips can feature anywhere from 0 to n heads, there are $n+1$ structure descriptions, each with probability $1/(n+1)$. Thus the probability of a specific state-description with k heads out of n flips is

$$\begin{aligned} \frac{1}{(n+1)\binom{n}{k}} &= \frac{1}{(n+1)\frac{n!}{(n-k)!k!}} \\ &= \frac{(n-k)!k!}{(n+1)!}. \end{aligned} \tag{1}$$

The second probability we need is for the same initial sequence, but with an additional heads on the next toss. That's a sequence with $k+1$ heads out of $n+1$ tosses. There are $n+2$ structure descriptions now, each with probability $1/(n+2)$. So the probability in question is

$$\begin{aligned} \frac{1}{(n+2)\binom{n+1}{k+1}} &= \frac{1}{(n+2)\frac{(n+1)!}{(n-k)!(k+1)!}} \\ &= \frac{(n-k)!(k+1)!}{(n+2)!}. \end{aligned} \tag{2}$$

Now, to get the conditional probability we're after, we take the ratio of the second probability (2) over the first probability (1):

$$\begin{aligned} \frac{\frac{(n-k)!(k+1)!}{(n+2)!}}{\frac{(n-k)!k!}{(n+1)!}} &= \frac{(n-k)!(k+1)!}{(n+2)!} \frac{(n+1)!}{(n-k)!k!} \\ &= \frac{k+1}{n+2}. \end{aligned}$$

This agrees with the rule of succession, as desired.

So far though, we've only shown the rule of succession for a specific, observed sequence. We've shown that $Pr(HTHH | HTH) = 3/4$, for example. But what if we don't know the particular sequence so far? Maybe we only know there were 2 heads out of 3 tosses. Shouldn't we still be able to derive the same result?

We can, with the help of a relevant theorem of probability: if $Pr(A | B) = Pr(A | C)$, and B and C are mutually exclusive, then

$$Pr(A | B \vee C) = Pr(A | B) = Pr(A | C).$$

In our case A specifies heads on flip $n + 1$, while B and C each specify some sequence for flips 1 through n . Although these sequences feature the same number of heads and tails, B and C specify different orderings. So they're mutually exclusive.

We've already shown that

$$Pr(A | B) = \frac{k+1}{n+2} = Pr(A | C).$$

So we just have to verify the theorem:

$$\begin{aligned} Pr(A | B \vee C) &= \frac{Pr(A \& (B \vee C))}{Pr(B \vee C)} \\ &= \frac{Pr(A \& B) + Pr(A \& C)}{Pr(B \vee C)} \\ &= \frac{Pr(A | B)Pr(B) + Pr(A | C)Pr(C)}{Pr(B \vee C)} \\ &= \frac{Pr(A | B) (Pr(B) + Pr(C))}{Pr(B \vee C)} \\ &= Pr(A | B). \end{aligned}$$

By applying this formula repeatedly to a disjunction of state-descriptions, we get the conditional probability on the structure description of interest.