Goodman's new riddle of induction in statistical terms

Uwe Saint-Mont, Nordhausen University of Applied Sciences

August 24, 2026

Abstract. Goodman's new riddle of induction has mainly been treated in verbal terms. Its core concept, projectabiliy, has much in common with statistical estimation. Therefore, understanding when and why the latter approach works, also sheds much light on Goodman's riddle.

Keywords

Goodman; riddle of induction; statistical estimation; Tukey's example

1 Introduction

During the last seventy years, philosophers of science have written much about the 'new riddle of induction' (Goodman 1955): all emeralds observed so far have been green, thus it is straightforward to project this property into the future. Writing '0' for the property of being green and indicating the prognosis with $\rightarrow$, we thus have:

$$0, 0, 0, 0, 0, 0, \ldots \rightarrow 0$$

Figure 1: a constant sequence

However, the same evidence also supports the property of being 'grue', i.e., that an object is green (0) up to time t and blue (1) afterwards. In a picture:

time:	1	2	3	4	...	t		$t+1$	
	0,	0,	0,	0,	...,	0	$\rightarrow$	0	'green'
							$\searrow$		
								1	'grue'

Figure 2: Goodman's riddle

In this guise the issue of induction becomes the question whether a certain property is *projectable* or not. The color of an emerald or the electrical conductivity of copper seem to be projectable, however (Goodman 1955), the 'fact that a given man now in this room is a third son does not increase the credibility of statements asserting that other men now in this room are third sons, and so does not confirm the hypothesis that all men now in this room are third sons.'

Of course, it is straightforward to generalize the last figure to more than two colors, to assume that t as a variable, or to introduce several moments in time $t_1, t_2, \ldots$ when an emerald might change its color. However, for the purpose of this article, it suffices to consider the above elementary model which represents the basic problem of induction in a very simple way.

Potential answers to Goodman's riddle have been rather verbal in nature and differ widely (Kovalenko 2022, Cohnitz et al. 2024). Thus, it seems to be a good idea to translate the riddle into a suitable formal framework, and to analyze it there with a greater amount of precision. Consistently, this article couches the problem in stochastic terms and employs basic statistical theory to reach a transparent conclusion.

2 A probabilistic framework

Although statistics deals with real-world data, i.e., numbers with a lot of context to them, it is based on probability theory, which, of course, has to be purely formal, void of the meaning any particular variable might have. In this setting, the fundamental problem of induction typically shows up in generalizing from a finite sample at hand to a larger not observed or even non-observable (e.g. infinite) population.

To this end, statistics' standard model is that of a random variable X with distribution $\mathcal{D}(X)$, corresponding distribution function F, density f, and specific parameters such

as the expected value μ or the variance σ^2. For example, in short notation $X \sim N(\mu, \sigma^2)$ if the distribution is a Normal with $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$. Given a sample from such a population, one would expect that is similar to the underlying distribution, and that the sample's mean $\bar{x} = (x_1 + \ldots, x_n)/n$ converges towards μ in a probabilistic sense.

This guess, a statistical analog to fig. 1, turns out to be correct (see figs. 3, 4). Given independent random variables X_i, all being distributed according to $N(\mu, \sigma^2)$, their mean $\bar{X}_n = (X_1 + \ldots + X_n)/n$ is a so-called *consistent* estimator of μ. That is, for every $\varepsilon > 0$, the probability that $\bar{X}_n$ deviates by more than ε from μ vanishes if n becomes large. In other words, for all $\varepsilon > 0$ we have $\mathrm{prob}(|\bar{X}_n - \mu| > \varepsilon) \to 0$ if $n \to \infty$.

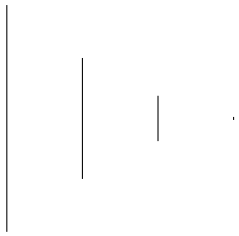

Figure 3: Convergence — diminishing uncertainty

Fig. 3 depicts this situation. The longer the lines, the more uncertainty there is. Sampling yields information, and thus the rightmost line indicates our final state of knowledge. Convergence means that the amount of information increases until, in the limit, it is a as large as possible (the deviation vanishes, i.e., the parameter is known). In the best-case scenario, the information growth is fast, monotonic and the final result is perfect knowledge.

In a nutshell, the crucial concept of convergence makes classical calculus and statistics work. It is only a slight exaggeration to say that without convergence, these and many other areas of mathematics would be at least incomplete, defective and perhaps almost non-existent. The basic problem of induction explains why: since the finite sample and the potentially infinite population are separated by a large gap, bridging the latter successfully is a major achievement, and that is exactly what concepts of convergence accomplish. In statistical terms, fig. 1 translates into a property of the random variable X (or a sequence of identical copies thereof) that is projectable to some limit:

$$\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5, \bar{x}_6, \ldots \to \mu$$

Figure 4: a convergent sequence

In order to establish an analog to fig. 2, i.e., to construct an elementary counterexample, it is straightforward to consider a situation where the above limit does not exist. To this end, suppose that $Z \sim N(0, 1)$, i.e., Z has a so-called Standard-Normal with density $\varphi(t) = e^{-t^2}/\sqrt{2\pi}$ for $t \in \mathbb{R}$ with expected value $\mu_Z = E(Z) = \int_{-\infty}^{\infty} te^{-t^2}/\sqrt{2\pi}\,\mathrm{dt} = 0$ and variance $\mathrm{Var}(Z) = E(Z^2) - \mu_Z^2 = 1$. Moreover, assume that the random variable Y is Standard-Cauchy, i.e. that it has the density $g(t) = 1/(\pi(1 + t^2))$, where $t \in \mathbb{R}$, and that Y and Z are independent.

Now consider a mixture of the above random variables in the sense that most observations come from the Normal. That is, suppose X_λ is given by (Hampel (1998), referring to J.W. Tukey, see fig. 5),

$(1 - 10^{-10})$ Standard-Normal + 10^{-10} Standard-Cauchy

More precisely, this means that the density f of X_λ is given by the convex combination $f(t) = (1 - \lambda)\varphi(t) + \lambda g(t)$, where $\lambda = 10^{-10}$. (In general, one has $0 \leq \lambda \leq 1$ for a convex combination.)

$$\bar{x}_1,\ \bar{x}_2,\ \bar{x}_3,\ \bar{x}_4,\ \bar{x}_5,\ \bar{x}_6, \ldots \quad \begin{matrix} \nearrow \\ \rightarrow \\ \searrow \end{matrix} \quad \begin{matrix} \text{`Normal'} \\ (1-\lambda) \cdot \text{Normal} + \lambda \cdot \text{Cauchy} \\ \text{`Cauchy'} \end{matrix}$$

Figure 5: Tukey's example

Although, for all practical purposes, X_λ behaves like a Normal, it is pointless to estimate $E(X_\lambda)$; the latter term does not exist, since $E(Y) = \infty$. In general, any arbitrarily small 'contamination' $\lambda > 0$ makes the estimation of $E(X_\lambda)$ impossible, since in all these cases $E(X_\lambda) = \infty$. In other words, in Tukey's example induction only works in the trivial case where $\lambda = 0$.

3 Alternating between success and failure

The integral $E(Y) = \int_{-\infty}^{\infty} t \cdot g(t)\,\mathrm{dt} = \int_{-\infty}^{\infty} t/(\pi(1+t^2))\,\mathrm{dt}$ diverges since the additional variable t in the numerator makes the integrand too large. Therefore, a straightforward remedy could consist in decreasing the numerator, which is for instance achieved with the help of the first logarithmic moment $E_I(Y) = \int_{-\infty}^{\infty}(-\ln|t|)/(\pi(1+t^2))\,\mathrm{dt}$. Since the latter value exists and vanishes, one can show that

$$E_I(X_\lambda) = (1-\lambda)E_I(Z) + \lambda E_I(Y) = (1-\lambda)(\gamma + \ln 2)/2,$$

where $\gamma \approx 0.57721$ is Euler's constant (Saint-Mont 2026). Thus the behavior of the first logarithmic moment is not pathological - quite the opposite: it just interpolates between $E_I(Z) = (\gamma + \ln 2)/2$ and $E_I(Y) = 0$.

The above calculations are based on the concrete functional forms of the Normal and the Cauchy. However, in order to get back to fig. 4, all that seems to matter is whether the expected value exists or not. As long as the expected value $E(X)$ of some random variable X exists, if there is a *single* point of attraction so to speak, the arithmetic mean of independent and identically distributed observations should approximate it. In other words, there should be a mathematical theorem that states this intuitively straightforward idea, and actually the first theorem of this kind, a 'law of large numbers', was proved by Bernoulli (1713).

Venturing further, notice that some parameter is just a number, derived from the underlying distribution. Put differently, the twin concepts of random variable and distribution are more fundamental than any number (a so-called parameter) associated with them. Thus a straightforward conjecture would be that if we collected data from a *fixed*

distribution, these values should point at its parent. More technically speaking, suppose $X_1, \ldots, X_n$ are independent and have the same distribution function $F(x)$. For every $x \in \mathbb{R}$, we may then count the number $n(x)$ of random variables for which $X_i \leq x$ holds. Thus we get a proportion $n(x)/n$, and altogether the empirical distribution function $\hat{F}_n(x) = \sum_{i=1}^{n} n(x)/n$. Given mild assumptions (e.g. that F is continuous), we expect that

$$\hat{F}_n(x) \to F(x)$$

in a strong mathematical sense. For obvious reasons this result is called the fundamental theorem of statistics (Glivenko (1933), Cantelli (1933)).

For instance, the theorem applies to Tukey's convex combination. The distribution function there is just $F(t) = (1-\lambda)\Phi(t) + \lambda G(t)$, where Φ and G are the distribution functions of Z and Y, respectively. That the sample's distribution function $\hat{F}_n$ converges toward F is completely natural, since the quota of observations that come from the Normal is $(1-\lambda)$ and that from the Cauchy is λ. No inconsistency occurs, rather, mathematicians have generalized the fundamental theorem in many different ways (see, e.g., van der Vaart and Wellner (1996)).

Considering the basic assumption of independent and *identically* distributed random variables, it is obvious that the latter assumption provides us with a reasonable limit (aka distribution function F). Of course, this assumption can be weakened, in particular upon assuming a series of functions converging towards F. However, that would only be a minor generalization not offering much insight.

The assumption of *independence* is much more interesting, since the basic idea is that information accrues with every observation. The more observations the better, but this does not guarantee in itself that we obtain enough information such that, for *every* $x \in \mathbb{R}$, the estimator $\hat{F}_n(x)$ converges toward $F(x)$. To this end, independence is a sufficient condition, but less could still do. Actually Etemadi (1981) proved that it suffices to consider *pairwise* independent random variables.

Of course, these convergence results are not the end of the story, and a Goodman-like counterexample is not far away. If Etemadi's basic assumption is violated, that is, if, for example, only the positive or the negative semi-axis is observed (for instance depending on the sign of the very first observation), we cannot learn anything about the other (unobserved) semi-axis. Therefore, the fundamental theorem of statistics does not hold given these circumstances. More generally speaking, independent observations provide a maximum amount of information about F, since they only depend on the parent distribution, whereas any kind of dependence amongst the random variables may slow the learning process. In the worst-case scenario, a single value x shows up indefinitely often, and we do not learn anything about the distribution of X.

4 Implications for Goodman's riddle

Theoretical statistics has to focus on numbers - and not their content or meaning. Thus, the number of observations is crucial, or more precisely, the amount of information these observations provide. The classical theorems of probability and statistics such as those just cited demonstrate that this approach works well: given mild assumptions, a finite sample can be generalized to a potentially infinite population. However, it is also rather

easy to construct counterexamples - one just has to violate a crucial assumption, that of independence in particular.

Goodman's rather artificial-looking concept of 'grue' is similar to Tukey's weird mixed distribution. Both do not really seem to be relevant in practice, nevertheless they pose intellectual riddles that should be addressed in order to clarify our understanding. Obviously, Tukey's strange distribution did not turn statistical estimation into a futile endeavor, and hopefully Goodman's riddle also does not lead to the conclusion that any kind of induction is impossible.

Considering green - why do we extend this property to all emeralds (seen or unseen) without any reservation? The reason is that we have learned much about physics and chemistry, i.e., the behavior of electromagnetic radiation and the structure of crystals, which means that we know how light interacts with transparent bodies such emeralds - in the end producing light of a certain wavelength. This is not just a psychological habit or a play with color terms (green, blue, grue, bleen, etc.) As long as the regularities remain as they are, emeralds will be green. In other words, our reason to believe that the next emerald will be green is extremely well-founded, since it is built on specific, well-explored *laws of nature*. There could hardly be a better reason for a sustained prognosis, and thus nobody in practice seriously checks on the color of emeralds.

Quite the opposite is true, however, with a spurious property such as being a third son (Goodman 1955). Even if by chance there are only third sons in a certain room at a certain time, the next man that enters the room could (and in all likelihood will) not be a third son. Thus the general idea that all men in a room have to be third sons vanishes even more quickly than the hypothesis that all platypuses live in zoos. In the terms of Swinburne (1968), 'locational' properties cannot be projected, whereas 'qualitative' properties, which are essentially independent of a particular setting, can be projected beyond the situation at hand. In the same vein, Carnap (1947) distinguishes between purely positional, purely qualitative and mixed properties.

In short, if the rules that govern a phenomenon are stable, or if there is some *prevailing structure*, we may extrapolate into the future. In the simplest case, 'being like' can be interpreted as the 'straight rule', which tells us that some phenomenon will repeat itself in the future. Thus, at times, it may be well-founded to add another zero to an long series of zeros, e.g. that the next emerald will be green or that the sun will rise tomorrow (fig. 1). However, and perhaps contrary to what Hume thought, we should not expect this to happen due to our habits, but because we have good theories about physical phenomena such as matter and the development of stars.

Therefore, by the same token, it is also often rational to assume that there will be some kind of change (fig. 2): neither should we expect the sun to shine forever, nor should we assume that life on planet Earth may exist for another ten billion years. Rather, because of our non-trivial scientific knowledge it is save to predict that the sun will turn into a red giant, and that life on the surface of planet Earth will cease to exist in just another billion years or so. The same train of thought informs us that it would be ridiculous to extend the property of 'being alive' endlessly into the future. Instead, we all know that - at least in the long run - we are all dead; and taking into account specific circumstances, the prognosis of one's residual lifetime may be considerably shorter. Actually, life expectancy at birth is a valuable idea, for it projects the present living conditions into an amount of time a newborn may expect to live.

5 A general model: the funnel

The above examples underline that induction is a matter of degree, and that inductive steps are gradual — given certain regularity conditions, they can be justified; otherwise they may fail. For these reasons, a funnel, indicating more or less information, is a reasonable model (see Saint-Mont (2022), and his fig. 2 in particular). With respect to making educated guesses about the future, all one has to do is to rotate the funnel, or fig. 3 specifically, thus arriving at:

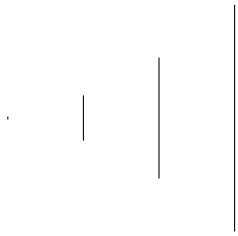

Figure 6: Divergence — growing uncertainty with respect to the future

The last figure should be interpreted as follows: We have a maximum amount of information about the present. (Again, a single point indicates perfect knowledge.) The further we venture into the future, the less we know, and thus the intervals become larger. How fast the errors multiply, i.e., how rapidly the depicted intervals grow is determined by our knowledge, which is based on the situation at hand.[1]

Following this train of thought, Knight (1921), p. 313, summarizes that "the existence of a problem in knowledge depends on the future *being different* from the past, while the possibility of a solution of the problem depends on the future *being like* the past" (my emphasis). Thus eclipses of the sun are easy to predict since the behavior of celestial objects is stable and also well understood. A weather forecast, on the other hand, is more difficult, since the physical laws lead to complex behavior, such that much data and powerful computers are necessary to produce a reliable (and precise) idea about tomorrow's weather. A prognosis becomes even more difficult if our knowledge and skills are meager (Tetlock and Gardner 2016). For example, due to countless and only partly understood influencing factors it is very difficult to predict a person's health or their life expectancy; and owing to the vague verbal theories which prevail in the social sciences, it is almost impossible to reach a consensus about next year's political developments, say.

Formally speaking, in the best case, the initial interval is small and the growth of successive intervals is slow, which means that we have a well-founded and precise educated guess, such as that all future emeralds will be green (even 100 million years from now). In the worst case, we cannot even predict what is going to happen in a few moments. ('Accidents will happen', as the proverb says.) Typical real-world examples are in-between: sometimes a sustained precise prognosis is possible, sometimes the error bars grow rapidly. Therefore, fluid dynamics - with laminar and turbulent flow in particular - seems to be an appropriate physical model for the philosophical problem of projectability.

[1]Instead of having the length of the lines represent error bars (lack of information), it may be convenient (and equivalent) that the lines' lengths represent the amount of information available, see Saint-Mont (2022), where much information corresponds to a long line.

More generally speaking, it is very natural to model problems of prediction and induction with funnel-like structures. These problems have a reasonable solution, if the corresponding funnel is not degenerate. In statistics, which is based on number(s), basic theorems assure that induction works if the observed information accrues toward some well-defined limit. In scientific settings, well-understood (general) laws, persistent structures and traceable mechanisms make predictions possible. By the same token, chaotic systems, quantum mechanics, and the theory of undecidability point at fundamental difficulties (Perales-Eceiza et al. 2025).

References

Bernoulli, J. (1713). Ars conjectandi.

Cantelli, F.P. (1933). Sulla determinazione empirica delle leggi di probabilità. Giorn. Ist. Ital. Attuari 4: 92–99.

Carnap, R. (1947). On the Application of Inductive Logic. Philosophy and Phenomenological Research. 8(1): 133–148.

Cohnitz, D.; and M. Rossberg (2024). Nelson Goodman. The Stanford Encyclopedia of Philosophy (Spring 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/spr2024/entries/goodman/>.

Etemadi, N. (1981). An elementary proof of the strong law of large numbers. Z. Wahrscheinlichkeitstheorie verw. Gebiete 55, 119–122.

Glivenko, V. (1933). Sulla determinazione empirica delle leggi di probabilità. Giorn. Ist. Ital. Attuari 4: 421–424.

Goodman, N. (1946). "A Query on Confirmation". The Journal of Philosophy 43(14): 383–385.

Goodman, N. (1955). Fact, Fiction, and Forecast. Cambridge, Massachusetts: Harvard Univ. Press.

Hampel, F. (1998). Is statistics too difficult? Canadian J. of Statistics, 26: 497–513.

Knight, F. (1921). Risk, Uncertainty, and Profit. *Houghton Mifflin, New York.*

Kovalenko, R. (2022). Kowalenko, R. (2022). The Putnam-Goodman-Kripke Paradox. Acta Analytica 37, 575–594.

Perales-Eceiza, Á., Cubitt, T., Gu, M., Pérez-García, D.; and M.M. Wolf (2025). Undecidability in physics: A review. Physics Reports 1138, 1-29.

Saint-Mont, U. (2022). Induction: A Logical Analysis. Found Sci 27, 455–487.

Saint-Mont, U. (2026). Information is Key. ISBN 979-8-249695-18-7

Swinburne, R.G. (1968). Grue. Analysis 28(4), 123-128.

Tetlock, P.E.; and D. Gardner (2016). Superforecasting: The Art and Science of Prediction. Broadway Books, New York.

van der Vaart, A.W.; and J.A. Wellner (1996). Glivenko-Cantelli Theorems. Springer.