The Ground of Prediction

On the Conditions Under Which Probabilistic Inference Is Valid, and Why Fluent Systems Obscure Them

Marcel Langjahr

Pantareon GmbH
https://pantareon.io/
langjahr@pantareon.io

Abstract

Probabilistic prediction — the inference of an unknown quantity from patterns in observed data — has become the dominant mode by which computational systems are asked to inform decisions, across domains from language to markets to public policy. This paper argues that the validity of such a prediction is not intrinsic to the method but borrowed from the process that generates the data: a probabilistic prediction is valid only to the degree that the quantity being predicted is genuinely governed by the regularities present in the data, and that the process producing those regularities is stable, densely sampled, and unperturbed by the prediction itself. Where these conditions hold — the deterministic limit, and the well-characterised stochastic middle — extrapolation is sound. Where they fail — in non-stationary, reflexive, or singular processes — the same machinery produces outputs formally indistinguishable from valid ones but without validity. We argue that contemporary large language models sharpen this problem in three ways: their output register is constant across the domains in which prediction is warranted and those in which it is not, rendering their uncertainty inaudible; their tendency to reproduce the centroid of their training distribution leads them to interpolate novel regimes as if they were familiar ones; and, for predictions about agents who act on them, they close a reflexive loop that corrupts the very quantity predicted. The paper is not a claim that probabilistic prediction is unsound, nor that such systems should not be used. It is an attempt to state the boundary between the domains in which such prediction carries and the domains in which it only appears to.

Keywords. probabilistic inference; prediction; stationarity; reflexivity; large language models; epistemic calibration; forecasting.

1 The Problem

1.1 The Moment

Prediction machinery is now deployed almost everywhere, and its validity is widely assumed to travel with the method rather than with the domain. Systems that estimate the next token, the next purchase, the next value of a series, the response of a population, are applied across domains that differ enormously in how predictable they actually are — and the systems do not distinguish among those domains in tone or in form. Competence in one domain is taken as licence to trust the same apparatus in another. This paper concerns the conditions under which that trust is warranted, and the specific way in which fluent systems make those conditions hard to see.

1.2 What This Paper Is Not

This paper does not argue that probabilistic prediction is invalid. It is the basis of a great deal of successful science and engineering, and nothing here is meant to unseat it. Nor does it argue that systems built on such prediction should not be used; used within their conditions, they are powerful. It is not a policy paper and advances no regulatory position. It is a position paper by an independent researcher, articulating a distinction that has proven necessary in the author's own work: between the domains in which a probabilistic prediction inherits validity from its data and the domains in which it does not, and the reasons the difference is presently obscured.

1.3 What This Paper Attempts

We attempt four things. First, to locate the source of a probabilistic prediction's validity in the process that generates its data, rather than in the estimation method. Second, to state the conditions under which a prediction inherits that validity. Third, to characterise the regimes in which those conditions fail while the outward form of the prediction does not change. Fourth, to show why fluent, pattern-based generative systems specifically obscure the difference between the two cases.

This paper is a companion to a separate framework on the structure of experience developed by the author. There, an epistemic point is made about the self-reports of such systems: that in their default operation their outputs track the centroid of their training distribution rather than any position, so that a report may describe what such a system would plausibly be expected to report rather than anything it underwent. That is the inward-facing counterpart of the outward-facing argument made here. One structural property of these systems — output as trained central tendency, expressed without a position — yields two distinct problems: an inward one for what such a system can be taken to report about itself, and an outward one for what it can be taken to predict about the world. This paper concerns the second.

2 Where Validity Comes From

2.1 A Prediction Borrows Its Ground

A probabilistic prediction is a statement about an unobserved quantity, licensed by regularities among observed quantities. The licence is not unconditional. It holds only insofar as the unobserved quantity stands in the same lawful relation to the observed ones that the observed ones stand in to each other. The prediction does not create this relation; it presupposes it. Where the presupposition holds, the prediction is as sound as the relation is lawful. Where it fails, the arithmetic of the prediction is unchanged but its warrant is gone.

2.2 The Deterministic Limit

Consider the simplest case. Given that $1 \times 2 = 2$ and $2 \times 2 = 4$, a prediction of the value of 2×3 is warranted — not because three examples make a large sample, but because the quantity predicted is governed, without residue, by the same rule that produced the examples. Here "prediction" is really derivation: the underlying structure is known and exceptionless, and extrapolation along it is certain. This is the limit at which probabilistic prediction is maximally valid — and, revealingly, the limit at which it is least needed, because where the law is known one does not estimate, one computes.

2.3 Probability as a Substitute for Structure

This reveals what probabilistic prediction actually is: a substitute for structural knowledge one does not possess. One estimates from patterns precisely when the generating law is unknown, or too complex to apply directly. The estimate is a wager that the unknown law is stable enough, and the sampling of its effects dense enough, that the shape of the past constrains the shape of the next instance. The wager is sometimes excellent and sometimes worthless, and — this is the crux — its quality depends on the process, not on the estimation technique. No sophistication of method can extract a valid prediction from a process whose future is not governed by its past.

2.4 The Validity Condition

We can state the condition under which a probabilistic prediction inherits validity from its data. Several properties of the generating process must hold, to a degree:

- *Stationarity.* The process that produced the data is the same process that will produce the target. If the regime has shifted, the data describes a world that no longer exists.

- *Dense sampling.* The region of possibility in which the target falls has been observed often enough that its regularities are actually present in the data, rather than absent or represented by a handful of unrepresentative instances.

- *Non-reflexivity.* The act of predicting does not alter the quantity predicted. If issuing the prediction changes the behaviour of what is predicted, the target moves as the prediction is made.

- *Repeatability.* The target is an instance of a recurring kind, not a singular event, so that "probability" has a referent at all. A one-off has no frequency.

Where these hold, probabilistic prediction is not merely legitimate but powerful — the actuarial table, the well-characterised physical measurement, the calibrated weather forecast all live here. Where they fail, the method continues to return an answer, formatted identically, and the answer is unfounded.

3 The Regimes Where the Condition Fails

3.1 Non-Stationarity

Many processes of interest change their own rules. Markets shift regime; institutions reorganise; a technology alters the very behaviour it was built to measure. A model fit to the prior regime then describes a vanished world with undiminished formal confidence. The economic record is dense with instances of models that predicted well until the regime that made them valid ended — and the density of the past data gives no warning, because the break is precisely what the past does not contain. Stationarity is the least examined of the conditions because its failure is invisible until after it has occurred.

3.2 Reflexivity

Where the target is an agent, or a population of agents, who can come to know the prediction, the prediction perturbs its object. A forecast of a market can move the market; a prediction about a person, conveyed to them, changes how they act; a projection of a policy's effect shapes the policy and thereby the effect. Prediction and target are no longer independent, and one can never afterward separate the accuracy of the forecast from its influence on what it forecast. This is the domain in which prediction is not merely uncertain but self-corrupting: it manufactures part of its own confirmation or refutation. Reflexivity is why prediction about human affairs is categorically harder than prediction about indifferent nature. Nature does not read the forecast.

3.3 Singularity

Some questions are not instances of a repeating kind. The outcome of a particular unprecedented event, the trajectory of a specific novel situation, the behaviour of a system in a regime it has never occupied — these have no frequency to estimate from. A probability offered for a singular, unprecedented outcome is a number without a denominator: it borrows the form of frequency-based inference while lacking the repeated trials that would give a frequency meaning. The estimate may still be produced, and will look no different from one resting on a million trials.

3.4 The Dissolving Target

There is a subtler failure in which the quantity predicted is not a fact about any case but an artefact of the aggregate — a central tendency that no individual instance embodies, and that dissolves the moment one acts on it as though it were a fact about the instance at hand. To predict "the average case" and then to meet a particular case is a routine way in which a formally valid aggregate prediction is invalid for the very instance to which it is applied. The average is real as a property of the population and absent as a property of the member.

In all four regimes the essential point is the same, and worth stating plainly: the failure is invisible in the output. A prediction drawn from a non-stationary, reflexive, singular, or aggregate-only process is formatted exactly like a prediction drawn from a stationary, independent, densely-sampled, repeatable one. The form of the answer is constant across the boundary of its own validity.

4 Why Fluent Systems Sharpen the Problem

Everything above holds of probabilistic prediction in general, by any estimator, human or mechanical. Contemporary large language models add three specific aggravations.

4.1 The Constant Register

A human forecaster who is guessing usually sounds like it — hedged, hesitant, visibly reaching. A statistical model at least emits a variance, a confidence interval, an error bar. A fluent generative system emits neither the human's audible hesitation nor the statistician's explicit variance. Its output in a domain where prediction is valid is linguistically indistinguishable from its output in a domain where it is baseless: the same measured, coherent, competent prose. The uncertainty is not represented in the channel the user actually reads. This is, for the present argument, the single most consequential property of such systems: they cannot make their own uncertainty audible.

4.2 Interpolation Over the Novel

Such systems are trained to reproduce the central tendencies of their corpus. Presented with a genuinely novel regime — one absent from, or sparse in, the training data — the system does not report absence; it interpolates, rendering the unfamiliar in the shape of the familiar. A regime break, precisely the case in which the past is the wrong guide, is therefore the case such a system is least equipped to flag, because flagging it would require recognising that the present lies outside the distribution the system has internalised — and its default disposition is to treat the incoming case as drawn from that distribution.

4.3 The Reflexive Loop at Scale

When such systems are used to predict the behaviour of the very agents who consult them — consumers, voters, participants in a market, an individual seeking guidance about another individual — the reflexive corruption of Section 3.2 operates at scale and at speed. The prediction is acted upon; the action moves the predicted quantity; the movement is read back as data; and accountability for the resulting decision diffuses into the system, which bears none. What was a wager on a stationary world becomes an intervention in a non-stationary one, mislabelled as an observation of it.

> *Question to hold open: if a system's confidence is constant across the domains in which it is warranted and those in which it is not, on whom does the burden fall to know the difference — and is that burden dischargeable by the very users least equipped, by situation or by expertise, to carry it?*

5 The Honest Counterweight

The argument must not prove too much. It would be false to conclude that probabilistic prediction requires certainty in its data, or that it is valid only in the deterministic limit. A great deal of uncertain, noisy data supports excellent prediction: mortality tables built on populations no member of which is individually predictable; weather forecasts, probabilistic and imperfect, that are nonetheless well-calibrated and genuinely informative; physical measurements carrying honest error bars. What these share is not certain data but a governed relation between data and target — stationary enough, densely enough sampled, non-reflexive, repeated — together with the discipline of calibration against outcomes, which continually tests whether the estimates track reality.

The claim of this paper is therefore not that prediction needs certainty. It is narrower, and we think more exact: prediction inherits validity from the structure of the generating process; some uncertain domains possess the required structure and some only appear to; and the appearance is identical from the inside. The task is to know which domain one is in — and it is precisely this discrimination that a fluent, confident, constant-register system removes from view.

There is a corresponding discipline, which we state without pretending it is easily institutionalised: match the confidence of a prediction to the knownness of the process that generates it, not to the fluency with which it can be expressed; make uncertainty audible where the channel would otherwise hide it; and hold the boundary between describing what is present in the data and forecasting what is not — the former being where such systems are strong, the latter being where their strength becomes hazard.

6 Where This Leaves Us

We have argued that the validity of a probabilistic prediction is borrowed, not intrinsic; that it is inherited from a generating process only when that process is stationary, densely sampled, non-reflexive, and repeatable; that four broad regimes violate these conditions while leaving the prediction's outward form unchanged; and that fluent generative systems aggravate the situation by holding their register constant across the boundary, interpolating novelty as familiarity, and closing reflexive loops at scale.

None of this is an argument against the use of such systems where their conditions are met, which is often and valuable. It is an argument against a specific and now-common error: mistaking the constancy of a system's confidence for the constancy of its validity. The multiplication that opens Section 2 is the emblem of the domain where prediction is sound — a known law, extended without residue. The domains where the stakes are highest — the shifting market, the reflexive population, the singular decision, the particular person — are precisely the ones least like multiplication; and a system that speaks of them in the same voice it uses for arithmetic is not thereby entitled to the same trust.

The distinction between description and prediction, between the governed and the ungoverned relation, between the audible and the inaudible uncertainty, is not one that current systems draw for us. Until they do, it must be drawn by those who use them — and the first step is to recognise that the boundary exists, and that nothing in the surface of a confident answer reveals on which side of it the answer falls.

> *Question to hold open: could a system be built whose expressed confidence varied with the knownness of the generating process rather than with the fluency of the expressible answer — and would such a system be adopted, given that its chief visible effect*

would be to say "I do not know" precisely where present systems are most reassuring?