Ambiguity and nonidentifiability in the statistical analysis of neural codes

Asohan Amarasingham, Stuart Geman, and Matthew T. Harrison

Among the most important open questions in neurophysiology are those regarding the nature of the code that neurons use to transmit information. Experimental approaches to such questions are challenging, because the spike outputs of a neuronal subpopulation, as typically recorded in behaving animals, are influenced by a vast array of factors. Such factors span all levels of description, from the microscopic (e.g., ion fluctuations and states of presynaptic neurons) to the macroscopic (e.g., sensation and attention), but only a small fraction of these is measured, or even understood. As a consequence, it is not clear to what degree variations in unknown and uncontrolled variables alternately reveal or confound the underlying signals that observed spikes are presumed to encode. Another consequence, very much related, is that these uncertainties also disturb our intuitive comfort with common models of statistical repeatability in neurophysiological signal analysis. In this context, there is an increasingly popular strain of thought in the neural-coding literature that “doubly stochastic point processes” (1–8) provide a way to think about and model fundamental questions about the relationship between sources of “trial-to-trial variability” (4, 9) and the observed variability of spike responses.

Imagine an experiment consisting of repeated presentations of a sensory stimulus. In a typical probabilistic model of spike responses, the probability that a particular neuron emits (fires) a spike at one time is described by a theoretical firing rate function (a function of time relative to stimulus onset). [Note that theoretical firing rates are distinct from the basic “observed” or “empirical” firing rate, which is a report of how many spikes occur in a window of a specified time length. Here, we discuss the theoretical firing rate. Observed firing rates are used to infer theoretical firing rates or their properties (SI Appendix, section S1)]. For example, in a generic firing rate model, spikes are treated as completely random, beyond the structure induced by the time-varying firing rate function. An analogy can be made with coin tossing: every neuron is associated with a sequence of weighted coins; each coin is associated with a moment in time, representing the absence or presence of a spike. Trials consist of tossing the coins. In a doubly stochastic spiking model, there is another level of complexity in that the firing rate function is itself random and varies from trial to trial (illustrated in SI Appendix, Fig. S1).

Thus, in such doubly stochastic models, there are “two layers of variability” (1). For each trial and neuron, a firing rate function (analogously, a sequence of coins) is randomly chosen. The variation in the firing rate function across trials is the first layer of variability. The firing rate function, in turn, generates the observed spike train (analogously, the coins are tossed), contributing the second layer of variability. Importantly, there is a mental affinity—never quite pinned down, perhaps, but nevertheless influential—between rate-coding and firing rate functions. In this view, the firing rate function is the “rate” of the “code.” For example, the affinity is sometimes exploited in experimental attempts to contrast rate codes from temporal codes, in which precise temporal structures or dependencies in spike patterns play a role in the neural code.

It is then natural to ask if we can use statistics to relate this probabilistic picture of two-layered variability to real, experimentally observed spike trains. Answers might then be related to questions about neural coding. In the words of Churchland and Abbott, “Experimentalists often attempt to segregate response variability into firing rate variability and spiking noise. It is generally assumed that the former can influence behavior or perception, where the latter effectively acts as measurement noise” (1).

This image is seductive (generative, easy to picture, and easy to simulate) but also potentially misleading, especially when we turn to inference. The root of the problem is that the definition of firing rate functions in the image that we have sketched—essentially, that a firing rate function specifies the probability of

Significance

Among the most important open questions in neurophysiology are those regarding the nature of the code that neurons use to transmit information. Experimental approaches to such questions are challenging because the spike outputs of a neuronal subpopulation are influenced by a vast array of factors, ranging from microscopic to macroscopic scales, but only a small fraction of these is measured. Inevitably, there is variability from trial to trial in the recorded data. We show that a prominent conceptual approach to modeling spike-train variability can be ill-posed, confusing the interpretation of results bearing on neural codes. We argue for more careful definitions and more explicit statements of physiological assumptions.

Author contributions: A.A., S.G., and M.T.H. conceived the ideas; A.A. and M.T.H. performed research; and A.A. and M.T.H. wrote the paper.

Reviewers included: M.A., Bar-Ilan University.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

1A.A. and M.T.H. contributed equally to this work.

2To whom correspondence may be addressed. Email: aamarasingham@ccny.cuny.edu or stuart_geman@brown.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1506400112/-/DCSupplemental.
a spike—is underconstrained, and therefore, firing rate functions are usually not unique. The six columns in Fig. 1 provide an illustration in which six strikingly different firing rate processes give rise to spike train processes which are, in fact, identical (i.e., statistically indistinguishable). This example is not singular. In fact, as exemplified by the sixth random firing rate process in Fig. 1, every (discrete time) spiking process can be perfectly described by a doubly stochastic model that attributes all variability to firing rate variability. Additional modeling assumptions would be needed to resolve these ambiguities. For example, we may require firing rates to be slowly varying in time, prohibit the firing rates from being too high at any instant, or invoke other kinds of statistical constraints on the firing rate process. Modeling assumptions of some kind are necessary to make reasonable use of the doubly stochastic concept.

Consider typical neurophysiological settings in which a theoretical firing rate is invoked. For example, do a pair of neurons exhibit more synchronous firing than one would expect from their (theoretical) firing rates? For another example, how much of a neuron’s observed spiking variability can be attributed to the variability of the (theoretical) firing rates, and how much can be attributed to spike timing variability? Theoretical firing rates specify the probability of a spike in a neuron. However, the probability of an event is not meaningful unless conditions of observation are specified (10–13). Rényi himself found it necessary to emphasize this: “Every probability is in reality a conditional probability.” (10). This observation might appear elementary—indeed, Rényi added: “[t]his evident fact is somewhat obscured by the practice of omitting the explicit statement of the conditions if it is clear under what conditions the probability of an event is considered” (10)—but in the neural-coding literature for various reasons such conditions may not be specified. This state of affairs causes at least two kinds of serious problems. In the first kind, different investigators both use the term firing rate, but in reality, they have different conceptions of firing rate in mind and are not discussing comparable quantities. In the second kind, an investigator assumes that the firing rate in a doubly stochastic model is unique or well-defined independent of a precise account of conditioning and then, using apparently ordinary statistical reasoning, draws conclusions about neural coding. This oversight can demonstrably lead to mathematical inconsistencies. A recurring problem stems from this basic fact that firing rate functions in doubly stochastic models may not be unique and, in any case, should be defined with more care.

Our aim here is to illustrate these ambiguities in the context of typical and recent neurophysiological questions. We choose these particular examples entirely for the sake of concreteness; the general issues are widespread in the literature of neural coding, if often unarticulated (14, 15). The manuscript is organized as follows. Our main point is developed in Fig. 1, where the essential ambiguity of inference is related to the statistical concept of identifiability. In Trial-to-Trial Variability and Doubly Stochastic Decompositions, we show how this ambiguity is present in, and dooms, prototypical model-free attempts to interpret the contribution of trial-to-trial firing rate variability to spike count and spike timing variability. In Implicit Modeling Assumptions and Solutions, we point out that closely related issues concerning modeling assumptions are pervasive in neurophysiology, and we provide a brief overview of potential solutions in this context. We summarize our conclusions in Summary and Discussion. Technical details are
provided in SI Appendix. Our chief conclusion is that the assumptions justifying statistical models should be prominently described and scrutinized when applied to the quantification of neural-coding questions, particularly in regards to trial-to-trial variability and related phenomena.

**Trial-to-Trial Variability and Doubly Stochastic Decompositions**

Here, we will first develop definitions and notation to facilitate a precise discussion of theoretical firing rates. The key property that a theoretical firing rate clearly must have is that the expected number of spikes per unit time is equal to the theoretical firing rate when the theoretical firing rate is known. Mathematically, we can write this as

$$E[\text{rate of spiking}|\text{firing rate}] = \text{firing rate},$$

where $E$ denotes expected value. (We will now typically use firing rate in place of theoretical firing rate.) More formally, if the term $N$ denotes the observed data (such as the number of spikes per time bin) and $\lambda$ denotes the corresponding (theoretical) firing rate, then we must have

$$E[N|\lambda] = \lambda.$$  \[1\]

This equation is read: “the expected value of $N$, given knowledge of $\lambda$, is itself $\lambda$.” Eq. 1 simply states that $\lambda$ is a firing rate for $N$.

For example, if we know that the firing rate in a 1-s interval is uniformly 10 Hz, so that $\lambda = 10$, then the expected number of spikes in that interval given that knowledge, $E[N|\lambda]$, should be 10.

The actual number of spikes may differ from 10 on a particular trial, but if we repeat the experiment enough, always with a 10-Hz firing rate, then the average number of spikes in the interval, averaged across the repetitions, will approach 10.

Allowing $N$ and $\lambda$ to further depend on time does not change this basic relationship. We simply have that $E[N(t)|\lambda(t)] = \lambda(t)$, where $t$ indexes time.

Complexities involving trial-to-trial variability, as described in the Introduction, often motivate modeling $\lambda$ as random, in which case the pair $(N, \lambda)$ becomes a doubly stochastic model: both $N$ and $\lambda$ are random variables (variables for which values are specified by the outcome of a random experiment). $\lambda$ itself is not observed. It is sometimes treated, for example, as the signal of interest (“the quantity that the neuron is supposed to represent”) (3).

When $\lambda$ is random, then it also has an expected value $E[\lambda]$ and variance $\text{Var}[\lambda]$ [the latter quantity denotes the variability in the (random) firing rate]. Furthermore, when $\lambda$ is random, quantities like $E[N|\lambda]$ and $\text{Var}[N|\lambda]$ become random as well, because they inherit the randomness in $\lambda$. Because $\text{Var}[N|\lambda]$ is a random variable, it also has an expected value, $E[\text{Var}[N|\lambda]]$, which quantifies the average conditional variability in spiking, given the firing rate. SI Appendix, section S2 reviews the notation regarding expectation $E[\cdot]$, variance $\text{Var}[\cdot]$, conditional expectation $E[\cdot|\cdot]$, and conditional variance $\text{Var}[\cdot|\cdot]$.

**Ambiguity of the Firing Rate.** There are many random variables that satisfy Eq. 1. This fact immediately dooms model-free attempts to estimate firing rates from experimental data. For example, given any random variable $R$ (or roughly, any piece of information), the random variable $\lambda$ defined by

$$\lambda = E[N|R]$$  \[2\]

will satisfy Eq. 1. (In fact, the following definition is equivalent to Eq. 1: $\lambda$ is a firing rate if $\lambda = E[N|R]$ for some $R$.) It is, thus, clear that, by considering different random variables $R$, a spectrum of different theoretical firing rates can be obtained in general. For example, in one traditional experimental formulation, $R$ is simply taken to specify time since the onset of an experimental trial. Alternatively, $R$ can draw from an essentially limitless combination of measured or unmeasured information sources. These sources range from relatively coarse, such as attributes of the experimental stimulus or the spiking activity of a simultaneously measured population of neurons (16–18), to finely-detailed, such as extracellular ion concentrations or somatic voltage. Consequently, there are many firing rates (Fig. 1). Furthermore, strikingly distinct firing rates can give rise to the same observed data. Thus, the observed data cannot be used to distinguish between candidate firing rate models. In statistical parlance, we say that such models are nonidentifiable (a formal definition is in SI Appendix, section S3): there is no correct or true version of the theoretical firing rate.

Essentially, the way to address this ambiguity is to introduce constraints on $\lambda$. However, not all constraints will work. For example, consider the spike time series $N(t)$ (a sequence of spike counts or a binary spike train) as a time series with the associated firing rate $\lambda(t)$ satisfying $E[N(t)|\lambda(t)] = \lambda(t)$. Then, $\lambda(t)$ is ambiguous in the sense discussed above (Fig. 1). An additional constraint, which is reminiscent of certain descriptions of rate coding, treats $N(t)$ as conditionally independent across time (8) given the time series $\lambda(t)$, but the constraint is also too weak. Processes 4–6 in Fig. 1 illustrate this point. Another way to look at this is as follows: consider a discrete spike train of length $T$ bins and discretize probability into $M > 2$ levels. A distribution on $\lambda(1), \lambda(2), \ldots, \lambda(T)$ is specified by $M^T - 1$ parameters. Moreover, the distribution on $\lambda(1), \lambda(2), \ldots, \lambda(T)$ (continuously) determines the distribution on $N(1), N(2), \ldots, N(T)$, which is characterized by $2^T - 1$ parameters. This observation alone implies nonidentifiability and suggests its perservativeness.

We treat more subtle characteristics of such ambiguities in detail below.

**Spike Count Variability.** Consider the following interesting statistics problem (related, in varying degrees, to refs. 3, 7, and 19–21). Fix a recording interval (of, for example, 100 ms, with some start time specified relative to a stimulus onset), according to which we record multiple observations of spike counts, $N_1, N_2, \ldots, N_m$, from a neuron corresponding to $m$ trials of a repeated experiment. For each of the $m$ trials, there is a separate firing rate (i.e., $\lambda_i$ is the firing rate on trial $i$). Trial-to-trial variability is here modeled by the fact that $\lambda_1, \lambda_2, \ldots, \lambda_m$ are not necessarily identical. A typical neural-coding intuition is that of rate coding, in which $\lambda_i$ represents an “intensity command” or “the quantity that the neuron is supposed to represent” in trial $i$ (3). In practice, one only observes $N_1, N_2, \ldots, N_m$, and it is variable. How much of the observed variability is because of variability in the “intensity command” (3) (firing rate)?

To formalize this question, suppose that $N$ and $\lambda$ are the spike count and firing rate, respectively, in a randomly chosen or future trial. The variability in $N$ can be decomposed as follows (2, 3, 7, 19):

$$\text{Var}[N] = \text{Var}[E[N|\lambda]] + E[\text{Var}[N|\lambda]]$$  \[3\]

where Eq. 3 is a standard decomposition valid for any pair of finite random variables, and Eq. 4 follows from the definition that $\lambda$ is a firing rate for $N$ (Eq. 1). It is natural to ask how we can use observations of $N$ to distinguish the two variability terms $\text{Var}[\lambda]$ (“firing rate variability”) and $E[\text{Var}[N|\lambda]]$ (“point process variability”) (3, 7, 19).

To consider this last question, suppose now that we return to our observations of $N_1, \ldots, N_m$ in trials and use these to estimate $\text{Var}[N]$ in the usual manner. In practice, some approximations are made to also estimate $E[\text{Var}[N|\lambda]]$ from $N_1, \ldots, N_m$. This estimate can then be used (3) to infer $\text{Var}[\lambda]$ using Eq. 4. Even without examining those approximations directly, it is clarifying to take a broader perspective on the statistical model underlying this decomposition. The model is fairly minimal: $N_1, \ldots, N_m$ and
Case I. \(\lambda_1, \lambda_2, \ldots, \lambda_m\) are independent and identically distributed (iid) Poisson random variables with parameter \(c\). \(N_i = \lambda_i\). Thus, \(N_1, N_2, \ldots, N_m\) are iid Poisson(c). Under this generative model, \(\text{Var}[\lambda] = c\) and \(E[\text{Var}[\lambda]] = 0\).

Case II. \(\lambda_1 = \lambda_2 = \ldots = \lambda_m = c\). (The \(\lambda_i\) values are a constant.) Each \(N_i\) is an independent Poisson random variable with parameter \(\lambda_i\). Thus, as in case I, \(N_1, N_2, \ldots, N_m\) are iid Poisson(c). Therefore, under this generative model, \(\text{Var}[\lambda] = 0\) and \(E[\text{Var}[\lambda]] = c\).

In both cases I and II, \(N_1, \ldots, N_m\) has exactly the same spiking distribution. Thus, the underlying variance decomposition is nonidentifiable: no use of the data (i.e., \(N_1, \ldots, N_m\)) could possibly distinguish \(\text{Var}[\lambda]\) from \(E[\text{Var}[\lambda]]\), regardless of how much data we collect. [Mixing cases I and II produces intermediate models that are equally indistinguishable. In fact, our use of a Poisson example is just a pedagogical choice; the nonidentifiability phenomenon is completely general (SI Appendix, section S4).] The problem could only be worsened if we introduce sampling variability, such as when the (probability) distribution of \(N\) is inferred from the data.

Additional constraints are needed. For example, refs. 3 and 4 invoke another modeling assumption that \(\text{Var}[N|\lambda] = \phi \lambda\), where \(\phi\) is an unknown constant parameter that does not vary with \(\lambda\) or time. Ref. 3 includes an extensive discussion of this assumption as well as its motivation from the model of rate-modulated renewal processes, in which spike trains are the outcomes of a renewal process that has been rescaled in time (SI Appendix, section S5). However, this strong assumption is not strong enough to ensure identifiability, which the counterexamples show, because under this model, \(\phi\) is itself not identifiable (despite the fact that it is crucial to conclusions drawn about \(\text{Var}[\lambda]\)) (5). [Note, for example, that \(\phi\) is zero and one, respectively, in cases I and II above (additional discussion of this nonidentifiability of \(\phi\) and the nonidentifiability of the constancy of \(\text{Var}[N|\lambda] = \lambda\) is in SI Appendix, section S6).] Indeed, the rate-modulated renewal process can generate any discrete time spike process, even when \(\phi\) is restricted to zero (SI Appendix, section S5). This observation alone implies the nonidentifiability of \(\phi\) and theoretical firing rates in general. The supplemental information in ref. 3 further describes how different assumptions about the numerical value of \(\phi\) can influence scientific results.

However, if \(\phi\) is known and this assumption holds a priori, then \(\text{Var}[\lambda]\) is identifiable, and it is straightforward to estimate from the spike counts (a mathematical explanation is in SI Appendix, section S7). In terms of scientific considerations, the modeling assumption and a priori knowledge of \(\phi\) are the key assumptions to be debated.

Indeed, if \(\phi\) is not known or \(\text{Var}[N|\lambda] / \lambda\) is not a constant, \(\text{Var}[\lambda]\) and \(E[\text{Var}[\lambda]]\) are entangled, and their distinction does not even quite make sense.

**Spike Train Variability and Cross-Correlogram Statistics.** Another prominent case is synchrony analysis and other analyses based on the cross-correlogram. Is it possible to disambiguate the relative contributions of firing rate variability and spike timing variability to the observed synchrony between a pair of neurons? This question arises naturally from experimental attempts to distinguish rate-coding from more complex temporal-coding hypotheses.

Consider the following generative prescription for a doubly stochastic spike train model. First, in each trial, a firing rate function is randomly chosen. As above, the randomness of the firing rate function is a mechanism for modeling complexities of trial-to-trial variability. Second, spikes are randomly generated, with probabilities governed, in an appropriate sense, by the associated firing rate function. Then, we can ask (2, 3, 7) whether the variability of the firing rate functions can be separated from the variability in spike timing, given the firing rate. If so, the goal is to quantify the contribution of each component of variability to the observed synchrony (or time-lagged synchrony).

Suppose that we model time as discrete so that we are dealing with (otherwise arbitrary) binary time series. Let \(N\) and \(\lambda\) refer to the observed spiking and firing rate, respectively, of one neuron in a (e.g., random) time bin, and let \(M\) and \(\gamma\) refer to the observed spiking and firing rate of a second neuron, possibly lagged by a fixed amount \(\tau\). Note that \(N\) and \(M\) are now binary random variables signifying the presence or absence of a spike in a time bin, in neurons one and two, respectively. It is also possible to include time directly in the equations so that these quantities are related to an experimental clock, but our conclusion—nonidentifiability—will be the same (SI Appendix, section S8).] We can study the covariability between \(N\) and \(M\) using the covariance \(\text{Cov}[N, M]\), which measures the covariability in firing for two neurons (2, 3) (SI Appendix, section S2).

Firing rates \(\lambda\) and \(\gamma\) are hidden ("latent") random variables that have the property that knowledge of the firing rate function specifies the likelihood of a spike:

\[
E[N|\lambda, \gamma] = E[N|\lambda] = \lambda \\
E[M|\lambda, \gamma] = E[M|\gamma] = \gamma.
\]

Verbally, (i) given knowledge of a neuron’s firing rate function, the knowledge of other neurons’ firing rate functions is irrelevant to the (conditional) probability of a spike, and (ii) the (conditional) probability of a spike in neuron 1 or 2 at a randomly selected moment, given knowledge of \(\lambda\) or \(\gamma\), respectively, is \(\lambda\) or \(\gamma\), respectively.

A covariance decomposition can then be applied:

\[
\text{Cov}[N, M] = \text{Cov}[E[N|\lambda, \gamma], E[M|\lambda, \gamma]] + E[\text{Cov}[N, M|\lambda, \gamma]]
\]

where Eq. 6 is the analogous covariance decomposition valid for any pair of finite random variables, and Eq. 7 follows from definitions in Eq. 5. The reasoning, now based on covariance, is the same as in the previous section (compare Eq. 7 with Eq. 4).

For example, Staude et al. (2) conclude: “The decomposition achieved [i.e., Eq. 7] shows that spike coordination and rate covariation are mathematically distinguishable components of the cross-correlation function.”

In other words, the setup is just as in the previous section, except that, now, \(N\) and \(M\) indicate the event of a spike in a time bin for each neuron, and \(\lambda\) and \(\gamma\) are the associated (random) probabilities of a spike in the bin. By the same logic, this decomposition suggests the strategy (2) of quantifying the "spike coordination" component of variability by subtracting an estimate of "firing rate covariability" from an estimate of \(\text{Cov}[N, M]\).

However, once again and for the same reasons, the fundamental difficulty is that \((\lambda, \gamma)\) is not well-defined in terms of the random variables \(N\) and \(M\). In particular, for any pair of (discrete-time) random spike trains, there is always an interpretation of firing rate variability available that is (i) completely consistent with the definitions in Eq. 5, but (ii) under which, there is no spike coordination in the sense of Eq. 7. More generally, any population of (discrete time) random spike trains can be represented as a doubly stochastic model, in which spikes are conditionally independent, given their firing rate functions. (This fact is the implication of the example of
Column 6 in Fig. 1 extended to multiple neurons; SI Appendix, section S8 has an explicit demonstration and expanded discussion.) Thus, the decomposition expressed in Eq. 7 is nonidentifiable, because firing rates are not unique.

**Scope of Critique.** Our intent is not to diminish the importance of the questions and phenomena discussed in the papers that we happened to choose for the discussion above. These (and related) works rightly serve to motivate the importance of developing proper interpretations of trial-to-trial variability in neuroscience. Such issues are indeed routinely tackled in leading laboratories. Rather, we want to underscore the complexities involved in analyzing the variability of spike trains. We feel that these complexities are intrinsic to the interpretation of electrophysiological measurements, but also that their articulation in quantitative and statistical terms is hampered by an imprecision that we have also come across more broadly in the literature.

**Implicit Modeling Assumptions.** We concluded above that, under these doubly stochastic spike models, several parameters of apparent interest are not identifiable unless stronger modeling assumptions are made. In their absence, such estimators cannot have well-defined targets. What then is being estimated in practice by procedures that evoke the notion of trial-varying firing rates?

Many procedures for estimating firing rate functions typically involve some form of smoothing. Smoothing effectively amounts to another implicit modeling assumption, built into the definition of firing rate covariation. The model is that firing rates specify the probability of a spike as well as additional constraints, such as that firing rates vary at a particular (slower) timescale, as governed by details of the smoothing procedure. This interpretation extends to other approaches (22). That is, the typical situation is that statistically modeling assumptions are not completely explicit but implicitly built in to an estimation or inference procedure (23). In addition to the well-known dangers that implicit assumptions lead too easily to misinterpretation, another important concern, particularly with sophisticated smoothing techniques (e.g., cross-validation, penalized likelihood, and Bayesian models), is that the degree of smoothing may depend on the quantity and quality of the data, and therefore, the same procedure is essentially using different modeling assumptions for different datasets. Another difficulty is that estimation of quickly varying and trial-varying firing rates will be more error-prone (or even hopeless), because there are few data for estimating each probability. Approximations, like the bootstrap, can break down in this regime. For example, Staude et al. (2) observe this phenomenon in simulation but misdiagnose it as the main obstruction to effective firing rate estimation in practice. Rather, it seems to us that the main obstruction is that firing rates are subjective and not identifiable without additional modeling assumptions. An infinite amount of data would not salvage the situation.

Several works (24, 25) [including our own (26)] have suggested that one of the fundamental distinctions in neural-coding problems is that of timescale. In line with this observation, many procedures in the literature, even if they purport to avoid modeling assumptions, are often tested or developed in simulations in which the timescales of interest are clearly distinguished. For example, Staude et al. (2) contrast two classes of generative models in their simulation studies. In the first class, processes are built as doubly stochastic models with firing rate functions that are random and piecewise constant (constant in intervals of fixed temporal length) and with spikes otherwise independent, conditioned on those firing rate functions. The temporal lengths set a timescale of firing rate variability for the process. In addition, the firing rate functions of different neurons may be correlated in the sense that the constants are correlated. This model provides the archetype for firing rate variability. In the second class, spike coordination is built on top of the former model. Here, coincident spikes are injected jointly into the pair of neurons modulo some noisy scattering. The width of the scattering sets the timescale of spike coordination. This simulation model is presumably the motivation for the statistics problem of isolating the contribution of the spike coordination component, given only the observed spike times. In fact, this contribution is not identifiable unless the two timescales are distinct. Because the key distinction in such a simulation model is, thus, one of timescale, it makes sense to organize inference explicitly around this distinction as well.

**Solutions**

Explicit Modeling Assumptions. Nonidentifiability effectively means that many of the questions that we have discussed are ill-posed as stated. Progress would, thus, seem to favor the introduction of constraints that are scientifically relevant and also explicit.

As an example, a common way to incorporate constraints is to model spike times on the continuum and then assume both that spike counts in nonoverlapping temporal intervals are independent and that trials are independent and identically distributed. This model is the Poisson process. However, doubts about such a simple form of constraint are largely responsible for the more complex models that motivated our discussions here. A relaxation of the Poisson assumption is to assume that the spike process is conditionally a Poisson process, conditioned on a random rate function, again with independent trials (4, 7). Such a doubly stochastic model is called a Cox process (27). As an example, suppose that we assume that the spike process is Cox and that the random variable \(N\) is the spike count in an interval, and we identify the firing rate with the random rate function of the Cox process, \(\lambda\). In this case, the firing rate is identifiable (28). Furthermore, in this case, \(\phi = \text{Var}[N]/\lambda = 1\) by definition, and therefore, \(\text{Var}[N]\) can be estimated as in SI Appendix, section S7. (However, see also the concerns in refs. 5 and 29 as well as the discussion and warnings in SI Appendix, section S9 on identifiability and the Cox model.) Refer to ref. 23 for a broader discussion of widely used constraints.

**Timescale.** As remarked above, distinctions concerning timescale are central to questions about the neural code. One effective way to handle such distinctions is to model questions about a spike’s timing in terms of its conditional probability, conditioned on a temporal coarsening of observed spike trains (26). For example, let \(X\) represent a pair of spike trains. Let \(C_\Delta(X)\) be a temporal coarsening of \(X\): divide time into disjoint windows of length \(\Delta\), and encode the spike counts in those windows for the two spike trains as \(C_\Delta(X)\). We conceptualize \(C_\Delta(X)\) as discarding spike timing information on timescales finer than \(\Delta\). The idea is to focus statistical questions about time-timescale interactions in \(X\) onto the conditional distribution:

\[
\Pr(X|C_\Delta(X)).
\]  
[8]

Degrees of temporal coarsening are specified by \(\Delta\) and provide a rigorous basis for defining timescales quantitatively. (The idea generalizes in many directions, including to an arbitrary collection of spike trains.) A consequence of this definition is that the estimation of coarse-timescale parameters is unnecessary for inference of fine-timescale parameters. In this way, nonidentifiability and other obstructions to estimation of firing rate, defined on coarse timescales, become irrelevant. This modeling approach leads to permutation-based and other inference procedures that are, in some cases, exact (i.e., approximation-free) and thus, sidestep nonidentifiability. Variations on the jitter principle are one example (8, 26, 30–32) that have been applied to a variety of neurophysiological problems (26, 33–38). Thus, conditional modeling provides an example of a practical and explicit way to model trial-to-trial variability robustly when the scientific goal is to separate neuronal interactions across timescales.

**Summary and Discussion**

In other words, what is a firing rate? We argue that this question, often a centerpiece in systems neurophysiology, deserves a more thorough and ongoing interrogation in the literature. Theoretical firing rates are subjective quantities. We have discussed...
some of the problems that arise when insufficient attention is devoted to the question. These problems include inconsistent statistical reasoning, an incomplete understanding of quantitative conclusions regarding trial-to-trial variability, and even nonidentifiability.

Most likely, the source of these problems is that we are collectively misled by intuitions that are overgeneralized from simple generative models. For example, consider again the example of defining spike coordination. Firing rate functions have a natural definition for inhomogeneous Poisson processes. In that setup, it is natural to define spike coordination in terms of violations of independence beyond the firing rates. However, intuitions about trial-to-trial variability reasonably challenge Poisson-like assumptions (f), leading one to further parameterize the firing rate function with trial-varying parameters. Then, one follows the same approach of associating departures from independence with spike coordination. However, the analysis is delicate, because the probabilistic structure that we invoke to model trial-to-trial variability will have significant implications for what we mean by spike coordination. We cannot avoid such modeling assumptions, although they may be implicit (see Implicit Modeling Assumptions). This fact is most exposed by the extreme observation that, if we simply generalize the Poisson-like assumptions to accommodate arbitrary trial-to-trial variability, then spike coordination and trial-to-trial variability, in such a model, are conflated.

Ambiguous language might also bear some of this blame: many things are called a firing rate in the neurophysiology literature (23). For context, recall the spike count measurements discussed in Trial-to-Trial Variability and Doubly Stochastic Decompositions, Ambiguity of the Firing Rate, where \( \lambda_i \), \( N_i \) forms a doubly stochastic model producing spike counts \( N_i, N_2, \ldots, N_m \) across \( m \) trials. Consider that (i) the firing rate of the process is \( E[N_i] \) vs. (ii) the firing rate is \( \lambda_i = \lambda_{i,n} \). Or, accommodating trial-to-trial variability, consider that the firing rate is defined individually on each trial: (i) the firing rate on trial \( i \) is \( N_i \) vs. (ii) the firing rate on trial \( i \) is \( E[N_i | \lambda_i] \) vs. (iii) the firing rate on trial \( i \) is \( E[N_i] \). All four definitions of firing rate are well-defined in any of the generative models discussed in Trial-to-Trial Variability and Doubly Stochastic Decompositions, Ambiguity of the Firing Rate and have relationships to the different ways in which the term firing rate is used in the literature. Some correspond to random variables, and some correspond to parameters. However, they are completely different. Perhaps different notions of firing rate are vague or inadvertently interchanged, or simply misinterpreted.

What is more, perhaps, even the relationship between firing rate functions and firing rate codes, although arguably a principal justification for firing rate analysis in the literature, is a cloudy one (5, 39–41). (The neural-coding schemes proposed in refs. 42–44, for instance, are counterexamples to a traditional view of the relationship between a conventional firing rate code and noise in neural computation.) We suspect that such analyses are often motivated by buried but arguably underjustified assumptions regarding exactly how neurons are representing information to downstream targets.

In terms of neural coding, another way to look at this is as follows. There is no a priori separation between a signal and noise. Absent a model, the difference is semantic. It follows that variability decompositions cannot solely be based on mathematical, as opposed to neuroscientific, distinctions.

ACKNOWLEDGMENTS. We thank B. Anderson, D. Butts, G. Buzsáki, P. Cloud, H. Ejaab, S. Fujisawa, A. Grosmark, J. Levitt, R. Kass, S. Kelly, W. Levy, G. Maimon, M. Paradiso, A. Renart, J. de la Rocha, E. Schomburg, L. Sjulson, W. Truccolo, and B. Watson, as well as M. Abeles and an anonymous referee for helpful comments on earlier versions of this manuscript. We thank Chip Levy for encouraging our collaboration. Research supported in this publication was, in part, on work supported by NIH Grant R01MH102840 and NIH Institutional Development Award P20GM103645, a City College of New York City Seed Grant, National Science Foundation Grants DMS-1007593 and DMS-1309004, Office of Naval Research Contract ONR N000141010923, and Defense Advanced Research Projects Agency Contract FA8650-11-1-7151.
**Ambiguity and non-identifiability in the statistical analysis of neural codes**

Supporting Information

Asohan Amarasingham  Stuart Geman  Matthew T. Harrison

---

**Supplementary Figure 1**: Point processes and doubly stochastic point processes. The numerical values of the firing rates (in units of spikes/bin) at each instant in time are shown in boxes. Given these firing rates, spikes (arrows) are independent. (In this example, there can be at most one spike per bin and firing rates correspond directly to probabilities of spiking.) The left column shows repeated trials of a classical spiking model where the firing rate function is identical across trials. Variability in observed spiking is modeled as only arising from spiking noise. The right column shows repeated trials of a doubly stochastic spiking model where the firing rate function varies randomly across trials. Variability in observed spiking is modeled as arising partly from spiking noise and partly from variation in firing rates. The generative model in both cases is illustrated in the bottom box: first choose the firing rate — fixed (left panel) or random (right panel) — and then randomly generate a spike given the chosen firing rate.

**S1. Theoretical firing rates are distinct from observed firing rates.**

A theoretical firing rate is distinct from the basic "observed" or "empirical" firing rate, which is a report of how many spikes occur in a window of a specified timelength. In this paper, we discuss the theoretical firing rate, which we will sometimes refer to simply as ‘firing rate.’ Observed firing rates are used to infer theoretical firing rates, or their properties. The analogy with elementary statistics is that the observed firing rate corresponds to a sample mean (a random variable), whereas the theoretical firing rate corresponds to a population mean (a parameter). In the specification of doubly stochastic models in particular, however, there are several kinds of “means”, which adds a layer of subtlety to the issue and multiplies the confusion absent a more precisely-specified model.

**S2. Some elementary probability concepts.**

A random variable is a variable the value of which is specified by the outcome of a random experiment. A random (or stochastic) process, or simply process, is a collection of random variables indexed by time. For a discrete-valued random variable $N$, $\Pr(N = n)$ denotes the probability that $N$ takes on the value $n$. $E[N]$ and $\text{Var}[N]$ denote the expectation and variance, respectively, of random variable $N$. $E[N|\lambda]$ and $\text{Var}[N|\lambda]$ are the conditional expectation and conditional variance, respectively, of random variable $N$ given knowledge of random variable $\lambda$. For example,
E[N|\lambda] can be read: “The conditional expectation of \( N \), given knowledge of \( \lambda \).” Note that, when \( \lambda \) is random, \( E[N|\lambda] \) is itself a random variable, because it is a function of \( \lambda \). One useful interpretation of \( E[N|\lambda] \) is as the unique “best guess” for \( N \) based on knowledge of \( \lambda \) – where “best” is in the sense of minimal average square-error. Since \( E[N|\lambda] \) is a random variable, it has a variance, denoted \( \text{Var}[E[N|\lambda]] \). Similarly, the covariance between two random variables \( N \) and \( M \), written \( \text{Cov}[N, M] \) is defined as \( \text{Cov}[N, M] = E[(N - E[N])(M - E[M])] = E[NM] - E[N]E[M] \), and measures how variation in one of the random variables is predictive in a linear sense of variation in the other random variable. Independent random variables have zero covariance. Conditional covariance is defined in the analogous way.

**Expectation and Probability are equivalent when outcomes are binary.** In the case that \( N \) is a binary \( \{0, 1\} \) variable (indicating spike or no spike), note that \( \text{Pr}(N = 1) = E[N] \) so probabilities and expectations are equivalent in this case, as are the associated conditional probabilities and conditional expectations. (This equivalence is used several times in the text.)

See [1], for example, for more information.

**S3. Formal definition of non-identifiability.**

Here is a technical definition of non-identifiability for discrete data sets [2]. Model a data set \( X \) as a random variable, with \( \Theta \) a space of parameters (of any dimension). This means the probability of outcome \( x \) is \( \text{Pr}_\theta(X = x) \), with \( \theta \in \Theta \). In inference, \( \theta \) is conceptualized as unknown. \( \theta \) is not identifiable if there is a \( \theta_1 \in \Theta \) and a distinct \( \theta_2 \in \Theta \) with the property that \( \text{Pr}_{\theta_1}(X = x) = \text{Pr}_{\theta_2}(X = x) \) for all possible outcomes \( x \). (Note the implications are the same regardless of Bayesian or frequentist setting.)

**S4. Non-identifiability of the \( (\text{Var}[\lambda], E[\text{Var}[N|\lambda]]) \) decomposition.**

In the text, our use of a Poisson example is just a matter of choice, as the same example generalizes to any arbitrary distribution on the spike counts. Suppose the distribution for \( N \) is specified by \( (p_0, p_1, p_2, \ldots) \) in the sense that the probability mass function (pmf) for \( N \) is \( \text{Pr}(N = k) = p_k \).

**Case 1.** Let \( \lambda \) be a sample from pmf \( \text{Pr}(\lambda = k) = p_k \), and set \( N = \lambda \). In this case the doubly stochastic spike count model \((N, \lambda)\) has \( E[N|\lambda] = \lambda \) with \( \text{Var}[E[N|\lambda]] = \text{Var}[N] \) and \( E[\text{Var}[N|\lambda]] = 0 \).

**Case 2.** Let \( \lambda = \sum_{k=0}^{\infty} kp_k < \infty \) and \( N \) be a sample from the pmf. Then \((N, \lambda)\) also has \( E[N|\lambda] = \lambda \) but \( \text{Var}[E[N|\lambda]] = 0 \) and \( E[\text{Var}[N|\lambda]] = \text{Var}[N] \).

In fact, there will typically be many such doubly stochastic models, lying along a spectrum, which preserve the pmf \((p_k)\) for \( N \), while \( \text{Var}[E[N|\lambda]] \) spans 0 to \( \text{Var}[N] \).

**S5. Any collection of discrete-time random spiking processes can be represented as a collection of rate-modulated renewal processes with random rate functions.**

The reasoning is very similar to S8.

A renewal process \( X_1, X_2, \ldots \) is defined by construction from a collection of nonnegative iid random variables \( I_1, I_2, \ldots \) via the relations \( X_0 = 0 \) and \( X_j = I_1 + X_{j-1} \). (Sometimes \( X_0 \) is constructed in another way but this does not change any conclusions.) In our examples, \( X_j \) is the time of the \( j \)th spike. The distribution of \( I_j \) is called the ISI (“inter-spike interval distribution.”) We will define \( \mu = E[I_1] \) and \( \sigma^2 = \text{Var}[I_1] \). A rate-modulated renewal process \( S_1, S_2, \ldots \) is a time-rescaled version of \( X_1, X_2, \ldots \), defined by \( S_i = \Lambda^{-1}(X_i) \) where \( \Lambda \) is any invertible monotonically-increasing function. Here we assume \( \mu = 1 \). (This does not constrain the class of processes \( S_1, S_2, \ldots \) that are considered.) In this construction, \( \lambda(t) = \frac{d\Lambda}{dt}(t) \), when appropriate, is referred to as an intensity function. (The representation through the intensity function is commonly used. It specifies firing rates in various senses. To connect this with the main text, for example, fix a time interval \([a, b]\) and define the random variable \( N \) as the spike count in \([a, b]\). Then \( N = \lambda(b) - \lambda(a) = \int_a^b \lambda(t) dt \) is a firing rate in the main text’s sense: \( E[N|\lambda] = E[N] = \lambda \).)

Let us generalize rate-modulated renewal processes to rate-modulated renewal processes with random rate functions. The main point is that every discrete-time spike time process is included in this generalization: a random time-rescaling function can be used to model any spiking process. To see this (see also S8) we will ignore continuum pathologies and assume time is discretized (suitably finely), and let \( \omega \) represent a sample outcome so that all probabilities are specified by \( P(\omega) \). In this case, the rescaling function depends on \( \omega \) – we will denote it \( \Lambda^{-1}(t, \omega) \). Consider an arbitrary random spike train and let \( S_j(\omega) \) specify the time of the \( j \)th spike in the train. Fix a sample outcome \( \omega \).
Using the renewal process $X_1 = 1, X_2 = 2, \ldots$ constructed from the (deterministic) ISI distribution $I_1 = 1$, consider any $\Lambda^{-1}(t, \omega)$ that satisfies

$$\Lambda^{-1}(j, \omega) = S_j(\omega) \quad \forall j \in \{1, 2, \ldots\} \forall \omega.$$ 

Thus, trivially, we have $\Lambda^{-1}(X_j, \omega) = S_j(\omega)$. Since the process $S_1, S_2, \ldots$ was chosen arbitrarily, this shows that random rate-modulated renewal processes can model any spiking process. In this construction, $\phi = \text{Var}[N|\lambda]/\lambda = 0$. Since, in general, randomly-modulated renewal processes can also be constructed without taking $\phi = 0$, the non-identifiability of $\phi$ and of the distribution of the random rate process follows.

Returning to the main text (Trial-to-Trial Variability and Doubly Stochastic Models, Spike Count Variability), one motivation for using $\text{Var}[N|\lambda]/\text{E}[N|\lambda] = \text{Var}[N|\lambda]/\lambda$ comes from the theory of renewal processes. The approximate logic appears to be as follows. Start with an ordinary renewal process $X_1, X_2, \ldots$ and let $N$ be the corresponding spike count in a fixed observation window. Provided $\sigma^2 > 0$ and the window is sufficiently wide (i.e., asymptotically), then $\text{Var}[N]/\text{E}[N]$ depends (asymptotically) only on $\sigma/\mu$ [3]. This suggests a heuristic that $\text{Var}[N]/\text{E}[N]$ is independent of the choice of window for a rate-modulated renewal process. (The validity of this heuristic would depend on the details – for example, the smoothness – of the time-rescaling function and the size of the observation window considered.) Now consider a rate-modulated process with rate $\lambda$. Sticking with the heuristic, it might appear that $\text{Var}[N|\lambda]/\text{E}[N|\lambda]$ should be independent of the choice of window, when $\lambda$ is random. The universal construction in the previous paragraph, which takes $\sigma = 0$, alone illustrates how tricky this interpretation is. (One way to approximately retain the heuristic interpretation, and identifiability, might be to put a lower bound on $\sigma$ and smoothness conditions on the rescaling process $\Lambda^{-1}(t)$. But then these become the crucial parameters of the model.)

In summary, subtle constraints on the class of randomly-modulated renewal processes affect whether i) such a class implies any constraint on the class of spike processes, ii) whether rates are identifiable in such models and iii) whether, for $N$ the spike count in an observation window and $\lambda$ an associated firing rate, $\text{Var}[N|\lambda]/\lambda$ is independent of $\lambda$ and/or the choice of observation window.

**S6. Non-identifiability of $\phi$.**

Referring to Trial-to-Trial Variability and Doubly Stochastic Models, Spike Count Variability in the main text, let $(N, \lambda)$ be a doubly-stochastic model, and define

$$\phi = \phi(\lambda) = \text{Var}[N|\lambda]/\text{E}[N|\lambda] = \text{Var}[N|\lambda]/\lambda$$

where the first identity expresses the assumption that $\phi(\lambda)$ is constant in $\lambda$. Consider repeated samples from $N$. In this case, $\phi$ is not identifiable. The cases I and II in Trial-to-Trial Variability and Doubly Stochastic Models, Spike Count Variability demonstrate this. In Case I, $\phi = 0$. In Case II, $\phi = 1$. In both Case I and II, the distribution of $N$ is the same.

Furthermore, the fact of whether $\phi$ is independent of $\lambda$ is itself not identifiable. Consider generating $(N, \lambda)$ from an equiprobable mixture of Case I and Case II. The distribution of $N$ remains the same, but now $\phi(\lambda)$ is random and depends on $\lambda$. Adjusting the mixture probability alters $\text{E}[\phi(\lambda)]$. Regardless, the distribution of $N$ is the same, independently of the mixture probability.

Thus, repeated samples from $N$ can neither identify $\phi$ nor whether $\phi(\lambda)$ is a constant.

**S7. Estimation of $\text{Var}[\text{E}[N|\lambda]]$ if $\phi$ is known.**

Start with the modeling assumption

$$\text{Var}[N|\lambda] = \phi \text{E}[N|\lambda] \quad \forall \lambda,$$  \hspace{1cm} (S.1)

where $\phi$ is a known constant [4]. Take expectations on both sides to obtain $\text{E}[\text{Var}[N|\lambda]] = \phi \text{E}[N]$. Then substitute this into the decomposition $\text{Var}[N] = \text{Var}[\text{E}[N|\lambda]] + \text{E}[\text{Var}[N|\lambda]]$. This gives $\text{Var}[\text{E}[N|\lambda]] = \text{Var}[N] - \phi \text{E}[N]$, which demonstrates identifiability of $\text{Var}[\text{E}[N|\lambda]]$ and is straightforward to estimate from repeated samples (see Figure 1 from [4] for an example).

**S8. Any collection of discrete-time random spiking processes can be represented as a doubly stochastic model in which spikes are independent, given their firing rate functions.**

The basic observation is simple. In the setting of discrete time under a fine discretization, let a neuron’s spike train be its own firing rate. For any spiking distribution, this provides a version of a firing rate which is consistent with
the firing rate definition in the text (i.e., Equation [5]; equation numbers in this Supporting Information refer to the main text, with the exception of Equations (S-1)-(S-3) defined in Section S10.). For this version of firing rate, spiking is independent, conditioned on the firing rate. Thus, there is no “spike coordination,” for example, in the sense of Equation [7]. (See the 6th column of Figure 1 in the main text.)

(The restriction to discrete time is a mathematical qualification, but not a real limitation in neurophysiology applications. Neurophysiological phenomena can be modeled arbitrarily well by using a suitably fine temporal discretization.)

To be complete, here is a precise formal demonstration. Time is discrete, and the discretization is sufficiently fine that spike trains can be represented by binary variables. The idea is particularly simple in the setup considered in the main text, where experimental time does not play a role. Thus we are interested in the probability distribution of \((N, \lambda, M, \gamma)\). We can think of these random variables as referring to the spiking in a randomly-chosen time bin, or we can simply assume, as in [5], that this joint distribution is invariant to experimental time. Let \(\omega\) represent a sample outcome, so that with \(\omega\) fixed, \(N(\omega) \in \{0, 1\}\) and \(M(\omega) \in \{0, 1\}\) are the spike outcomes associated with sample outcome \(\omega\) (see [6]). Probabilities are assigned directly to each \(\omega\). Every discrete probabilistic spike train model can be set up this way – the only condition is that the probability assignment satisfies \(0 \leq \Pr(\omega) \leq 1\) for all \(\omega\) and \(\sum_{\omega} \Pr(\omega) = 1\). Now, simply consider the “firing rates” specified by \(\lambda(\omega) = N(\omega)\) and \(\gamma(\omega) = M(\omega)\), which are thus well-defined random variables in our probability space. Then \((N, \lambda, M, \gamma)\) describe a random time bin for a pair of doubly stochastic spiking processes with i) \(E[N|\lambda, \gamma] = E[N|\lambda] = \lambda\) and \(E[M|\lambda, \gamma] = E[M|\gamma] = \gamma\) and ii) \(E[Cov[N, M|\lambda, \gamma]] = E[0] = 0\). That is, \(\lambda\) and \(\gamma\) are “firing rates” in the sense of Equation [5], and there is no “spike coordination” in the sense of Equation [7].

The same reasoning applies when experimental time is part of the picture. Think of an experiment of duration \(T\) (in units of time bins). Again we assume time is discrete, so that for two neurons we have a pair of random time series on \(\{1, ..., T\}\). Let \(\omega\) represent a sample outcome and \(t\) index time, so that with \(\omega\) fixed \(N(t, \omega) \in \{0, 1\}^{T}\) and \(M(t, \omega) \in \{0, 1\}^{T}\) are the spike trains associated with sample outcome \(\omega\). Again, simply consider the “rate functions” specified by \(\lambda(t, \omega) = N(t, \omega)\) and \(\gamma(t, \omega) = M(t, \omega)\). Then \((N(t), \lambda(t), M(t), \gamma(t))\) are a pair of doubly stochastic spiking processes, with (random) rate functions \(\lambda(t)\) and \(\gamma(t)\). This reasoning applies to any distribution on spiking processes (i.e., any probability assignment), no matter how “structured” or “structureless.” It extends to any number of neurons. In this way, every multineuron random spiking process can be represented as a doubly stochastic model in which spikes are independent, conditioned on their firing rate functions.

S9. Identifiability of firing rate in the Cox model.

The Cox model brings up an apparent oddity. In the Cox model, spike times are modeled in the continuum, and the spike process is conditionally a Poisson process, conditioned on a random rate-function. In this model, the distribution on this random rate-function is identifiable, and therefore so are any parameters that it determines (see arguments in [7] for example). Contrast this with the random binary time series model (also called a random Bernoulli process), which is in discrete time. Here, there is again a random rate function, and the spike process is conditionally an independent binary (Bernoulli) process, conditioned on the random rate function. In contrast to the Cox model, in the random Bernoulli model, the distribution on random rate-functions is not identifiable. This might appear odd, given the close association between the two models. Indeed, the Poisson process can be defined as a suitable limit of Bernoulli processes, where the limit is taken as the discretization becomes finer. Technical mathematical distinctions in the sense of convergence determine the transfer of intuitions between random Bernoulli and random Poisson processes. In particular, the differences between discrete time and continuous time arise from the peculiarities of allowing spikes to occur arbitrarily close together in time and from the modeling constraints that must be placed on a sequence of Bernoulli processes in order for them to converge properly to a Poisson process. Our view is that the Poisson process should be treated as an extreme idealization. On one hand, it arises naturally from elementary intuitions about independence and homogeneity for point processes. On the other, it is in some senses actually quite constrained and may be inappropriate in many modeling contexts.

S10. Computational Experiments in Figure 1.

Figure 1 in the main text generates repeated trials of spike trains over the interval \([0, 1]\) seconds for 6 different random processes. We note that the examples developed here were chosen for pedagogical rather than scientific reasons, and many alternatives would be equally illuminating. Further, the 6 firing rate processes described here are fully on a
continuum, so that intermediate examples could also be produced using the same ideas. It is not difficult to suggest biological analogies (e.g., stimulus attributes, attentional state, synaptic noise, etc.), at least at the level of a thought experiment, to provide motivation for the information encoded in hidden random processes \( R_1(r, \cdot) \), \( R_2(r, \cdot) \), ..., and \( R_6(r, \cdot) \) (see below).

We discretize time into 1000 disjoint 1 ms time bins centered at 0.0005, 0.0015, 0.0025, ..., 0.9995. A spike on the \( r \)th trial in the \( m \)th time bin of random process \( m \) is indicated by \( N_m(r, t) \), which is 1 for a spike and 0 for no spike, for \( r = 1, \ldots, 20 \) and \( t = 1, \ldots, 1000 \) and \( m = 1, \ldots, 6 \). We also created several different hidden variables \( R_m(r, t) \) where \( r \) indexes trial, \( t \) indexes time bin, and \( m \) indexes the process. Define

\[
\lambda_m(r, t) = E[N_m(r, t)|R_m(r, t)] = Pr(N_m(r, t) = 1|R_m(r, t)).
\]  

(S.2)

The \( \lambda \)'s are implicitly functions of the \( R \)'s, so if the \( R \)'s are random, then the \( \lambda \)'s are random. The \( \lambda \)'s are also firing rates in units of spikes per time bin, because

\[
E[N_m(r, t)|\lambda_m(r, t)] = \lambda_m(r, t).
\]

This follows from the general fact that \( E[N] = E[N|R] = E[N|R] \) for any random variables \( N \) and \( R \). In Figure 1 in the main text, the \( r \)th trial of the \( m \)th column shows the function \( 1000\lambda_m(r, \cdot) \), which is a random firing rate in units of spikes per second (Hz). The rasters at the bottom of each column show the spike times in the corresponding \( N_m(\cdot, \cdot) \).

Here is how we created the \( N \)'s. We first created 100 functions \( p_1(t), \ldots, p_{100}(t) \) defined over the time bins \( t = 1, \ldots, 1000 \). Each of the 100 functions is shown in supplementary figure S2. They take values between 0.003 and 0.030. The shapes of four of the 100 functions (multiplied by 1000) can also be seen in the fourth column (random process 4) of Figure 1 in the main text. The details of how we created these functions are not important, but they are provided at the end of this section. These same 100 functions are used by each of the 6 different processes described below. For each process and each trial we independently and uniformly selected one of these functions. Let \( L_m(r) = \{1, \ldots, 100\} \) be the index of the function that was selected on trial \( r \) for process \( m \), for \( r = 1, \ldots, 20 \) and \( m = 1, \ldots, 6 \). Also, for each process \( m = 1, \ldots, 6 \), trial \( r = 1, \ldots, 20 \), and time bin \( t = 1, \ldots, 1000 \), we independently generated a random variable \( Y_m(r, t) \) taking values in the interval \([-0.0025, 0.0500]\). These random variables have the property that \( E[Y_m(r, t)] = 0 \) for each \( r, t \). The distribution of the random variables does not depend on \( m \) or \( r \), but it does vary with \( t \). Again, the details are not important, but they will be provided at the end of this section. Supplementary Figure S3 shows some examples of samples from \( Y_m(\cdot, \cdot) \). Finally, for each process, trial, and time bin we independently chose \( N_m(r, t) \) where the probability that \( N_m(r, t) = 1 \) was \( p_{L_m(r)}(t) + Y_m(r, t) \). In words, given the \( p \)'s, the \( L \)'s, and the \( Y \)'s, the spiking probabilities on trial \( r \) were generated by the randomly chosen function for that trial (indexed by \( L_m(r) \) and chosen from among the 100 possible \( p \)'s) plus some additive noise (specified by \( Y_m(r, \cdot) \)). Let \( \Theta_m \) denote the collection of all the pieces used to specify the spiking probability, namely, \( \Theta_m = (p_j(t), L_m(r), Y_m(r, t)) : j = 1, \ldots, 100; r = 1, \ldots, 20; t = 1, \ldots, 1000 \) for each \( m = 1, \ldots, 6 \).

Note that the spikes (i.e., the \( N \)'s) for all 6 processes are constructed using the identical random mechanism, so the spiking processes have the same probability distribution and, consequently, are statistically indistinguishable. Any differences in spike times are solely the result of sampling variability. A statistical procedure that detects a difference between the two processes is making a mistake. We will model firing rates (the \( \lambda \)'s) differently in each process, however, by choosing a different \( R \) for each process; see supplemental equation (S.2). For processes 1–5, our \( R_m \) is a function only of \( \Theta_m \), which we emphasize notationally by writing \( R_m(r, t) = R_m(r, t, \Theta_m) \). This allows us to expand the definition of \( \lambda_m \) to obtain

\[
\lambda_m(r, t) = E[N_m(r, t)|R_m(r, t)] = E[N_m(r, t)|R_m(r, t, \Theta_m)] = E[E[N_m(r, t)|\Theta_m]|R_m(r, t, \Theta_m)] = E[p_{L_m(r)}(t) + Y_m(r, t)|R_m(r, t, \Theta_m)] + E[Y(r, t)|R_m(r, t, \Theta_m)],
\]

(S.3)

where we have used the law of total expectation, namely, \( E[U] = E[E[U|V]] \), and its conditional version, \( E[U|W(V)] = E[E[U|V]|W(V)] \), and we have also used fact that \( E[N_m(r, t)|\Theta_m] = Pr(N_m(r, t) = 1|\Theta_m) = p_{L_m(r)}(t) + Y_m(r, t) \).

Here’s how we created the \( R \)'s.

1. Random process 1. \( R_1 \) specifies the average value of \( p \) in each time bin if we know the collection of \( p \)'s but we do not know which \( p_j \) was chosen, i.e., we do not know the value of \( L_1(r) \). It does not vary across trials.
Supplementary Figure 2: The 100 different functions \( p_1(\cdot), \ldots, p_{100}(\cdot) \) over the interval \([0, 1]\) used in the construction of Figure 1 in the main text. The vertical axis in each case is \([0, 0.03]\) spikes/ms or, equivalently, \([0, 30]\) spikes/s.

Formally,
\[
R_1(r,t) = \bar{p}(t) = \frac{1}{100} \sum_{j=1}^{100} p_j(t).
\]

Using (S.3) gives
\[
\lambda_1(r,t) = E[p_{L_1(r)}(t)|\bar{p}(t)] + E[Y_1(r,t)|\bar{p}(t)] = \bar{p}(t) + 0 = \bar{p}(t),
\]
which also does not vary across trials. 1000\(\bar{p}(\cdot)\) is plotted for each trial in column 1 of Figure 1.

2. Random process 2. \( R_2 \) specifies the average value of \( p \) in each time bin if we know the collection of \( p \)'s and we have partial information about which \( p_j \) was chosen. In particular, we know the value \( L_2'(r) = 20\lceil L_2(r)/20 \rceil \). Formally,
\[
R_2(r,t) = \frac{1}{20} \sum_{j=L_2'(r)-19}^{L_2'(r)} p_j(t),
\]
giving
\[
\lambda_2(r,t) = R_2(r,t).
\]
This does vary across trials and is shown (scaled) in column 2 of Figure 1.
3. Random process 3. This is the same as process 2, but instead we know $L_3(r) = 4[L_3(r)/4]$, which is finer information about $L$, giving

$$\lambda_3(r, t) = R_3(r, t) = \frac{1}{4} \sum_{j=L_3(r)-3}^{L_3(r)} p_j(t),$$

and shown (scaled) in column 3 of Figure 1.

4. Random process 4. This is again similar to above, but now we know $L_4(r)$ exactly, i.e., we know which $p_j$ was chosen on each trial. So

$$\lambda_4(r, t) = R_4(r, t) = p_{L_4(r)}(t),$$

as shown (scaled) in column 4 of Figure 1.

5. Random process 5. Now we know $L_5(r)$, but also $Y_5(r, t)$ (the additive, zero-mean noise). Formally,

$$R_5(r, t) = p_{L_5(r)}(t) + Y_5(r, t).$$

Using the next to last step of equation (S.3) gives

$$\lambda_5(r, t) = E[p_{L_5(r)}(t) + Y_5(r, t)|R_5(r, t)] = p_{L_5(r)}(t) + Y_5(r, t).$$

These are shown (scaled) in column 5 of Figure 1. They are noisy versions of the functions in column 4.

6. Random process 6. Finally, we take $R_6(r, t) = N_6(r, t)$ so that $\lambda_6(r, t) = N_6(r, t)$ itself, and the sharp peaks in the firing rate match the spike times in the rasters in column 6 of Figure 1. Note that equation (S.3) does not apply in this case, since $R_6$ is not a function of $\Theta_6$.

To complete the specification of the simulation, we need to specify how we created the $p$’s and the $Y$’s. Let $(\mu_{ij} : i = 1, \ldots, 30; j = 1, \ldots, 100)$ be independent and identically distributed uniform$(0, 1)$ random variables and define

$$p_j(t) = A + 10^{-7}B t^2 e^{-C t} + 10^{-4}D \sum_{i=1}^{30} \sum_{e=5}^{5} \sum_{t=0.5}^{1000} e^{-D} \left( t - \mu_{ij} \right)$$

where $A = .00275, B = 5, C = 0.0075$, and $D = 40$. For the random $\mu$’s in our simulations we always had $0.003 \leq p_j(t) \leq 0.03$. For each $m = 1, \ldots, 6$, $Y_m(r, t) = 0.0525B_m(r, t) - 0.0025$ where the $B_m(r, t)$ are independent $Beta(a(t), 20a(t))$ random variables, for $r = 1, \ldots, 20$ and $t = 1, \ldots, 1000$. Since the mean of a $Beta(a, b)$ random variable is $a/(a + b)$, we have $E[B_m(r, t)] = 1/21$ and $E[Y_m(r, t)] = 0.0525/21 - 0.0025 = 0$. Also, Beta random variables take values in $[0, 1]$, so $Y_m(r, t)$ takes values in $[-0.0025, 0.0500]$. We chose $a(t)$ so that the variance of $B_m(r, t)$ was $\sigma(t) = \max(-\cos(10\pi(t/1000 - 0.05)), 10^{-5})/882$, which has periodic bursts of high variability and regions of very low variability. (The number 882 was chosen to ensure that the target variance was possible for some choice of $a(t)$.) The specific relationship between $a(t)$ and $v(t)$ in this case is $a(t) = 20/(21^3v(t)) - 1/21$.

References