1999

INTERNATIONAL SERIES IN
OPERATIONS RESEARCH & MANAGEMENT SCIENCE

Frederick S. Hillier, Series Editor
Stanford University

Saigal, R.: LINEAR PROGRAMMING: A MODERN INTEGRATED ANALYSIS
Nagurney, A. & Zhang, D.: PROJECTED DYNAMICAL SYSTEMS AND VARIATIONAL INEQUALITIES WITH APPLICATIONS
Padberg, M. & Rijal, M.: LOCATION, SCHEDULING, DESIGN AND INTEGER PROGRAMMING
Vanderbei, R.: LINEAR PROGRAMMING: FOUNDATIONS AND EXTENSIONS
Jaiswal, N.K.: MILITARY OPERATIONS RESEARCH: QUANTITATIVE DECISION MAKING
Gal, T. & Greenberg, H.: ADVANCES IN SENSITIVITY ANALYSIS AND PARAMETRIC PROGRAMMING
Prabhu, N.U.: FOUNDATIONS OF QUEUEING THEORY
Fang, S.-C., Rajasekera, J.R. & Tsao, H.-S.J.: ENTROPY OPTIMIZATION AND MATHEMATICAL PROGRAMMING
Yu, G.: OPERATIONS RESEARCH IN THE AIRLINE INDUSTRY
Ho, T.-H. & Tang, C.S.: PRODUCT VARIETY MANAGEMENT
El-Taha, M. & Stidham, S.: SAMPLE-PATH ANALYSIS OF QUEUEING SYSTEMS
Miettinen, K.M.: NONLINEAR MULTIOBJECTIVE OPTIMIZATION
Chao, H. & Huntington, H.G.: DESIGNING COMPETITIVE ELECTRICITY MARKETS
Weglarz, J.: PROJECT SCHEDULING: RECENT MODELS, ALGORITHMS & APPLICATIONS
Sahin, I. & Polatoglu, H.: QUALITY, WARRANTY AND PREVENTIVE MAINTENANCE
Tavares, L.V.: ADVANCED MODELS FOR PROJECT MANAGEMENT
Tayur, S., Ganeshan, R. & Magazine, M.: QUANTITATIVE MODELING FOR SUPPLY CHAIN MANAGEMENT
Weyant, J.: ENERGY AND ENVIRONMENTAL POLICY MODELING
Shanthikumar, J.G. & Sumita, U.: APPLIED PROBABILITY AND STOCHASTIC PROCESSES

APPLIED PROBABILITY AND STOCHASTIC PROCESSES

edited by

J. G. Shanthikumar
and
Ushio Sumita

Kluwer Academic Publishers
Boston/London/Dordrecht

Distributors for North, Central and South America:
Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, Massachusetts 02061 USA
Telephone (781) 871-6600
Fax (781) 871-6528
E-Mail <kluwer@wkap.com>

Distributors for all other countries:
Kluwer Academic Publishers Group
Distribution Centre
Post Office Box 322
3300 AH Dordrecht, THE NETHERLANDS
Telephone 31 78 6392 392
Fax 31 78 6546 474
E-Mail <orderdept@wkap.nl>


Electronic Services <http://www.wkap.nl>

Library of Congress Cataloging-in-Publication Data

Applied probability and stochastic processes / edited by
J. G. Shanthikumar and Ushio Sumita.
p. cm. — (International series in operations research & management science: 19)
Includes bibliographical references and index.
ISBN 0-7923-8439-3 (acid-free paper)
1. Probabilities. 2. Stochastic processes. I. Shanthikumar, J. George. II. Sumita, U. (Ushio), 1949– . III. Series.
QA273.18.A55 1999
519.5—dc21 99-17291
CIP

CONTENTS

TRIBUTE TO JULIAN KEILSON

J. G. Shanthikumar and Ushio Sumita

More than 45 years have passed since Professor Keilson began contributing to probability theory. His intellectual roots are in mathematics and physics. At Brooklyn College's mathematics department, which nurtured such substantial mathematicians as Richard Bellman, he fell in love with mathematical analysis and with the beauty of nineteenth-century potential theory. In the belief that theoretical physics would provide a home for his mathematical needs, he switched his major to physics in his senior year and entered Harvard's doctoral program in the summer of 1947. Further strength in mathematical physics developed under Julian Schwinger, from whom he received his Ph.D. in June of 1950. Since then, starting first as a theoretical physicist, he has been making tremendous contributions in various areas, including applied probability, stochastic processes, and computational probability, among others. In what follows, his major contributions are briefly summarized by classifying them into several categories. Paper references can be found in the list of publications by Professor Keilson that follows this summary.

From Physics to Probability

Discomfort with the formal, heuristic, and pragmatic character of quantum mechanics and quantum electrodynamics led him back to mathematics. As a postdoctoral Fellow in Electronics at Harvard (1950–1952), he became interested in Brownian motion and the theory of electrical noise. His first paper [1] with J. Storer addressed the bridge between the jump character of particle motion sample paths described by Boltzmann's integral equation and the continuous sample path character of the Fokker–Planck partial differential equation. Related papers [2–5] at M.I.T.'s Lincoln Laboratory (1952–1956) studying diffusion in semiconductors were tied to transistor technology, then in its infancy. Just as Professor Philip Morse's role in early aspects of operations research was based on his strength in mathematical physics, these diffusion and noise studies became the foundation for Professor Keilson's subsequent activity in applied probability.

[14] Takács, L. Fluctuation problems for Bernoulli trials. *SIAM Rev.* **21**, 222–228, 1979.

[15] Trotter, H. A property of Brownian motion paths. *Illinois J. Math.* **2**, 425–433, 1958.

[16] Wolfram, S. *Mathematica. A System for Doing Mathematics by Computer*, 2nd ed. Addison-Wesley, Redwood City, CA, 1991.

CHAPTER 5

PROBABILISTIC TOKEN CAUSATION: A BAYESIAN PERSPECTIVE

Elja Arjas

5.1 Introduction

Many authors (e.g., Good [8,9] and Eells [5,6]) distinguish between two kinds of probabilistic causality: *the tendency of C to cause E* and *the degree to which C actually caused E*. The former, a generic form of causation, can be discussed by comparing two prediction probabilities, one conditional on the occurrence of C and the other on its "counterfactual" event, where C does not occur. The latter, a singular form, is often called *token causality* and corresponds to finding a causal explanation of the occurrence of an event after it has been observed to happen. The purpose of this chapter is to formulate token causality by using the mathematical framework of marked point processes (MPPs) and their associated prediction processes. The same framework was used by Arjas and Eerola [2] for considering predictive causality. Therefore, this chapter can also be seen as an attempt to bridge the gap between these two types of causality reasoning.

In order to explain these ideas in concrete terms, and to compare them with those presented by others, we consider the following three examples.

5.1.1 Example 5.1

In Example 5.1, suppose children in a certain country are routinely inoculated with a vaccine that completely prevents the onset of an infectious disease. The disease is often fatal, and, in an unvaccinated population, one can expect approximately 2% of the children to die from the disease before age 5. Unfortunately, some children, approximately 1 in 100,000, are allergic to the vaccine. The inoculation of such a child leads to an allergic reaction that is fatal in approximately half the cases. Clearly, the effect of inoculation is beneficial when considering the five-year survival of the children. Consider, however, a particular child who quickly after the inoculation developed the symptoms of the allergic reaction and died. It seems obvious that the inoculation actually caused the death.

5.1.2 Example 5.2

(Example 5.2 is taken from Rosen [11], quoted directly from Eells [5]). A golfer makes a shot, the ball hits the limb of a tree, and then, improbably enough, the ball is deflected directly into the cup for a birdie. It is supposed that the probability for a birdie is low given the shot in the first place, but when you add the information that the ball will hit the tree, the probability of a birdie becomes even lower. So, even though the ball's hitting the tree is what *actually caused* the birdie, its hitting the tree nevertheless *lowers the probability* of a birdie.

5.1.3 Example 5.3

(Example 5.3 is taken from Cartwright [4], quoted directly from Eells [5]). Consider a normal, healthy plant. The probability of its surviving and being in good health in a year is 0.98. However, Nancy sprayed it with defoliant, which the can label said was 90% effective in killing plants. So the probability of the plant's surviving for a year drops from 0.98 to 0.1. Nevertheless, the plant survives, and it is again healthy in a year.

Although this third example has much in common with the other two, our intuition will probably not say that the spraying token caused the plant's survival. Why would this be so?

Our plan is to analyze each of these examples in intuitively plausible terms. After doing so, we will show how the mathematical framework introduced in Arjas and Eerola [2] can be applied in order to provide a formal explanation for our intuitions and to build a bridge between singular (token) causality and generic (predictive) causality.

5.2 A Preliminary Analysis

In Example 5.1, it is plain that the two situations — before the inoculation and after the allergic reaction was observed — are different in that, in the latter instance (*a posteriori*), it is *known* that the child who died had the particular form of allergy. If the child had been tested for the allergy before the inoculation, he or she would not have been inoculated in the first place, but if the inoculation had been administered, the allergic reaction would have been fully predictable. According to the information given in Example 5.1, the short-term survival probability of a child who is known to have the allergy factor is approximately 0.5 if he or she is inoculated and close to 1.0 if not. This difference in the short-term predictions certainly justifies the view that inoculation was the true cause of the observed death.

A more formal way of expressing the above idea is to say that the indicator marking the presence of the allergy factor was a latent (unobservable) variable and that, by observation of the child's behavior after inoculation, the distribution of the variable was updated from the earlier assigned probability of 10^{-5} for the presence of the allergy factor and $1-10^{-5}$ for no allergy factor to the posterior assigned probability of 1.0 for the presence of the allergy factor.

Consider then Example 5.2. Here the claim that the birdie was caused by the ball

hitting a tree limb comes from the idea that, before the ball hit the limb, its speed and/or direction of travel must have been such that, if the limb hadn't been there, it would have been impossible for the ball to go directly into the cup. (Essentially the same idea is expressed by a number of authors; see Eells [5] for references).

As in Example 5.1, we can think of Example 5.2 in terms of a latent variable, viz. the physical motion parameters just mentioned. Given the earlier information about the values of these parameters, and supposing that there was no possibility to alter the direction in which the ball was traveling (either by a tree limb or otherwise), a birdie would have had zero probability. When the ball hit the limb, the situation changed, and soon afterward one would have given a birdie a small but positive probability. This difference explains our intuition that the token of hitting the limb caused the birdie.

In Example 5.3, it would be wrong to say that the plant survived *because* defoliant was sprayed or that the spraying caused its survival. Rather, one should say that the plant survived *despite* the spraying [5]. However, putting aside the issue of the direction in which an attributed cause changes the predictions, we are again led to consider a latent variable model. As is often assumed in toxicology, different individuals can have different tolerances towards toxic agents: a given dose level of a toxic agent kills those individuals whose tolerance level is lower than the dose level administered, whereas others survive. We can now assume that the plant that Nancy sprayed survived because its tolerance level exceeded the administered dose level. At the time the plant was sprayed, its actual tolerance level was unknown but was thought to be less than the dose level, with probability 0.9. After one year, the tolerance level was actually known to have exceeded the dose level.

5.3 A Mathematical Framework for Token Causality

Although one would perhaps view most physical processes as being in some sense continuous in real time, it is often helpful to simplify such processes and think of them as series of events. Each event then becomes a pair (T, X), where T is the occurrence time and X is a description of the event that happened at time T. Often one might view such events as landmarks in an essentially continuous development (e.g., in human life, the commencement and completion of education, marriage, birth of first child, retirement, death). This kind of simplification seems also essential for causal reasoning: One typically thinks of a chain of events, with the earlier events influencing the later ones in the sense that a change in the former would very likely have changed the outcome of the latter.

We now make an attempt to express the key ideas regarding token causality by using the general framework of marked point processes (MPPs) and their associated predictions. Here we build on the filtering ideas reviewed in Arjas [1], and presented in more detail in Arjas, Haara, and Norros [3]. In order to concentrate on the ideas that are central to token causality, we limit the mathematical details to the minimum possible.

Consider some evolution that is not fully observable and that is modeled in terms of an MPP, say $(\mathbf{T}, \mathbf{X}) = (T_n, X_n)_{n\geq 1}$. We call this process *the underlying MPP*. Let $(\hat{\mathbf{T}}, \hat{\mathbf{X}}) = (\hat{T}_n, \hat{X}_n)_{n\geq 1}$ be another MPP, which is fully observable and is determined by

(T, X) in the sense that at each time t the observer knows the pre-t history $\hat{H}_t = \{(\hat{T}_n, \hat{X}_n) : T_n \le t\}$ and that $\hat{H}_t$ is fully determined by the (partially unknown) underlying pre-t history $H_t = \{(T_n, X_n) : T_n \le t\}$. The observation of $\hat{H}_t$ may therefore be said to constitute partial information about H_t.

A systematic description of what is known about the underlying pre-t history, based on what has been observed up to time t, is given by the conditional probability distribution

$$\hat{\pi}_t(\cdot) = P\left(H_t \in \cdot \,\middle|\, \hat{\mathcal{F}}_t\right) \tag{5.1}$$

on the space $(\mathbb{H}, \mathcal{H})$ of all histories. Here $\hat{\mathcal{F}}_t = \sigma\{\hat{H}_t\}$, the σ-field generated by the pre-t observations, represents the observed information up to time t. The distribution $\hat{\pi}_0(\cdot)$ can be viewed as a prior, which is then updated according to what has been observed. This updating can be done recursively, essentially by using Bayes' rule (see [1,3]). In the Bayesian spirit, $\hat{\pi}_t(\cdot)$ can then be interpreted as a historian's view, at time t, of what happened in the underlying process in the past.

As in [2], we describe all elements in a causal chain in terms of an MPP. Two of the marked points in $(\hat{\mathbf{T}}, \hat{\mathbf{X}})$, say (T_C, C) and (T_E, E), are given a special role in this description: that of a *cause* and that of an *effect* or *response*. We assume that $T_C < T_E$ on $\{T_C < \infty\}$, with both T_C and T_E being stopping times with respect to the observed filtration $(\hat{\mathcal{F}}_t)$. Intuitively speaking, both C and E are observed as they happen. In the problems relating to token causality, we assume that there is a further marked point, say (T_L, L), such that $T_L < T_C$ on $\{T_L < \infty\}$. We assume that (T_L, L) is a marked point in the underlying process $(\mathbf{T}, \mathbf{X})$ but *not* in the observed process $(\hat{\mathbf{T}}, \hat{\mathbf{X}})$. In order to keep the present model as simple as possible, we assume here that (T_L, L) is the only such point. The interpretation is that L is a latent status variable and T_L is the time at which it is manifested. In many concrete examples it is natural to choose $T_L = 0$, with L being an indicator variable taking values 0 and 1.

The basic idea of Arjas and Eerola [2] was to consider the conditional distribution of the response (T_E, E) as a stochastic process, with the conditioning corresponding, progressively in time t, to the observation of the pre-t history of an MPP. Here, since we have hypothesized the existence of two different MPPs, one *underlying* and the other *observed*, we are led to consider two such prediction processes. As explained in [2], the predictions can be realized in terms of transition kernels from the space of histories into the values of the response variable. In particular, we can find kernels $\mu_t^*(\cdot,\cdot))_{t\ge 0}$ and $\hat{\mu}_t^*(\cdot,\cdot))_{t\ge 0}$ satisfying suitable regularity properties and such that the representations $P((T_E, E) \in A|\mathcal{F}_t) = \mu_t^*(H_t, A)$ and $P((T_E, E) \in A|\hat{\mathcal{F}}_t) = \hat{\mu}_t^*(\hat{H}_t, A)$ hold. The former prediction is hypothetical in the sense that it is based on information $\mathcal{F}_t$, which includes knowledge of the value of the latent variable (T_L, L) if $T_L \le t$, whereas the latter prediction is "real" since $\hat{\mathcal{F}}_t$ is actually observed.

Consider now the situation arising at time T_E, i.e., when the response (T_E, E) has just been observed. We have then also observed (T_C, C) at an earlier time T_C, and these two marked points both belong to the observed history $\hat{H}_{T_E}$. However, the latent variable (T_L, L) could not be observed, and knowledge about its value must therefore be described by using the conditional distribution $\hat{\pi}_{T_E}(\cdot)$. (Sometimes, by

having observed the history $\hat{H}_{T_E}$, we have effectively come to know (T_L, L) as well; in that case, $\hat{\pi}_{T_E}$ places unit mass on the history $H_{T_E} = \{(T_L, L)\} \cup \hat{H}_{T_E}$.)

Let us now pretend, for a while, that we know what (T_L, L) is. In that case, the role of C as a cause of (T_E, E), acting at time T_C, was assessed (see [2] by considering the relationship between the predictions $\mu_{T_C}^*(H_{T_C}; \cdot)$ and $\mu_{T_C}^*(H_{T_C-}; \cdot)$. Both predictions are made at time $t = T_C$, but they are based on different assumptions about the occurrence of C: The former corresponds to conditioning on the "true" underlying pre-T_C history, in which C occurred at time T_C, and the latter on a history that is otherwise identical except that now one is making the *counterfactual* assumption that "nothing occurs at t," i.e., there is no marked point at T_C. C is then viewed as a (*prima facie*) *cause* (relative to $(\mathcal{F}_t)$) of the event $\{(T_E, E) \in A\}$ if the difference

$$\mu_{T_C}^*(H_{T_C}; A) - \mu_{T_C}^*\left(H_{T_C-}; A\right) \tag{5.2}$$

is positive.

In reality, however, since (T_L, L) is not known to the observer at time T_C, only the predictions $\hat{\mu}_{T_C}^*(\hat{H}_{T_C}; \cdot)$ and $\hat{\mu}_{T_C}^*(\hat{H}_{T_C-}; \cdot)$ can be considered then. Moreover, it is easy to find examples in which (5.2) is positive but the difference

$$\hat{\mu}_{T_C}^*\left(\hat{H}_{T_C}; A\right) - \hat{\mu}_{T_C}^*\left(\hat{H}_{T_C-}; A\right) \tag{5.3}$$

is zero or negative. In other words, the occurrence of C at T_C would be viewed as a prima facie cause if (T_L, L) were known, but no longer when it is not. Similarly, C can sometimes be viewed as a cause in the latter case and not in the former.

Let us then turn to the actual token causality. It is fundamental to all reasoning in probabilistic causality that one should compare conditional probabilities of events of the form $\{(T_E, E) \in A\}$, under different circumstances that are made explicit in the conditioning. Token causality, being concerned with finding an explanation of the response after it was actually observed, must somehow make use of the after-the-fact information available at time T_E. But (T_E, E) itself is an element of $\hat{H}_{T_E}$, so it would be foolish to consider (trivial) conditional probabilities of $\{(T_E, E) \in A\}$, given that (T_E, E) is actually known. This leads us to the idea of *backprediction*, where one uses the knowledge of $\hat{H}_{T_E}$ in making an assessment of the temporally earlier but unobserved latent variable (T_L, L). Based on that assessment, one then makes predictions concerning $\{(T_E, E) \in A\}$ under two different scenarios: the true development in which (T_C, C), the potential cause, actually occurs, and the counterfactual one from which (T_C, C) is removed. (Note here that, while (T_C, C) does not appear in the counterfactual conditioning concerning $\{(T_E, E) \in A\}$, it is nevertheless used in the probability assessment concerning (T_L, L), which is always conditional on $\hat{H}_{T_E}$.)

We now make an attempt to express these ideas more formally. As we saw above, the knowledge about the latent variable (T_L, L) at time T_E, when the response has already been registered, can be expressed in terms of the distribution $\hat{\pi}_{T_E}(\cdot)$ of underlying pre-T_E histories. Let us now make the convention of working on the canonical path space of the underlying MPP $(\mathbf{T}, \mathbf{X})$, in which case $\pi_t \equiv \pi_t \circ (T_L, L)^{-1}$ is the corresponding $\hat{\mathcal{F}}_t$-conditional distribution of the latent variable (T_L, L). Provided that

we had this knowledge (represented by $\tilde{\pi}_{T_E}$) about (T_L, L) at time T_C, and C occurs then, our prediction probability that an event of the form $\{(T_E, E) \in A\}$ would happen would be

$$\tilde{\pi}_{T_E}\left[\mu^*_{T_C}(\{\cdot\} \cup \hat{H}_{T_C}; A)\right] \underset{def}{=} \int \tilde{\pi}_{T_E}(dt \times dl)\mu^*_{T_C}(\{(t, l)\} \cup \hat{H}_{T_C}; A), \tag{5.4}$$

where the integration is over the range of (T_L, L). This is nothing but the conditional expected value of the predictions $\mu^*_{T_C}(H_{T_C}; A)$ with respect to the distribution $\tilde{\pi}_{T_E}$. When formulating the counterfactual alternative in which nothing occurs at time T_C, we are led to consider the analogous expectation

$$\tilde{\pi}_{T_E}\left[\mu^*_{T_C}(\{\cdot\} \cup \hat{H}_{T_{C^-}}; A)\right] \underset{def}{=} \int \tilde{\pi}_{T_E}(dt \times dl)\mu^*_{T_C}(\{(t, l)\} \cup \hat{H}_{T_{C^-}}; A) \tag{5.5}$$

The difference between (5.4) and (5.5) is merely that $\hat{H}_{T_C}$, which includes the point (T_C, C), is replaced by $\hat{H}_{T_{C^-}}$, which does not. We then interpret (T_C, C) as a token cause of the response (T_E, E) if (5.4) is positive and (5.5) is zero. Another possibility is to require that the difference between (5.4) and (5.5) be positive.

Note that the predictive probabilities $\hat{\mu}^*_{T_C}(\hat{H}_{T_C}; A)$ and $\hat{\mu}^*_{T_C}(\hat{H}_{T_{C^-}}; A)$ appearing in (5.3) can also be viewed as conditional expected values and can be written in an integral form similar to (5.4) and (5.5). We have (cf. [1])

$$\hat{\mu}^*_{T_C}(\hat{H}_{T_C}; A) = E\left(\mu^*_{T_C}(H_{T_C}; A)\middle|\hat{\mathcal{F}}_{T_C}\right) = \tilde{\pi}_{T_C}\left[\mu^*_{T_C}(\{\cdot\} \cup \hat{H}_{T_C}; A)\right], \tag{5.6}$$

and similarly,

$$\hat{\mu}^*_{T_C}(\hat{H}_{T_{C^-}}; A) = \tilde{\pi}_{T_C}\left[\mu^*_{T_C}(\{\cdot\} \cup \hat{H}_{T_{C^-}}; A)\right]. \tag{5.7}$$

In other words, the difference between (5.4) and (5.6) (respectively, (5.5) and (5.7)) is that in the former the knowledge about the latent variable (T_L, L) is based on $\hat{\mathcal{F}}_{T_E}$ and in the latter on $\hat{\mathcal{F}}_{T_C}$. This reflects the difference between the explanatory (posterior) and predictive types of reasoning. One could perhaps say that (5.4) and (5.5) represent a pretension, or hindsight, on the part of the observer: "Had I known the underlying conditions that produced the observed response as well as I know them now, I would have assessed the role of C to be. . . ."

5.4 An Analysis of the Examples

Our plan in this section is to connect the earlier analysis of Examples 5.1 to 5.3, presented in Section 5.1, to the above mathematical formalism.

5.4.1 Example 5.1

It is natural to choose the latent variable L as the indicator of the allergy factor and to take T_L to be some fixed early age (e.g., one year) at which the corresponding status is manifested. We let T_C be the time of inoculation, with C (inoculation) being

the suspected token cause. Then $\hat{H}_{T_C} = \{(T_C, C)\}$ and $\hat{H}_{T_{C^-}} = \emptyset$ (history with no points). Let also

$$T_E = \min(T_C + 5 \text{ years}, \text{time of death}),$$

$$E = \begin{cases} \text{"alive" if alive at } T_E, \\ \text{"observed symptoms" if dead at } T_E. \end{cases}$$

Consider then $A = (0, \infty) \times \{\text{"symptoms of the allergic reaction to the vaccine in question"}\}$ and suppose that the event $\{(T_E, E) \in A\}$ actually occurred. Then the posterior $\tilde{\pi}_{T_E}$ assigns unit mass to the marked point $(T_L, 1)$ corresponding to the event $\{L = 1\} = \{\text{"allergy factor present"}\}$, whereby the prediction probability (5.4) of $\{(T_E, E) \in A\}$ becomes

$$\tilde{\pi}_{T_E}\left[\mu^*_{T_C}(\{\cdot\} \cup \{(T_C, C)\}; A)\right] = \mu^*_{T_C}(\{(0, 1), (T_C, C)\}; A) = 0.5.$$

By comparison, since the same symptoms cannot be obtained in the counterfactual event that the child is not inoculated, we have

$$\tilde{\pi}_{T_E}\left[\mu^*_{T_C}(\{\cdot\}; A)\right] = \mu^*_{T_C}(\{(0, 1)\}; A) = 0.$$

In other words, a posterior analysis shows that the observed response was only possible in the event that inoculation indeed took place. It is therefore justified to call inoculation a token cause of the observed death, in agreement with our earlier discussion.

Recall, however, that given only the prior (population-level) information about the presence of the allergy factor, inoculation increases the five-year survival probability. This is reflected in the difference

$$\hat{\mu}^*_{T_C}(\{(T_C, C)\}; A) - \hat{\mu}^*_{T_C}(\emptyset; A),$$

which is positive for $A = (0, \infty) \times \{\text{"alive"}\}$.

5.4.2 Example 5.2

Here we choose T_C as the time the ball hit the limb and T_E as the time the ball stopped moving. Let E be the indicator of a birdie. It is then convenient to let T_L be the time when the ball had reached some short distance, say an inch, from the limb, and to let L be a vector of physical motion parameters at T_L describing the speed and rotation of the ball and the direction in which it traveled.

Without having specified the physical circumstances more closely, it is hard to give credible numerical values to the various prediction probabilities. However, in the event that $\{E = 1\}$ (a birdie) was observed, and when it is known that hitting the limb always alters the direction in which the ball travels, we should obviously assign zero posterior probability to any set of motion parameters that, given the counterfactual event that the limb was suddenly moved aside, would then have resulted in

a birdie. Therefore, considering $A = (0, \infty) \times \{1\}$, we have that $\tilde{\pi}_{T_E}[\mu^*_{T_C}(\{\cdot\} \cup \{(T_C, C)\}; A)]$ is some still fairly small but positive number, whereas $\tilde{\pi}_{T_E}[\mu^*_{T_C}(\{\cdot\}; A)]$, corresponding to the counterfactual removal of the limb, must be zero.

5.4.3 Example 5.3

It is already clear from the discussion in Section 5.1 that we should choose T_C = the time of spraying, C = the actual dose level, T_L= some time shortly before T_C, L = the unknown tolerance level of the plant, $T_E = T_C + 1$ year, and finally, E = the indicator of being alive at T_E.

Let $A = (0, \infty) \times \{\text{"alive"}\}$. The prior probabilities $\hat{\mu}^*_{T_C}(\{(T_C, C)\}; A)$ and $\hat{\mu}^*_{T_C}(\emptyset; A)$ that a sprayed or unsprayed plant, respectively, survived until T_E were assessed to be 0.1 and 0.98. On the other hand, the posterior analysis shows that, because the plant survived, we must have $L > C$. But for such a posterior we can assess $\tilde{\pi}_{T_E}[\mu^*_{T_C}(\{(T_C, C)\}; A)]$ to be some positive number close to one, maybe somewhat lower than 0.98, and $\tilde{\pi}_{T_E}[\mu^*_{T_C}(\emptyset; A)] = 0.98$. The difference that, considering survival, is even in the "wrong direction" would certainly not justify the idea that the token of spraying caused the response.

A particular property of this example is that, although spraying the plant with defoliant was a strong prima facie cause contributing to its death within a year, spraying was not a token cause when the plant did not die. The formal calculations above support this intuition. Moreover, observing the plant's condition day by day after it was sprayed must have progressively changed Nancy's prediction about what was going to happen in a year's time. Based on such observations, and being a clever person, she must have updated her predictions in time according to

$$P\left((T_E, E) \in \cdot \,\middle|\, \hat{\mathcal{F}}_t\right) = \hat{\mu}^*_t\left(\hat{H}_t; \cdot\right).$$

This, in our view, corresponds to "the probabilities changing in time" as considered by Eells [5,6] in his discussion of Example 5.3.

5.5 A Discussion

Causality, no doubt, is a basic element in how we interpret and structure the world around us. Quoting from Pearl ([10], p. 383):

> Human beings exhibit an almost obsessive urge to mold empirical phenomena conceptually into cause-effect relationships. This tendency is, in fact, so strong that it sometimes comes at the expense of precision and often requires the invention of hypothetical, unobservable entities (such as the ego, elementary particles, and supreme beings) to make theories fit the mold of causal schemata.

In the above analysis of token causality, it has been crucial that both (T_C, C) and (T_E, E) have actually been observed to happen. Therefore, the question of whether or not (T_C, C) occurred is not the issue; the issue is whether we should interpret its occurrence as a cause of the observed response (T_E, E). Likewise, it seems meaningless to talk about the *probability* that "(T_C, C) token caused (T_E, E)," since $\{(T_C, C)$ *caused* $(T_E, E)\}$ is not a well-defined event, at least not in our probability framework. Neither do we attempt here to define a concept such as the *degree of causation* [7,9].

On the other hand, there are situations in which it is perfectly natural to look for a cause, among two or more different alternatives, of some observed response, and to consider the probability that a particular one among these was "the true cause." This is the case, for example, when one is pondering about what caused the failure of a device, or the death of an individual. In such instances it is natural to interpret the cause, when it is unknown to the observer, as a latent random variable. In other words, (T_C, C) takes a similar role to that of (T_L, L) above, and one may well consider its posterior distribution $P((T_C, C) \in \cdot \,|\, \mathcal{F}_{T_E}) = \hat{\pi}_{T_E} \circ (T_C, C)^{-1}$ when the response (T_E, E) has been observed. A technical inspection of a failed device, or autopsy of a dead individual, may then reveal the true value of C.

It is important to distinguish between these two cases where (T_C, C) has, or has not, been observed. This becomes very clear from example 5.2: Should the observer not know that the ball had hit the limb of a tree, it would be quite farfetched to even think of that as a possible explanation for an observed birdie.

A final question, which has so far been only implicit in our discussion, is how the statements involving probabilities should be interpreted. Our formal analysis has been technically neutral in the sense that we have only used the widely accepted rules of probability calculus. Most authors seem to think that causality is something that exists as an objective physical concept, and then it is obviously desirable that probabilities, which appear in the key definitions here, would exist in some similar sense. On the other hand, the intuition behind the formalism presented here is that of *backprediction*, which is essentially a mental operation in which the observer puts himself back in time, using then information that in reality was not available at that time. In this, the subjective interpretation of probability as a quantitative expression of an individual's uncertainty seems to be the natural one to use.

References

[1] Arjas, E. Survival models and martingale dynamics (with discussion). *Scand. J. Stat.* **16**, 177–225, 1989.

[2] Arjas, E., and Eerola, M. On predictive causality in longitudinal studies. *J. Stat. Planning Inference* **34**, 361–386, 1993.

[3] Arjas, E., Haara, P., and Norros, I. Filtering the histories of a partially observed marked point process. *Stoch. Proc. Appl.* **40**, 225–250, 1992.

[4] Cartwright, Nancy. Regular associations and singular causes. In: Skyrms, Brian, and Harper, William L. (eds), *Causation, Chance, and Credence*. Kluwer Academic Publishers, Dordrecht, 1988, pp. 79–97.

[5] Eells, Ellery. Probabilistic causal levels. In: Skyrms, Brian, and Harper, William L. (eds), *Causation, Chance, and Credence*. Kluwer Academic Publishers, Dordrecht, 1988, pp. 109–133.

[6] Eells, Ellery. *Probabilistic Causality*. Cambridge University Press, Cambridge, 1991.

[7] Good, I. J. A causal calculus. *Br. J. Phil. Sci.* **11**, 305–318, 1961; **12**, 43–51, 1961; **13**, 88, 1962.

[8] Good, I. J. Causal propensity: a review. In: Asquith, P. D., and Kitcher, P. (eds), *PSA* **2**, 829–850. Philosophy of Science Association, East Lansing MI, 1984.

[9] Good, I. J. Causal tendency: a review. In: Skyrms, Brian, and Harper, William L. (eds), *Causation, Chance, and Credence*. Kluwer Academic Publishers, Dordrecht, 1988, pp. 23–50.

[10] Pearl, Judea. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Mateo, 1988.

[11] Rosen, D. A. In defense of a probabilistic theory of causality. *Phil. Sci.* **45**, 604–613, 1978.

[12] Suppes, Patrick. *A Probabilistic Theory of Causality*. North Holland, Amsterdam, 1970.

CHAPTER 6

ON A STATISTICAL ALGORITHM TO DECODE HEAVILY CORRUPTED LINEAR CODES

I. N. Kovalenko and M. N. Savchuk

6.1 Problem Statement

As is well known, decoding general linear codes is an NP-hard problem [1]. Thus, the design of decoding algorithms that have a lower complexity in special cases compared to that of decoding by the maximum likelihood method is of major interest. Levitin and Hartmann [3] present a decoding algorithm having the complexity of order $2^{F(\rho)n}$ where $F(\rho)$ is a function of the code rate ρ with $F(\rho) < 1$ for $\rho > 0.1887$ and $F(\rho) = 1$ for $\rho \le 0.1887$. If codes are heavily corrupted (with means that the probability of the distortion of a symbol is not considered as a small variable for large n), the complexity of decoding by this method equals 2^n and coincides with the complexity of the maximum likelihood method. In [2], it is shown that there exists an algorithm to decode heavily corrupted codes with complexity of order $2^{n/c\log_2 n}$ as $n \to \infty$, $c = \text{const} > 0$.

In this chapter, a simple algorithm to decode heavily corrupted codes is presented that is especially effective for disperse generating matrices.

Consider a linear code over GF(2) (Galois Field; simply speaking, additions are considered modulo 2) with a generating matrix A that transforms an information word $X = (x_1, \ldots, x_n)$ to a codeword $Y = (y_1, \ldots, y_{N_0})$. The latter is transmitted over a memoryless binary symmetric channel with the errorless transmission probability for the ith symbol equal to $p_i > 1/2$ and turns to a received word $\tilde{Y} = (\tilde{y}_1, \ldots, \tilde{y}_{N_0})$:

$$A = (a_{ij}), \quad i = 1, \ldots, N_0, \quad j = 1, \ldots, n;$$
$$P(\tilde{y}_i = y_i) = p_i > \frac{1}{2}, \quad i = 1, \ldots, N_0. \tag{6.1}$$

Consider also a series of M linear codes over GF(2) with the same generating matrix and different information words $X^l = (x_1^l, \ldots, x_n^l)$ that get transformed to codewords $Y^l = (y_1^l, \ldots, y_{N_0}^l)$ and result in received words $\tilde{Y}^l = (\tilde{y}_1^l, \ldots, \tilde{y}_{N_0}^l)$ on having been transmitted over a memoryless binary symmetric channel with the errorless transmission probability of the ith symbol in the lth codeword equal to p_i^l.