Title

Growing a Mind with Statistical Inference

Permalink

https://escholarship.org/uc/item/5n24n2g0

Journal

Berkeley Scientific Journal, 27(2)

ISSN

1097-0967

Author

Niyogi, Anjuli

Publication Date

2023

DOI

10.5070/BS327262047

Copyright Information

Undergraduate

Growing a Mind with Statistical Inference

BY ANJULI NIYOGI

As we go about our daily lives, our minds are constantly building concepts and formulating theories about the world around us. These mental processes often rely on making inferences drawn from limited data. The process of statistical inference, which involves drawing conclusions from data, provides a scientific basis for this cognitive process.

In the early 2000s, Joshua Tenenbaum, a professor at MIT, introduced a new approach to understanding this theory of learning. According to Tenenbaum, humans learn and reason using three components: statistical inference, abstraction (the ability to generalize with limited data based on commonalities), and structured representations (methods of mentally organizing information).[1] These processes allow us to perform sophisticated tasks and make decisions every day.

Statistical Inference

The heart of statistical inference relies on conditional probability, which is the likelihood of an event occurring given another event has already taken place. To illustrate this, suppose you roll two fair six-sided dice. What is the probability of the sum of the two dice totaling 12? Conversely, what is the probability, given that the first roll was a six? As you can see, the probability changes significantly from 1/36 to 1/6 once you are given the condition of the first roll.

The probability of an event like this can be estimated using Bayes' rule, a mathematical formula named after statistician and philosopher Thomas Bayes. Bayes' rule enables the estimation of the probability by taking into account prior knowledge about its likelihood. For instance, one could use Bayes' rule to determine the probability of a dice roll, given the result of a previous roll.

$$P(\text{the person is unfriendly} \mid \text{the person is shy}) = \frac{P(\text{a person is shy} \mid \text{unfriendly}) * P(\text{an unfriendly person})}{P(\text{a person is shy})}$$

Figure 1: Bayes Rule indicates the conditional probability that someone you may encounter is unfriendly, given that they were shy when you met them.

To illustrate this, consider the following example modeled after a thought experiment presented in the award-winning book *Thinking Fast and Slow* by Daniel Kahneman. Imagine you meet someone who is very shy and withdrawn. Your natural initial impression might be that they are unfriendly or unapproachable. However, if you know from prior experience that shy people are often friendly once you get to know them, you can adjust your initial impression using Bayes' rule. In this case, the formula is used to update your impression of the person based on your more generalized, abstract knowledge about the behavior of shy people, the limited amount of data you observed, and your prior beliefs (see Fig. 1).[2]

### Incorporating Abstract Knowledge

While statistical learning plays a crucial role in our creation of world models, we are also constantly taking in, updating, and using abstract knowledge in our everyday lives. Developing abstract knowledge is an iterative process that often involves incorporating new information and experiences to refine previously constructed models. This process relies on both top-down and bottom-up processes, where top-down processes use previously acquired knowledge and structures to guide the learning of new information, and bottom-up processes involve constructing new structures and knowledge from sensory input and experiences.[3]

To demonstrate the use of both abstract knowledge and conditional probability (see Fig. 2), consider the following scenario: imagine you see an animal on the street with short legs, a lengthy body, and a droopy mouth, while on a leash. You know from general abstract principles that because it is on a leash and walking around your neighborhood street, it is highly likely to be a dog. Now let's say you're wondering what breed of dog it is. You have a structured probabilistic model about what breed this dog might be. It has short legs, so it's likely a dachshund, a basset hound, or a corgi. You've now narrowed it down based on your model. Finally, you notice it has unusually long ears. Based on conditional probability, you can deduce that it must be a basset hound since you know that dachshunds and corgis have short ears. You've got it—this dog is a basset hound.

Knowledge Representations

The representations we form using
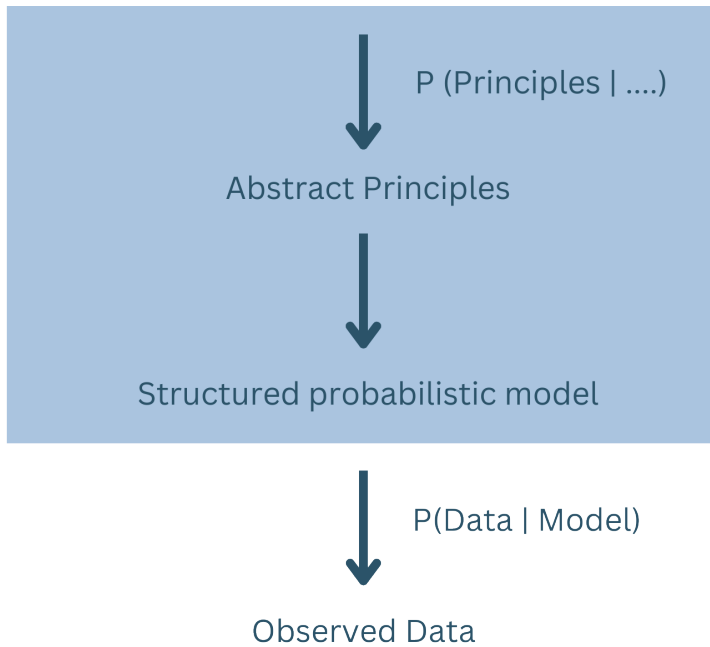
Figure 2: Theory-based induction can be used to explain observed data; a learner may generate their hypotheses given abstract principles and structured probabilistic models.
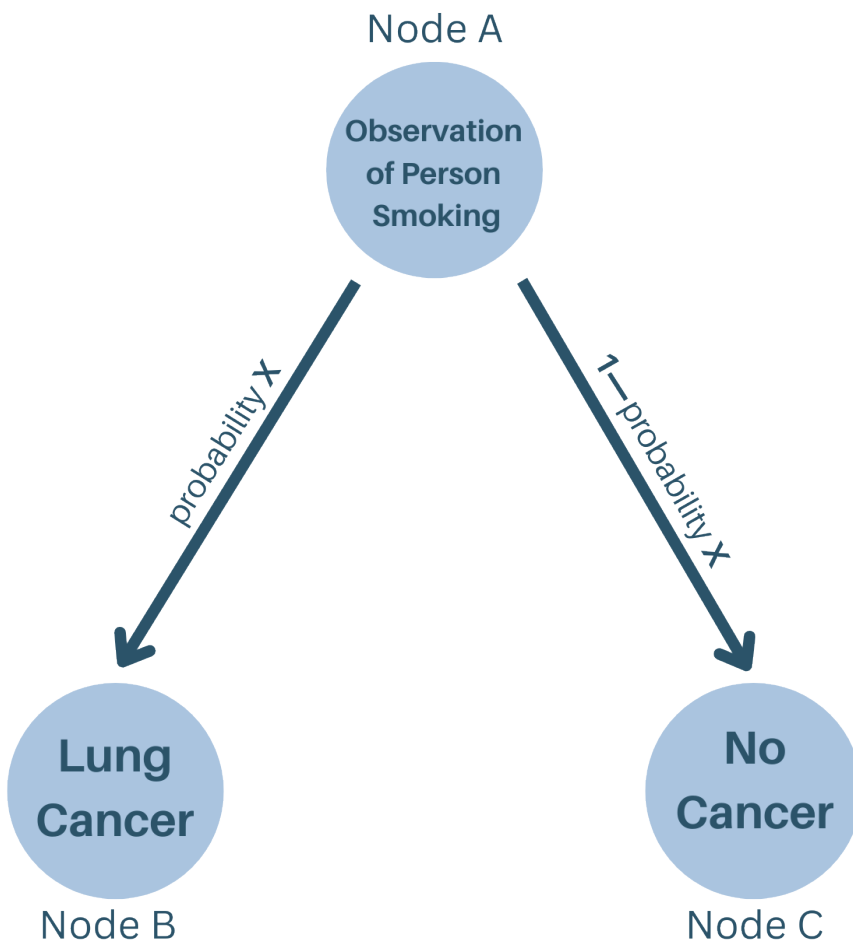
Figure 3: Tree representations can be used to describe the causal relationships between events, with varying probabilities between nodes A to B and A to C.

abstraction and probability can be depicted with models commonly used in the field of computer science that incorporate our understanding of causal relationships. These models include graphs, which visualize relationships between data, and conditional inference trees, which show conditional relationships between variables. The choice of representation is dependent on the knowledge being captured and other levels of abstraction.

One of the most common causal structures is a graph representation. A directed acyclic graph is a graphical model that incorporates events or states of the world (points on the graph) with directed edges, which represent the causal relationships between those variables. The individual building blocks of the structure are referred to as nodes, and the directed edges show the causal structure of an event. For example, consider event A; we know that A can cause either event B or C, but the relationship may not work in the other direction. Since only event A can cause B or C, but neither B nor C can cause A, this structure is directional by nature and lacks any cycles between nodes (see Fig. 3). The exact structure of a causal graph depicts the probability of the variables in that graph, particularly the conditional dependencies of those variables.[4]

Trees are one of the most widely used graph structures and are built with a sequence of nodes; a root node on top is connected with as many layers of nodes as necessary to capture the information represented until the bottom layer is reached. The simplest version of a tree may look like a singular root node with two nodes connecting it, similar to children in an ancestral family tree.

In this context, the nodes of the tree can represent events or decisions, while the edges lead to possible outcomes. For example, suppose you notice a person nearby smoking a cigarette. We can build a tree representation that models the probability of this person being diagnosed with lung cancer based on this observation (see Fig. 3). The observation of them smoking can be represented by the root node, and the numerical probabilities of them getting lung cancer in the future or not can be represented by the left and right branches connected to the two child nodes, respectively.

Incorporating wide amounts of abstract

knowledge with more than just one type of condition can lead to a much more complex tree representation. This complexity arises when we need to consider multiple factors that influence the probability of an event or decision. In a medical diagnosis, for instance, we may need to consider not only the presence of a symptom but also the patient's medical history, lifestyle, and genetics. Such complex decision trees can be built using Bayesian networks, which allow for the representation of multiple dependencies and conditional probabilities.[5] These networks have been used in various applications, such as predicting the risk of heart disease and analyzing the effectiveness of different marketing strategies.[6, 7]

Using other kinds of representations, such as two-dimensional chains, may be better for capturing information with different properties. A single chain may represent a linear relationship of a certain characteristic changing across different variables. Combining this with another chain may add to the level of detail you have. For instance, suppose you are at a restaurant and want to imagine what something will taste like. There are two dimensions you can model this–how savory or sweet something is, as well as its spice level. In this case, two linearly sequenced chains create a clearer understanding of what each menu item will taste like. The mango salsa may be a sweet and spicy combo, whereas the carne asada may be a rather savory, but mild choice. This learning may result from observations of previous foods you've eaten and your prediction of an item's taste qualities may result from the model that has been created.

Bayesian models allow us to represent more complex probabilistic models that reflect our beliefs over all possible outcomes, events, and properties.[8] These models usually use some form of a graph structure, with numerical parameters to capture the quantitative details of the information. The numerical parameters are based on our previous knowledge and observations, as well as upcoming data through Bayesian updating. Above this level of probabilistic models are more abstract principles based on our previous knowledge and observations. Inferences involve using pre-existing models to infer unobserved aspects of the data. This process can work both ways: developing models with abstract principles and theories, or learning abstract principles given data and theories.

Applications in Developmental Psychology & Beyond

UC Berkeley Professor of Psychology Alison Gopnik* has researched extensively on the ties between adolescent learning and building causal structures. Children essentially infer causal structures from statistical information through their interactions with the environment and observations of others.[9] In an experiment with young children, two-and-a-half-year-olds could discriminate between conditional independence and dependence. That is, after viewing various combinations of objects that either did or did not light up when placed on a machine, they could use that information to make judgments about which objects caused the machine to light up.[10] By the age of 18 months, children can also infer the likelihood of linguistic syllables in an artificial language and even do the same with visual stimuli.[11] While children learn how to do this from an early age, adults continue to build, challenge, and update these structures to guide decision-making processes.

In the early 2010s, Bayesian inference provided a novel perspective on human cognition. Applications have been developed across many different fields in cognitive science, including linguistics, developmental psychology, and even medicine.[12] A better understanding of how humans learn and reason has also led to significant advancements in artificial intelligence, as seen in more recent models like chat GPT, which rely on a similar framework of statistical inference.[13] While the potential of Bayesian inference has already been proven, it is only the beginning of what is to come.

REFERENCES

1. Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. Science, 331, 1279--1285. doi: 10.1126/science.1192788

2. Kahneman, Daniel. (2011). Thinking, Fast and Slow. Penguin UK.

3. Klahr, D., & Dunbar, K. (1988). Dual Space Search During Scientific Reasoning. Cognitive Science, 12(1), 1–48. https://doi.org/10.1016/0364-0213(88)90007-9.

4. Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. Trends in Cognitive Sciences, 8(8), 371–377. https://doi.org/10.1016/j.tics.2004.06.005

5. Koller, D. K., & Friedman, N. F. (2009). Probabilistic Graphical Models Principles and Techniques. The MIT Press.

6. Ordovas, J.M. & Rios, David & Santos-Lozano, Alejandro & Lucia, Alejandro & Torres-Barrán, Alberto & Kosgodagan, Alex & Camacho Rodríguez, Jose Manuel. (2023). A Bayesian network model for predicting cardiovascular risk. Computer Methods and Programs in Biomedicine. 231. 107405. 10.1016/j.cmpb.2023.107405.

7. Karakaya, Ç., Badur, B., & Aytekin, C. (2011). Analyzing the Effectiveness of Marketing Strategies in the Presence of Word of Mouth: Agent-Based Modeling Approach. Journal of Marketing Research & Case Studies, 1–17. https://doi.org/10.5171/2011.421059

8. Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. C. (2006). Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Sciences, 10(7), 309–318. https://doi.org/10.1016/j.tics.2006.05.009

9. Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. Psychological Bulletin. https://doi.org/10.1037/a0028044

10. Gopnik, A., Sobel, D. M., Schulz, L., &

In our Spring 2022 issue of Flux, we featured Professor Alison Gopnik in an interview about the course "Sense and Sensibility and Science," for which she is a member of the instructional team.

Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. Developmental Psychology, 37(5), 620–629. https://doi.org/10.1037/0012-1649.37.5.620

11. Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. Science, 274(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

12. Spiegelhalter, D. J., Myles, J. P., Jones, D. R., & Abrams, K. R. (1999). Methods in health service research. An introduction to bayesian methods in health technology assessment. BMJ (Clinical research ed.), 319(7208), 508–512. https://doi.org/10.1136/bmj.319.7208.508

13. What Is ChatGPT Doing . . . and Why Does It Work?—Stephen Wolfram Writings. (2023, February 14). Retrieved April 2, 2023, from https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/

IMAGE REFERENCES

1. Figure 1: Niyogi, Anjuli. Bayes' Rule. 2023.

2. Figure 2: Adapted from Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. C. (2006). Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Sciences, 10(7), 309–318. https://doi.org/10.1016/j.tics.2006.05.009

3. Figure 3: Niyogi, Anjuli. Tree Representations. 2023.

FEATURES