Brain and Cognition 112 (2017) 98–101

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Types of approximation for probabilistic cognition: Sampling and variational

Adam N. Sanborn

Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom

ARTICLE INFO

Article history:
Received 2 June 2015
Accepted 17 June 2015
Available online 27 July 2015

Keywords:
Probabilistic cognition
Rational process models
Sampling
Variational approximations

ABSTRACT

A basic challenge for probabilistic models of cognition is explaining how probabilistically correct solutions are approximated by the limited brain, and how to explain mismatches with human behavior. An emerging approach to solving this problem is to use the same approximation algorithms that were been developed in computer science and statistics for working with complex probabilistic models. Two types of approximation algorithms have been used for this purpose: sampling algorithms, such as importance sampling and Markov chain Monte Carlo, and variational algorithms, such as mean-field approximations and assumed density filtering. Here I briefly review this work, outlining how the algorithms work, how they can explain behavioral biases, and how they might be implemented in the brain. There are characteristic differences between how these two types of approximation are applied in brain and behavior, which points to how they could be combined in future research.

1. Introduction

Probabilistic cognition is a natural fit to the kind of problems posed by the environment: people are faced with noisy and ambiguous observations about the world, yet need to make good decisions. Probabilistic models allow for uncertainty and ambiguity to be dealt with appropriately, because instead of incorrectly assuming that imperfect information is known perfectly, these models can find the best possible action given that imperfect information.

These models have had broad success in explaining human data, accounting for how people are aware of their perceptual uncertainty and combine it appropriately with prior knowledge (Körding & Wolpert, 2004; Tassinari, Hudson, & Landy, 2006), and explaining how people can learn to represent an ambiguous environment in cognitive tasks (Griffiths, Steyvers, & Tenenbaum, 2007; Kemp & Tenenbaum, 2008). However despite these successes, probabilistic models have faced skepticism from two major sources: evidence of mismatches between human behavior and probabilistic cognition (Tversky & Kahneman, 1978), and the inherent computational complexity of these models. It just does not seem like we as humans can do the complex calculations necessary to arrive at the best answers, and so there must be shortcuts involved (Anderson, 1991; Simon, 1955; Van Rooij, 2008).

Fortunately the problem of working with complex probabilistic models in limited systems has received a lot of attention from computer scientists and statisticians. Researchers in these fields have developed algorithms that arrive at good solutions while minimizing computational and memory requirements. These algorithms then provide an interesting alternative to extant heuristics in psychology and neuroscience, and in cognitive science using these algorithms to explain behavior has been termed *rational process models* (Sanborn, Griffiths, & Navarro, 2010). The advantage of this approach is that when these algorithms are used in situations for which they are well-adapted, they make probabilistic cognition achievable, but when they are applied to situations for which they are poorly adapted, they can explain biases in behavior that cannot be explained by probabilistic models alone.

Computer scientists and statisticians have developed various types of approximations for probabilistic models, such as Laplace's method, sampling algorithms, variational approximations, and expectation propagation (Bishop, 2006; Doucet, de Freitas, & Gordon, 2001; Minka, 2001; Neal, 1993; Wainwright & Jordan, 2008). Here I focus on the two types that have been applied to approximate probabilistic cognition: sampling and variational approximations. Sampling algorithms are stochastic, randomly drawing samples to represent a probability distribution as a collection of points. While sampling algorithms asymptotically provide the correct answer, they are less accurate and can show biases for small numbers of samples. In contrast, variational algorithms

E-mail address: a.n.sanborn@warwick.ac.uk

http://dx.doi.org/10.1016/j.bandc.2015.06.008
0278-2626/

trade stochastic sampling for deterministic optimization. These algorithms can be very fast, but are asymptotically biased.

Researchers have used both sampling and variational algorithms as approximations to probabilistic cognition in behavior and the brain. However these investigations have tended to proceed separately, with little comparison between the work using the two types of algorithms. Below, I describe examples of both types of algorithms, how they can produce behavioral biases, and how they might be implemented in the brain. A comparison of the two types shows what each is good for, and how they could be profitably combined in future work.

2. Sampling approximations

Sampling algorithms are useful for approximating calculations that involve complex probability distributions because the collection of samples can simply stand in for the complex distribution in a calculation. These approximate calculations are asymptotically correct with an infinite number of samples, but there are generally no guarantees for smaller numbers of samples.

While it is ideal if samples can be drawn from the distribution directly, often this is not the case and more sophisticated methods are required. One commonly used variety of sampling is *importance sampling*, which avoids the problem of drawing samples directly from a complex distribution by first sampling from a similar but simpler distribution (Bishop, 2006). These samples are weighted so that they reflect the probability of the complex distribution and not the actual distribution from which they were drawn. Importance sampling works well when the simpler distribution is very similar to the complex distribution, but is inaccurate if these distributions are very different.

A generalization of importance sampling is *particle filtering* (Doucet et al., 2001). This algorithm extends importance sampling into sequential tasks in which decisions need to be made after each observation of data. The simplest version of particle filtering draws samples from the prior distribution and sequentially reweights these samples by the likelihood of the data as it is observed. However, this version of particle filtering quickly runs into trouble because it is likely that the weight for one sample will dominate all of the rest, effectively yielding only a single sample. More sophisticated particle filters add steps such as replacing the worst samples with better-performing samples or perturbing the samples to provide a better overall approximation.

Another commonly used sampling algorithm is Markov chain Monte Carlo (*MCMC*; Neal, 1993). MCMC starts at a particular set of values (the initial state) for each of the random variables and makes a series of stochastic transitions to new states. By clever choice of the transition function, the series of states produced are samples from the distribution of interest. The strength of MCMC is that not as much needs to be known about the complex distribution ahead of time, but some downsides are that the initial samples need to be discarded and that samples are autocorrelated: because most MCMC samplers preferentially transition to nearby states, transitions between far-apart states are slower.

2.1. Explaining behavioral biases

Importance sampling, particle filtering, and MCMC have all been used to explain biases in human behavior. Importance sampling has been formally linked to exemplar models, which are well-supported models of memory and categorization. This link generalizes exemplar models to new tasks and allows it to explain behavioral biases. For example, in reproduction tasks participants' responses are drawn toward the distribution of stimuli they have previously been shown. The form of the assimilative effect shows deviations from what probabilistic models predict, but these deviations can be explained by assuming participants use a restricted number of samples (Shi, Griffiths, Feldman, & Sanborn, 2010).

Particle filters have been used to explain human biases in a variety of sequential tasks. Because repeated reweighting effectively reduces the number of samples, particle filters are useful for explaining how behavior can be more strongly influenced by early than late observations: samples consistent with the early observations initially dominate, and for some types of particle filter this makes it impossible to draw samples consistent with the late observations. Particle filters have been used to explain how early observations can dominate in categorization (Sanborn et al., 2010), sentence processing (Levy, Reali, & Griffiths, 2009), and causal learning (Abbott & Griffiths, 2011). Particle filters have also been used to explain individual variability around the group mean in learning (Daw & Courville, 2008) and change point detection (Brown & Steyvers, 2009).

MCMC has been used to explain different kinds of behavioral biases. Samples generated by MCMC are autocorrelated, and this property is useful for describing how judgments change slowly over time. One application of this is to bistable perception, where the current percept of a figure can be cast as a sample from a bimodal probability distribution over interpretations, and sampling using MCMC can explain the transition times between percepts (Gershman, Vul, & Tenenbaum, 2012). Autocorrelation also means that MCMC is initially influenced by its start state, which has been used to explain how irrelevant self-generated anchors in reasoning problems can have an effect on later answers (Lieder, Griffiths, & Goodman, 2012).

2.2. Implementation in the brain

Proposals have been made for how each of the above sampling algorithms could be implemented in the brain. For importance sampling, Shi and Griffiths (2009) proposed that neural tuning curves were proportional to the likelihood and that the number of neurons with a particular tuning curve were proportional to the prior. This scheme was extended to perform inference in a hierarchical model, which the levels of the model mapped to hierarchically organized brain regions.

Lee and Mumford (2003) used a similar global organization, proposing that at each level in the cortical hierarchy probabilistic cognition was implemented with a particle filter. Messages were then passed between the levels so that the top-down effects of context and the bottom-up effects of the stimulus were both incorporated. More detailed neural implementation of particle filters are given by Huang and Rao (2014) and Legenstein and Maass (2014) using networks of spiking neurons.

Other researchers have described on how populations of neurons could implement MCMC. In these implementations, the state of the brain corresponds to a sample from a probability distribution and transitions between neural states correspond to the transitions that the MCMC algorithm makes (Fiser, Berkes, Orbán, & Lengyel, 2010). Currently there are separate kinds of MCMC implementations for sampling from continuous variables (Hennequin, Aitchison, & Lengyel, 2014; Moreno-Bote, Knill, & Pouget, 2011) and sampling from discrete variables (Buesing, Bill, Nessler, & Maass, 2011; Probst et al., 2015).

3. Variational approximations

Variational approximations are a second major type of approximation in computer science and statistics, and these algorithms trade the stochasticity of sampling for the determinism of optimization. Variational algorithms work by first defining a simpler