Posterior Inference and Soft Value Guidance in Sequential Models

Fine-tuning, controlled generation, and sampling in sequential models has attracted a flurry of recent attention in a variety of settings, particularly with the growing availability of powerful open-source pretrained models. For language modeling in discrete spaces, we would often like to align responses with human preferences or generate correct responses to complex reasoning questions. For diffusion models, we may be interested in steering generation to produce samples belonging a certain class, images which score highly on metrics such as realism, preference alignment, or text-to-image consistency, and proteins or molecules with desired properties such as synthesizability. Diffusion-based methods have also been applied for sampling from arbitrary target probability densities such as Boltzmann distributions, where we can only assume access to a unnormalized density or energy function.

In this blog post, we provide overview of these sampling or controlled generation tasks from a probabilistic perspective, which incorporates notions from soft reinforcement learning, stochastic optimal control, and Sequential Monte Carlo. A key role will be played by the soft value function, which yields both importance sampling weights and gradient guidance for diffusion processes. This perspective gives a single conceptual framework for guidance in discrete and continuous spaces, and highlights how methodologies can be shared across problem settings.

AUTHORS
Anonymous

AFFILIATIONS

PUBLISHED
April 28, 2025

Contents

Setting & Notation

Assume we are given a pretrained model p^{ref}, which we will seek to condition or modulate to achieve some target properties or distribution at the endpoint. We begin by defining shared notation which will encompass both the language and diffusion modeling settings, although the reader should feel free to skip ahead to concrete examples in Target Distributions and parse the notation within this context.

For autoregressive language models, we adapt our notation to provide a Markovian structure used throughout later exposition. We consider the state $\mathbf{x}_t = \mathbf{concat}(\mathbf{x}_0, x_1, x_2, \dots x_t) \in \mathcal{V}^{T_0+t}$ in an expanding state-space of tokens $x_\tau \in \mathcal{V}$ from a discrete vocabulary, which are generated in response to a prompt or initial state $\mathbf{x}_0 \in \mathcal{V}^{T_0}$ of maximum length T_0. We view a reference policy $p^{\text{ref}}_{\text{LM}}(a_t = x_{t+1}|\mathbf{x}_t)$ as selecting a next token x_{t+1} as the action a_t with the context $\mathbf{x}_t$ as the state , with deterministic environment transitions $p^{\text{env}}(\mathbf{x}_{t+1}|a_t = x_{t+1}, \mathbf{x}_t) = \mathbb{I}[\mathbf{x}_{t+1} = \mathbf{concat}(\mathbf{x}_t, x_{t+1})]$ that concatenate the generated token x_{t+1} with the context $\mathbf{x}_t$. The policy is usually given by an autoregressive model $\mathbf{x}_t \sim \prod_{\tau=0}^{t-1} p^{\text{ref}}_{\text{LM}}(x_{\tau+1}|\mathbf{x}_\tau)$. For convenience, we will write the full state transition as $p^{\text{ref}}_{t+1}(\mathbf{x}_{t+1}|\mathbf{x}_t) = p^{\text{ref}}_{\text{LM}}(x_{t+1}|\mathbf{x}_t)\mathbb{I}[\mathbf{x}_{t+1} = \mathbf{concat}(\mathbf{x}_t, x_{t+1})]$. This leads to a slight abuse of notation in which we can write the probability of a (partial) sequence $\mathbf{x}_t$ either using tokens $p^{\text{ref}}_t(\mathbf{x}_t) = \prod_{\tau=0}^{t-1} p^{\text{ref}}_{\text{LM}}(x_{\tau+1}|\mathbf{x}_\tau)$ or as a joint distribution over its prefixes $p^{\text{ref}}_t(\mathbf{x}_{0:t}) = \prod_{\tau=0}^{t-1} p^{\text{ref}}_{\tau+1}(\mathbf{x}_{\tau+1}|\mathbf{x}_\tau)$.

For diffusion processes, let $\mathbf{x}_t \in \mathbb{R}^d$ represent the current (noisy) state, where $\mathbf{x}_T$ corresponds to clean data. [1] We consider a reference stochastic differential equation with time-dependent drift b^{ref}_t , which may correspond to a physical force or pretrained score-based diffusion model

$$P^{\text{ref}}: \qquad d\mathbf{x}_t = b^{\text{ref}}_t(\mathbf{x}_t)dt + \sigma_t dW_t \qquad \mathbf{x}_0 \sim p^{\text{ref}}_0 \tag{1}$$

We can approximately model this continuous-time stochastic processes using discrete-time Gaussian kernels for small dt . We consider the reference drift as an action $a_t = b^{\text{ref}}_t(\mathbf{x}_t, t)$, with stochastic environment transitions drawn from $p^{\text{env}}(\mathbf{x}_{t+1}|a_t = b^{\text{ref}}_t(\mathbf{x}_t, t), \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t+1}; \mathbf{x}_t + b^{\text{ref}}_t(\mathbf{x}_t)dt, \sigma_t \mathbb{I}_d)$ via Euler discretization. For convenience, we combine action selection and state transition into the policy $p^{\text{ref}}_{t+1}(\mathbf{x}_{t+1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t+1}; \mathbf{x}_t + b^{\text{ref}}_t(\mathbf{x}_t)dt, \sigma_t \mathbb{I}_d)$.

Target Distributions

We will proceed to view many controlled generation or fine-tuning tasks as sampling from a target probability distribution at the final step T, where the target is only known up to a normalization constant. [1, 2, 3, 4].

To ease notation and facilitate posterior sampling interpretations, we define an observation random variable $\mathbf{y}$, which is emitted as a function of the final state according to $p(\mathbf{y}|\mathbf{x}_T)$, and attempt to sample from the posterior distribution over all states,

$$p^*(\mathbf{x}_{0:T}|\mathbf{y}) = \frac{1}{\mathcal{Z}^{\mathbf{y}}} p^{\text{ref}}(\mathbf{x}_{0:T}) p(\mathbf{y}|\mathbf{x}_T) \qquad \mathcal{Z}^{\mathbf{y}} = \int p^{\text{ref}}(\mathbf{x}_{0:T}) p(\mathbf{y}|\mathbf{x}_T) d\mathbf{x}_{0:T} \tag{2}$$

In particular, we would like our full language model responses or final diffusion states to be distributed according to the endpoint posterior marginal $p^*(\mathbf{x}_T|\mathbf{y})$. We will consider a flexible class of possible target posteriors defined in the following table.

Setting	$p(\mathbf{y}\|\mathbf{x}_T)$	$p^*(\mathbf{x}_T\|\mathbf{y})$
Constraint	$\mathbb{I}[\mathbf{x}_T \in \mathcal{B}]$	$\frac{1}{\mathcal{Z}^{\mathcal{B}}} p^{\text{ref}}(\mathbf{x}_T) \mathbb{I}[\mathbf{x}_T \in \mathcal{B}]$
Classifier or Observation	$p(\mathbf{y}\|\mathbf{x}_T)$	$\frac{1}{\mathcal{Z}^{\mathbf{y}}} p^{\text{ref}}(\mathbf{x}_T) p(\mathbf{y}\|\mathbf{x}_T)$
Reward or Energy Modulation	$\frac{1}{M} \exp\{\beta\, r(\mathbf{x}_T)\}$	$\frac{1}{\mathcal{Z}^{\beta r}} p^{\text{ref}}(\mathbf{x}_T) \exp\{\beta\, r(\mathbf{x}_T)\}$
Arbitrary Unnormalized Density	$\frac{1}{M} \frac{\tilde{\pi}_T(\mathbf{x}_T)}{p^{\text{ref}}(\mathbf{x}_T)}$	$\frac{1}{\mathcal{Z}} \tilde{\pi}_T(\mathbf{x}_T)$

A crucial challenge arises from the fact that conditioning information is only provided at the terminal state $\mathbf{x}_T$, whereas generation or sampling needs to be performed sequentially and forward in time according to

$$p^*(\mathbf{x}_{0:T}|\mathbf{y}) = p^*(\mathbf{x}_0|\mathbf{y}) \prod_{t=1}^{T} p^*(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}) \tag{3}$$

Before describing how soft value functions and stochastic optimal control can be used to address this challenge, we discuss several concrete examples below.

Examples

CONSTRAINTS

For the language modeling setting, constraints may filter responses which correspond to correct answers to reasoning questions [5], syntactically-valid outputs, or responses for which a scalar function meets an acceptability or rare-event threshold $\mathbb{I}[f(\mathbf{x}_T) \leq c]$.

For diffusion modeling, constraining the endpoint sample to fall within a certain set $\mathbb{I}[\mathbf{x}_T \in \mathcal{B}]$ corresponds to the traditional formulation of Doob's h-transform, which has been used for generative modeling on constrained domains [6] or with aligned data [7, 8] arising in biomolecular or chemical problems. In the case where p^{ref} is a diffusion with linear drift, the conditioned process ending at a particular point $\mathbb{I}[\mathbf{x}_T = \mathbf{x}]$ is available as a closed form linear interpolation. This observation underlies efficient optimization techniques for `bridge matching' methods [9, 10] which extend rectified flow matching [11, 12] to stochastic processes and Schrödinger Bridge problems for generative modeling or image translation.

CLASSIFICATION OR OBSERVATION RANDOM VARIABLES

Given a classifier $p(\mathbf{y} = c|\mathbf{x}_T)$, we can hope to condition our language or diffusion model to generate samples likely to be of a certain class, such as uncovering language model responses which are flagged by content moderation classifiers. In the Stochastic Optimal Control section below, we will see that class-conditioned diffusion processes characterize the optimal form of well-known classifier(-free) guidance techniques [13]. Finally, conditioning on a noisy observation $\mathbf{y} = \mathcal{A}(\mathbf{x}_T) + \epsilon$ finds extensive applications for solving inverse problems in imaging [14, 15, 16, 17].

REWARD OR ENERGY MODULATION

Reinforcement learning from human feedback has become a dominant paradigm for aligning pretrained language models with human preferences or task-specific applications [18], finetuning diffusion models to align with text prompts or user feedback [19], or generating proteins, molecules, or genetic sequences with particular properties such as stability, synthesizability, or downstream effectiveness . For our purposes, we will assume a reward model is given.

GENERAL UNNORMALIZED TARGET DENSITIES

Most generally, we can seek to sample from a given unnormalized target density $\tilde{\pi}_T(\mathbf{x}_T)$ over the final state, which includes reward modulation as a special case $\tilde{\pi}_T(\mathbf{x}_T) = p^{\text{ref}}(\mathbf{x}_T) \exp\{\beta\, r(\mathbf{x}_T)\}$. To facilitate a posterior interpretation in these cases, we would like to introduce a random variable $\mathbf{y}$ which reflects `optimality', or the fact that endpoint samples are distributed according to the endpoint target. To do so, we construct a hypothetical rejection sampling of the endpoint samples, where we accept samples with probability $p(\mathbf{y} = 1|\mathbf{x}_T) = \frac{1}{M} \frac{\tilde{\pi}_T(\mathbf{x}_T)}{p^{\text{ref}}(\mathbf{x}_T)}$, for $M = \max_{\mathbf{x}_T} \frac{\tilde{\pi}_T(\mathbf{x}_T)}{p^{\text{ref}}(\mathbf{x}_T)}$. The constant M, which ensures $p(\mathbf{y} = 1|\mathbf{x}_T) \leq 1$ and that accepted samples have the desired distribution, need not be estimated in practice, since it can be shown to vanish in the eventual posterior $p^*(\mathbf{x}_T|\mathbf{y})$.

Again, we emphasize that this construction is hypothetical, but is useful to add detail to presentation in the influential 2018 tutorial by Sergey Levine [20] and facilitate our unified viewpoint in terms of posterior inference.

Initial Sampling

An immediate question arises as to how to initialize sampling in (3), since $p^*(\mathbf{x}_0|\mathbf{y})$ is already likely to be intractable in general.

In language modeling settings, we are often given access to prompts $\mathbf{x}_0$ via data or user interaction, so it is natural to focus on the posterior over responses to particular prompts,

$$p^*(\mathbf{x}_{1:T}|\mathbf{x}_0, \mathbf{y}) = \frac{1}{\mathcal{Z}_0^{\mathbf{y}}(\mathbf{x}_0)} p^{\text{ref}}(\mathbf{x}_{1:T}|\mathbf{x}_0) p(\mathbf{y}|\mathbf{x}_T) \tag{4}$$
$$\mathcal{Z}_0^{\mathbf{y}}(\mathbf{x}_0) = \int p^{\text{ref}}(\mathbf{x}_{1:T}|\mathbf{x}_0) p(\mathbf{y}|\mathbf{x}_T) d\mathbf{x}_{1:T}$$

However, in diffusion models, we remain interested in $p^*(\mathbf{x}_{0:T}|\mathbf{y})$, and risk introducing bias if our initial sampling distribution differs from $p^*(\mathbf{x}_0|\mathbf{y})$. It may be possible to sample from $p^*(\mathbf{x}_0|\mathbf{y}) \approx p^{\text{ref}}(\mathbf{x}_0)$ in cases when the noising dynamics converge quickly to a stationary distribution, such as a standard Normal, independent of the initial distribution [16]. Alternatively, finetuning could be performed using a `memoryless' noise schedule which renders $p^{\text{ref}}(\mathbf{x}_T|\mathbf{x}_0) = p^{\text{ref}}(\mathbf{x}_T)$ and thus $p^*(\mathbf{x}_0|\mathbf{y}) = p^{\text{ref}}(\mathbf{x}_0)$ [19]. We proceed to assume $\mathbf{x}_0 \sim p^*(\mathbf{x}_0|\mathbf{y})$ and focus on subsequent sampling steps for $p^*(\mathbf{x}_{1:T}|\mathbf{x}_0, \mathbf{y})$ to encompass both language and diffusion settings.

Soft Value Function

We begin by characterizing the target posterior $p^*(\mathbf{x}_{1:T}|\mathbf{x}_0, \mathbf{y})$ via the solution to a variational optimization [21], which we will refer to as an Evidence Lower Bound (ELBO),

$$\begin{aligned} V_0^{\mathbf{y}}(\mathbf{x}_0) &= \max_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\big[\log p(\mathbf{y}|\mathbf{x}_T)\big] - D_{KL}\big[q(\mathbf{x}_{1:T}|\mathbf{x}_0) : p^{\text{ref}}(\mathbf{x}_{1:T}|\mathbf{x}_0)\big] \\ &= \log \mathcal{Z}_0^{\mathbf{y}}(\mathbf{x}_0) \end{aligned} \tag{5}$$

where $q(\mathbf{x}_{1:T}|\mathbf{x}_0) = p^*(\mathbf{x}_{1:T}|\mathbf{x}_0, \mathbf{y})$ achieves the maximum and the soft value is the log normalization constant $V_0^{\mathbf{y}}(\mathbf{x}_0) = \log \mathcal{Z}_0^{\mathbf{y}}(\mathbf{x}_0)$, which we expand upon below.

The optimal *soft value function* thus translates terminal target information to intermediate steps in order to facilitate sampling the exact posterior marginals along the entire trajectory. In particular, consider the optimization (5) starting from a given partial sequence or intermediate state $\mathbf{x}_t$,

$$V_t^{\mathbf{y}}(\mathbf{x}_t) = \max_{q(\mathbf{x}_{t+1:T}|\mathbf{x}_t)} \mathbb{E}_{q(\mathbf{x}_{t+1:T}|\mathbf{x}_t)}\big[\log p(\mathbf{y}|\mathbf{x}_T)\big] - D_{KL}\big[q(\mathbf{x}_{t+1:T}|\mathbf{x}_t) : p^{\text{ref}}(\mathbf{x}_{t+1:T}|\mathbf{x}_t)\big] \tag{6}$$
$$= \log \int p^{\text{ref}}(\mathbf{x}_{t+1:T}|\mathbf{x}_t) p(\mathbf{y}|\mathbf{x}_T) d\mathbf{x}_{t+1:T} \tag{7}$$
$$= \log p^*(\mathbf{y}|\mathbf{x}_t) \tag{8}$$

The soft value function measures the expected target likelihood under rollouts from the reference policy, which may involve generating tokens $x_{t+1:T}$ or running diffusion sampling until time T. In our setting with no intermediate reward or target information, we can recognize the expression for $V_{\mathbf{y}}^*(\mathbf{x}_t)$ in (7) as a conditional likelihood in (8) [2]

ONE-STEP OPTIMAL POLICY

Similarly, we can write the optimal one-step sampling distributions in terms of soft values,

$$\begin{aligned} p^*(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}) &= p^{\text{ref}}(\mathbf{x}_t|\mathbf{x}_{t-1}) \frac{p^*(\mathbf{y}|\mathbf{x}_t)}{p^*(\mathbf{y}|\mathbf{x}_{t-1})} \\ &= p^{\text{ref}}(\mathbf{x}_t|\mathbf{x}_{t-1}) \exp\{V_t^{\mathbf{y}}(\mathbf{x}_t) - V_{t-1}^{\mathbf{y}}(\mathbf{x}_{t-1})\} \end{aligned} \tag{9}$$

where $V_{t-1}^{\mathbf{y}}(\mathbf{x}_{t-1}) = \log \mathcal{Z}_{t-1}^{\mathbf{y}}(\mathbf{x}_{t-1})$ again serves as the log normalization constant.

INTERMEDIATE MARGINAL DISTRIBUTIONS

Finally, composing the optimal one-step policies, we can write the evolution of the intermediate target marginals in terms of the value function

$$p_t^*(\mathbf{x}_t|\mathbf{x}_0, \mathbf{y}) = \frac{1}{\mathcal{Z}_0^{\mathbf{y}}(\mathbf{x}_0)} p^{\text{ref}}(\mathbf{x}_t|\mathbf{x}_0) \exp\{V_{\mathbf{y}}^*(\mathbf{x}_t)\} \tag{10}$$

which can equivalently be expressed

$$\log \frac{p^*(\mathbf{x}_t|\mathbf{x}_0, \mathbf{y})}{p^{\text{ref}}(\mathbf{x}_t|\mathbf{x}_0)} = V_t^{\mathbf{y}}(\mathbf{x}_t) - V_0^{\mathbf{y}}(\mathbf{x}_0) = \log p^*(\mathbf{y}|\mathbf{x}_t) - \log \mathcal{Z}_0^{\mathbf{y}}(\mathbf{x}_0) \tag{11}$$

The central message is that the optimal soft value function provides a "backward message" summarizing future conditioning information relevant to sampling at time t.

Stochastic Optimal Control

Remarkably, the gradient of the soft value function can also be shown to provide the optimal drift for a controlled diffusion process guiding samples to the endpoint target distribution.

To build up to this connection, we note that in the continuous-time limit, the KL divergence in (5) is finite only for path measures or SDEs of the form

$$Q^u: \quad d\mathbf{x}_t = \big(b_t^{\text{ref}}(\mathbf{x}_t) + u_t(\mathbf{x}_t)\big)dt + \sigma_t dW_t, \tag{12}$$

where u_t satisfies mild regularity condtiions. In this case, the KL divergence can be written as the time-integral of the norm of u_t using the Girsanov theorem, and we can recognize the negative of the ELBO in (5) as a stochastic optimal control problem

$$-V_0^{\mathbf{y}}(\mathbf{x}_0) = \min_{Q^u(\mathbf{x}_{(0,T]}|\mathbf{x}_0)} \mathbb{E}_{Q^u(\mathbf{x}_{0:T})}\Big[-\log p(\mathbf{y}|\mathbf{x}_T) + \int_0^T \frac{1}{2\sigma_t^2}\|u_t(\mathbf{x}_t)\|^2 dt\Big] \tag{13}$$

subject to Q^u having the form of (12). Using variational calculus, [3] one can show that the solution takes the form

$$u_t(\mathbf{x}_t) = \sigma_t^2 \nabla_{\mathbf{x}_t} V_t^{\mathbf{y}}(\mathbf{x}_t) = \sigma_t^2 \nabla_{\mathbf{x}_t} \log p^*(\mathbf{y}|\mathbf{x}_t) \tag{14}$$

Using the probabilistic view of the value functions in (7)-(8), observe that the exponentiated value functions are related via expectations under the reference process

$$\exp\{V_t^{\mathbf{y}}(\mathbf{x}_t)\} = \mathbb{E}_{p^{\text{ref}}(\mathbf{x}_{t+s}|\mathbf{x}_t)}\left[\exp\{V_{t+s}^{\mathbf{y}}(\mathbf{x}_{t+s})\}\right] \tag{15}$$

This is known as a martingale condition in the stochastic process literature, where $h_t^{\mathbf{y}} = \exp\{V_t^{\mathbf{y}}\}$ is often known as Doob's h-function. The martingale condition ensures that conditional and marginals constructed from (9)-(10) are consistent with respect to marginalization, and results in the following remarkable theorem [23].

Theorem 1 For any function satisfying (15), the stochastic process

$$d\mathbf{x}_t = \left(b_t^{\text{ref}}(\mathbf{x}_t) + \sigma^2 \nabla V_t(\mathbf{x}_t)\right)dt + \sigma_t dW_t \tag{16}$$

realizes the transition dynamics

$$p^V(\mathbf{x}_{t+s}|\mathbf{x}_t) = \frac{\exp\{V_{t+s}(\mathbf{x}_{t+s})\}}{\exp\{V_t(\mathbf{x}_t)\}} p^{\text{ref}}(\mathbf{x}_{t+s}|\mathbf{x}_t) \tag{17}$$

This theorem is true for any function satisfying the martingale condition, including the optimal value function corresponding to a particular target p^*, and demonstrates the link between value functions, guidance drifts for controlled diffusion processes, and posterior or conditioned transition probabilities.

Twisted Sequential Monte Carlo Sampling

In both the language and diffusion cases, we can leverage Sequential Monte Carlo to resample a set of K partial sequences or intermediate states based on the (optimal) soft values, which has the effect of prioritizing sequences or states which we expect to achieve likelihood under the final-step target distribution.

To introduce this importance sampling technique, we consider the unnormalized $\tilde{p}^*(\mathbf{x}_{1:T}|\mathbf{x}_0, \mathbf{y}) = p^{\text{ref}}(\mathbf{x}_{1:T}|\mathbf{x}_0)p(\mathbf{y}|\mathbf{x}_T)$ (see (4)), which omits the intractable normalization constant $\mathcal{Z}_0^{\mathbf{y}}(\mathbf{x}_0)$ and thus is easy to evaluate. For a given proposal or approximate posterior $q(\mathbf{x}_{1:T}|\mathbf{x}_0)$ (which may be learned as in Objectives below, or simply set to p^{ref}), consider the importance weights in the extended space,

$$w_{1:T}(\mathbf{x}_{1:T}) = \frac{\tilde{p}^*(\mathbf{x}_{1:T}|\mathbf{x}_0, \mathbf{y})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}, \qquad \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\left[w_{1:T}(\mathbf{x}_{1:T})\right] = \mathcal{Z}_0^{\mathbf{y}}(\mathbf{x}_0) \tag{18}$$

The latter equality suggests that the weights are an unbiased estimator of the intractable normalization constant $\mathcal{Z}_0^{\mathbf{y}}$, assuming $w_{1:T} < \infty$ for all $\mathbf{x}_{1:T}$.

We would like to transform these weights into step-by-step *incremental* weights which will allow us to perform importance-weighting of intermediate states according to the optimal target posterior. While a naive forward factorization $w_{1:T}(\mathbf{x}_{1:T}) = p(\mathbf{y}|\mathbf{x}_T)\prod_{t=1}^T \frac{p^{\text{ref}}(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|\mathbf{x}_{t-1})}$ would only include target information at the final step, we should instead consider the posterior transitions in (3). Rewriting $p^*(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}) = \frac{p^*(\mathbf{y}|\mathbf{x}_t)}{p^*(\mathbf{y}|\mathbf{x}_{t-1})}p^{\text{ref}}(\mathbf{x}_t|\mathbf{x}_{t-1})$ using (9), we have

$$\begin{aligned} w_{1:T}(\mathbf{x}_{1:T}) &= \prod_{t=1}^T \frac{p^*(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y})}{q(\mathbf{x}_t|\mathbf{x}_{t-1})} \\ &= \prod_{t=1}^T \frac{p^*(\mathbf{y}|\mathbf{x}_t)}{p^*(\mathbf{y}|\mathbf{x}_{t-1})}\frac{p^{\text{ref}}(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|\mathbf{x}_{t-1})} = \prod_{t=1}^T \frac{\exp\{V_t^{\mathbf{y}}(\mathbf{x}_t)\}}{\exp\{V_{t-1}^{\mathbf{y}}(\mathbf{x}_{t-1})\}}\frac{p^{\text{ref}}(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|\mathbf{x}_{t-1})} \end{aligned} \tag{19}$$

Note that the numerator at the final step includes the given target conditional $p(\mathbf{y}|\mathbf{x}_T)$.

The weights in (19) suggest a sequential resampling scheme at intermediate steps. For a budget of K samples and looping over timesteps $1 \le t \le T$, we can proceed with the following steps:

- for $t \in [1, T]$:
  - for $k \in [1, K]$:
    - Sample $\mathbf{x}_t^{(k)} \sim q(\mathbf{x}_t|\mathbf{x}_{t-1}^{(k)})$
    - Update weights $w_{1:t}^{(k)} = w_{1:t-1}^{(k)} \frac{p^*(\mathbf{y}|\mathbf{x}_t^{(k)})}{p^*(\mathbf{y}|\mathbf{x}_{t-1}^{(k)})}\frac{p^{\text{ref}}(\mathbf{x}_t^{(k)}|\mathbf{x}_{t-1}^{(k)})}{q(\mathbf{x}_t^{(k)}|\mathbf{x}_{t-1}^{(k)})}$
  - (if resampling condition met, perform multinomial resampling):
    - Sample $i_k \sim \textbf{categorical}\left(\left\{\frac{w_{1:t}^{(j)}}{\sum_{j=1}^K w_{1:t}^{(\ell)}}\right\}_{j=1}^K\right)$ for $k \in [1, K]$
    - Copy or Reassign Samples: $\mathbf{x}_t^{(k)} \leftarrow \mathbf{x}_t^{(i_k)}$ (for all $k \in [1, K]$ in parallel)
    - Reset weights: $w_{1:t}^{(k)} \leftarrow \frac{1}{K}\sum_{j=1}^K w_{1:t}^{(j)}$

Note that resetting the weights means that only subsequent weights are used for resampling at future timesteps, which preserves the unbiasedness of the eventual weights in (18). See the blog post by Tuan Anh Le for a particularly simple proof [24]. More advanced resampling techniques such as systematic resampling might also be used.

Finally, we can use this resampling scheme even for approximate $V_t^\theta(\mathbf{x}_t)$ or $p^\theta(\mathbf{y}|\mathbf{x}_t)$ for $t < T$, although it is clear that the efficacy of this scheme will depend on the quality of these intermediate value functions or likelihoods.

LANGUAGE

For the language modeling setting, recall that we absorbed the autoregressive model into Markov transitions $p^{\text{ref}}(\mathbf{x}_t|\mathbf{x}_{t-1}) = p^{\text{ref}}_{\text{LM}}(x_t|\mathbf{x}_{t-1})\mathbb{I}[\mathbf{x}_t = \text{concat}(\mathbf{x}_{t-1}, x_t)]$ where the states expand with concatenation of next tokens. Rewriting the proposal in similar terms, we can think of the weights as evolving according to

$$w_{1:T}(\mathbf{x}_{1:T}) = \prod_{t=1}^{T} \frac{p(\mathbf{y}|\mathbf{x}_t)}{p(\mathbf{y}|\mathbf{x}_{t-1})} \frac{p^{\text{ref}}_{\text{LM}}(x_t|\mathbf{x}_{t-1})}{q_{\text{LM}}(x_t|\mathbf{x}_{t-1})} = \prod_{t=1}^{T} \frac{\exp\{V_t^{\mathbf{y}}(\mathbf{x}_t)\}}{\exp\{V_{t-1}^{\mathbf{y}}(\mathbf{x}_{t-1})\}} \frac{p^{\text{ref}}_{\text{LM}}(x_t|\mathbf{x}_{t-1})}{q_{\text{LM}}(x_t|\mathbf{x}_{t-1})}$$

where the likelihood or values are evaluated on the partial sequences $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$. See [25] or [1] for additional discussion.

DIFFUSION

Since diffusion process operate on states $\mathbf{x}_t \in \mathbb{R}^d$ in Markovian fashion, the weights in (19) can be used as is, where $q(\mathbf{x}_t|\mathbf{x}_{t-1})$ corresponds to the discretization of a stochastic process as in (12).

Objective Functions

We finally discuss several classes of objective functions for learning value functions and/or approximate posterior policies. We only attempt to give a high-level landscape of various methods, mostly in discrete time, and defer to references for algorithmic and technical details.

Evidence Lower Bound (Mode-Seeking KL)

Similarly to derivations in the case of standard variational inference, one can show that, for a given q, the gap in the ELBO in (5) is the mode-seeking KL divergence $D_{KL}\left[q(\mathbf{x}_{1:T}|\mathbf{x}_0) : p^*(\mathbf{x}_{1:T}|\mathbf{x}_0, \mathbf{y})\right]$. Thus, minimizing this KL divergence corresponds to maximizing (5). Notably, since $q(\mathbf{x}_{1:T}|\mathbf{x}_0)$ appears in the first argument, optimizing this objective requires taking gradients through the sampling procedure.

LANGUAGE

When $\log p(\mathbf{y}|\mathbf{x}_T) = \beta\, r(\mathbf{x}_T) - \log M$, we recognize (5) as a common objective for reinforcement learning from human feedback in language models, where $q(\mathbf{x}_{1:T}|\mathbf{x}_0)$ is optimized using policy gradient methods such as PPO [18] or REINFORCE [26]. While PPO maintains a value network to reweight policy gradients, the focus is on finetuning a policy $q^{\phi}(\mathbf{x}_{1:T}|\mathbf{x}_0)$, and an optimal policy $q = p^*$ will implicitly capture the value functions through the next-token logits in (9). A similar observation underlies token-wise interpretations of direct preference optimization parameterizations [27]. Nevertheless, learned value functions may also be used to guide generative sampling, either through Monte Carlo Tree Search [28] or Sequential Monte Carlo [1] (as above).

DIFFUSION

Methods for solving stochastic control problems have an extensive history dating back to [29]. Directly solving (13) using backpropagation through trajectories is known as the adjoint method [30, 31], for which improved gradient estimators have been recently proposed in [19]. The adjoint method was used for sampling from general unnormalized target densities in [32].

Cross Entropy (Mass-Covering KL)

While the ELBO and mode-seeking KL divergence was used to introduce the target distribution as the solution of a variational optimization in (5), we can perform optimization using any divergence minimization technique with the desired optimum. One example is to optimize the mass-covering KL divergence as in maximum likelihood training of energy-based models, where recognizing the form of the optimal target marginals in (10), we optimize

$$\min_{\theta} \sum_{t=1}^{T} D_{KL}\left[p^*(\mathbf{x}_{1:t}|\mathbf{x}_0, \mathbf{y}) : p^{\text{ref}}(\mathbf{x}_{1:t}|\mathbf{x}_0)\exp\{V_t^{\theta}(\mathbf{x}_t)\}/\mathcal{Z}_{V^{\theta}}(\mathbf{x}_0)\right] \tag{20}$$

Although exact samples from $p^*(\mathbf{x}_{1:t}|\mathbf{x}_0, \mathbf{y})$ are usually not available, one may use importance sampling approximations to reweight samples according to the endpoint target information $p(\mathbf{y}|\mathbf{x}_T)$, and reuse these weights for approximate sampling at intermediate t. [33, 1]

LANGUAGE

For full-sequence policy optimization, the distributional policy gradient algorithm [34, 35] amounts to optimizing the mass-covering KL at the final step T only, where the energy is parameterized directly via a normalized policy $q^{\phi}(\mathbf{x}_{1:T}|\mathbf{x}_0)$. For learning intermediate value functions, contrastive twist learning [1] optimizes a marginal KL divergence at each step, treating the value functions $V_t^{\mathbf{y}}$ as energies.

DIFFUSION

The contrastive energy prediction objective in [33] amounts to approximate energy-based training of the value functions $V_t^{\mathbf{y}}$ at each step, which can then be used to guide sampling using $\nabla V_t^{\mathbf{y}}$ as a guidance or control drift in (12).

For sampling from a general target density, [4] learn intermediate value functions for guidance and SMC resampling using a `target score matching' loss [36], which, as in the mass-covering KL, requires importance sampling corrections to draw approximate samples from the endpoint target distribution $p^*(\mathbf{x}_T|\mathbf{y})$.

Path Consistency

Path Consistency objectives [37] consider enforcing the first-order optimality conditions associated with the optimization in (5)-(6) using a squared error loss. Since this is a functional equality which should hold everywhere, we can optimize the loss over some off-policy sampling distribution $\pi_s(\mathbf{x}_{1:T}|\mathbf{x}_0)$. Taking the

variation of (5)-(6) with respect to q yields a KKT condition, which we can enforce using

$$\min_{\theta,\phi} \mathbb{E}_{\pi_s(\mathbf{x}_{1:T}|\mathbf{x}_0)}\left[\left(V_0^\theta(\mathbf{x}_0) - \log p(\mathbf{y}|\mathbf{x}_T) + \log \frac{q^\phi(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p^{\text{ref}}(\mathbf{x}_{1:T}|\mathbf{x}_0)}\right)^2\right] \tag{21}$$

This may also be viewed as minimizing the square of the log importance weights between full-sequence forward and reverse processes in (18)-(19). [1] Note that we may also construct one- or c-step consistency losses for any $1 \leq t \leq t+c \leq T$ using the compositional structure of the optimal values in (6)-(10) or decomposition of weights in (19).

LANGUAGE

Path consistency losses correspond to (partial) `trajectory balance' losses in the literature on Generative Flow networks (GFlowNets), and have been applied for inference [38] and finetuning [39] in autoregressive language models.

DIFFUSION

Trajectory balance or path consistency losses can also be applied for inference in diffusions models [40], see also [2]. In the sampling literature, a similar principle underlies the *log-variance* divergences studied in [41, 42], in which we enforce that the log likelihood ratio of *path-measures* or stochastic processes be constant or equal to zero. Recent work [43] has also married these losses with intermediate SMC resampling.

Denoising Mean Approximation for Diffusion Settings

Diffusion models parameterized via denoising mean prediction $\hat{\mathbf{x}}_T = D_\theta(t, \mathbf{x}_t)$ provide a particularly convenient, *training-free* estimator of intermediate value functions. Instead of fully estimating the expectation in (7) or (15), one can make a single-sample approximation by evaluating $p(\mathbf{y}|\hat{\mathbf{x}}_T)$ at the denoising mean prediction,

$$V_t^{\mathbf{y}}(\mathbf{x}_t) = \log \mathbb{E}_{p^{\text{ref}}(\mathbf{x}_T|\mathbf{x}_t)}\left[p(\mathbf{y}|\mathbf{x}_T)\right] \approx \log p(\mathbf{y}|\hat{\mathbf{x}}_T) \tag{22}$$

From this approximation, we can construct an approximate guidance drift $\nabla \hat{V}_t^{\mathbf{y}}(\mathbf{x}_t) \approx \nabla \log p(\mathbf{y}|\hat{\mathbf{x}}_T)$ (for differentiable likelihoods) along with targets $\hat{V}_t^{\mathbf{y}}(\mathbf{x}_t)$ for intermediate SMC resampling in (19) [44]. This approximation has found wide applicability for inverse problems [14], protein generation [44], and images [45] for continuous diffusion models, along with recent applications for discrete diffusion models [46]. However, given that this estimator can be crude even in simple cases [4], recent work [45] finds benefits to annealing the contribution of these terms for both guidance and SMC.

Conclusion

In this blog post, we have proposed to understand controlled generation, sampling, and guidance in both language and diffusion models through the lens of probabilistic inference. Through connections with soft reinforcement learning and stochastic optimal control, we obtain a rich design space of objective functions for learning both approximate posterior distributions and value functions, which can also be used within sequential importance sampling techniques to improve generation and estimation. We hope that this overview provides useful conceptual tools for newcomers to these rapidly-evolving areas, while also contributing to the continued cross-pollination of ideas between language and diffusion model literatures, between particular problem settings within the diffusion literature, or between sampling, RL, and finetuning literatures.

Footnotes

1. We focus on continuous diffusion models here. While many concepts introduced will be relevant to discrete diffusion guidance, this remains an active area of research. [↩]
2. We will find rich applications in our the setting of no intermediate reward or targets, but refer the interested reader to [20], [1], [3] for discussion of this case in various settings. [↩]
3. See [22] Sec. 2 and Appendix for accessible derivations. [↩]

References

1. **Probabilistic Inference in Language Models via Twisted Sequential Monte Carlo**
   Zhao, S., Brekelmans, R., Makhzani, A. and Grosse, R.B., 2024. Forty-first International Conference on Machine Learning.
2. **Understanding reinforcement learning-based fine-tuning of diffusion models: A tutorial and review**
   Uehara, M., Zhao, Y., Biancalani, T. and Levine, S., 2024. arXiv preprint arXiv:2407.13734.
3. **Guidance for twisted particle filter: a continuous-time perspective**
   Lu, J. and Wang, Y., 2024. arXiv preprint arXiv:2409.02399.
4. **Particle Denoising Diffusion Sampler**
   Phillips, A., Dau, H., Hutchinson, M.J., De Bortoli, V., Deligiannidis, G. and Doucet, A., 2024. Forty-first International Conference on Machine Learning.
5. **Step-by-Step Reasoning for Math Problems via Twisted Sequential Monte Carlo**
   Feng, S., Kong, X., Ma, S., Zhang, A., Yin, D., Wang, C., Pang, R. and Yang, Y., 2024. arXiv preprint arXiv:2410.01920.
6. **Learning Diffusion Bridges on Constrained Domains** [link]
   Liu, X., Wu, L., Ye, M. and liu, q., 2023. The Eleventh International Conference on Learning Representations .

7. **Aligned diffusion Schrodinger bridges**
   Somnath, V.R., Pariset, M., Hsieh, Y., Martinez, M.R., Krause, A. and Bunne, C., 2023. Uncertainty in Artificial Intelligence, pp. 1985--1995.

8. **Doob's Lagrangian: A Sample-Efficient Variational Approach to Transition Path Sampling**
   Du, Y., Plainer, M., Brekelmans, R., Duan, C., Noe, F., Gomes, C.P., Aspuru-Guzik, A. and Neklyudov, K., 2024. The Thirty-eighth Annual Conference on Neural Information Processing Systems.

9. **Diffusion Schrodinger bridge matching**
   Shi, Y., De Bortoli, V., Campbell, A. and Doucet, A., 2024. Advances in Neural Information Processing Systems, Vol 36.

10. **Diffusion bridge mixture transports, Schrodinger bridge problems and generative modeling**
    Peluchetti, S., 2023. Journal of Machine Learning Research, Vol 24(374), pp. 1--51.

11. **Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow**
    Liu, X., Gong, C. and Liu, Q., 2023. The Eleventh International Conference on Learning Representations (ICLR).

12. **Flow Matching for Generative Modeling**
    Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M. and Le, M., 2023. The Eleventh International Conference on Learning Representations.

13. **Adding Conditional Control to Diffusion Models with Reinforcement Learning**
    Zhao, Y., Uehara, M., Scalia, G., Biancalani, T., Levine, S. and Hajiramezanali, E., 2024. arXiv preprint arXiv:2406.12120.

14. **Diffusion posterior sampling for general noisy inverse problems**
    Chung, H., Kim, J., Mccann, M.T., Klasky, M.L. and Ye, J.C., 2022. arXiv preprint arXiv:2209.14687.

15. **Diffusion posterior sampling for linear inverse problem solving: A filtering perspective**
    Dou, Z. and Song, Y., 2024. The Twelfth International Conference on Learning Representations.

16. **DEFT: Efficient Finetuning of Conditional Diffusion Models by Learning the Generalised h-transform**
    Denker, A., Vargas, F., Padhy, S., Didi, K., Mathis, S., Dutordoir, V., Barbano, R., Mathieu, E., Komorowska, U.J. and Lio, P., 2024. arXiv preprint arXiv:2406.01781.

17. **A survey on diffusion models for inverse problems**
    Daras, G., Chung, H., Lai, C., Mitsufuji, Y., Ye, J.C., Milanfar, P., Dimakis, A.G. and Delbracio, M., 2024. arXiv preprint arXiv:2410.00083.

18. **Training language models to follow instructions with human feedback**
    Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A. and others,, 2022. Advances in neural information processing systems, Vol 35, pp. 27730--27744.

19. **Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control**
    Domingo-Enrich, C., Drozdzal, M., Karrer, B. and Chen, R.T., 2024. arXiv preprint arXiv:2409.08861.

20. **Reinforcement learning and control as probabilistic inference: Tutorial and review**
    Levine, S., 2018. arXiv preprint arXiv:1805.00909.

21. **An optimization-centric view on Bayes' rule: Reviewing and generalizing variational inference**
    Knoblauch, J., Jewson, J. and Damoulas, T., 2022. Journal of Machine Learning Research, Vol 23(132), pp. 1--109.

22. **Stochastic optimal control matching**
    Domingo-Enrich, C., Han, J., Amos, B., Bruna, J. and Chen, R.T., 2023. arXiv preprint arXiv:2312.02027.

23. **The markov processes of schrodinger**
    Jamison, B., 1975. Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete, Vol 32(4), pp. 323--331. Springer.

24. **A Better Proof of Unbiasedness of the Sequential Monte Carlo Based Normalizing Constant Estimator** [HTML]
    Le, T.A., 2023. Blog Post.

25. **Two Views of Sequential Monte Carlo** [HTML]
    Le, T.A., 2022. Blog Post.

26. **Back to basics: Revisiting reinforce style optimization for learning from human feedback in llms**
    Ahmadian, A., Cremer, C., Galle, M., Fadaee, M., Kreutzer, J., Pietquin, O., Ustun, A. and Hooker, S., 2024. arXiv preprint arXiv:2402.14740.

27. **From r to : Your Language Model is Secretly a Q-Function**
    Rafailov, R., Hejna, J., Park, R. and Finn, C., 2024. arXiv preprint arXiv:2404.12358.

28. **Making ppo even better: Value-guided monte-carlo tree search decoding**
    Liu, J., Cohen, A., Pasunuru, R., Choi, Y., Hajishirzi, H. and Celikyilmaz, A., 2023. arXiv preprint arXiv:2309.15028.

29. **Mathematical theory of optimal processes**
    Pontryagin, L.S., 1962. Interscience Publishers.

30. **Scalable gradients for stochastic differential equations**
    Li, X., Wong, T.L., Chen, R.T. and Duvenaud, D., 2020. International Conference on Artificial Intelligence and Statistics, pp. 3870--3882.

31. **Efficient and accurate gradients for neural sdes**
    Kidger, P., Foster, J., Li, X.C. and Lyons, T., 2021. Advances in Neural Information Processing Systems, Vol 34, pp. 18747--18761.

32. **Path Integral Sampler: A Stochastic Control Approach For Sampling**
    Zhang, Q. and Chen, Y., 2022. International Conference on Learning Representations.

33. **Contrastive energy prediction for exact energy-guided diffusion sampling in offline reinforcement learning**
    Lu, C., Chen, H., Chen, J., Su, H., Li, C. and Zhu, J., 2023. International Conference on Machine Learning, pp. 22825--22855.

34. **A distributional approach to controlled text generation**
    Khalifa, M., Elsahar, H. and Dymetman, M., 2020. arXiv preprint arXiv:2012.11635.

35. **Aligning Language Models with Preferences through f-divergence Minimization**
    Go, D., Korbak, T., Kruszewski, G., Rozen, J., Ryu, N. and Dymetman, M., 2023. International Conference on Machine Learning, pp. 11546--11583.

36. **Target Score Matching**
    De Bortoli, V., Hutchinson, M., Wirnsberger, P. and Doucet, A., 2024. arXiv preprint arXiv:2402.08667.

37. **Bridging the gap between value and policy based reinforcement learning**
Nachum, O., Norouzi, M., Xu, K. and Schuurmans, D., 2017. Advances in neural information processing systems, Vol 30.

38. **Amortizing intractable inference in large language models**
Hu, E.J., Jain, M., Elmoznino, E., Kaddar, Y., Lajoie, G., Bengio, Y. and Malkin, N., 2023. The Twelfth International Conference on Learning Representations.

39. **Efficient (Soft) Q-Learning for Text Generation with Limited Good Data**
Guo, H., Tan, B., Liu, Z., Xing, E. and Hu, Z., 2022. Findings of the Association for Computational Linguistics: EMNLP 2022, pp. 6969--6991.

40. **Amortizing intractable inference in diffusion models for vision, language, and control**
Venkatraman, S., Jain, M., Scimeca, L., Kim, M., Sendera, M., Hasan, M., Rowe, L., Mittal, S., Lemos, P., Bengio, E. and others,, 2024. arXiv preprint arXiv:2405.20971.

41. **Solving high-dimensional Hamilton--Jacobi--Bellman PDEs using neural networks: perspectives from the theory of controlled diffusions and measures on path space**
Nusken, N. and Richter, L., 2021. Partial differential equations and applications, Vol 2(4), pp. 48. Springer.

42. **Improved sampling via learned diffusions**
Richter, L. and Berner, J., 2023. The Twelfth International Conference on Learning Representations.

43. **Sequential Controlled Langevin Diffusions**
Chen, J., Richter, L., Berner, J., Blessing, D., Neumann, G. and Anandkumar, A., 2024. arXiv preprint arXiv:2412.07081.

44. **Practical and asymptotically exact conditional sampling in diffusion models**
Wu, L., Trippe, B., Naesseth, C., Blei, D. and Cunningham, J.P., 2024. Advances in Neural Information Processing Systems, Vol 36.

45. **Alignment without Over-optimization: Training-Free Solution for Diffusion Models** [link]
Anonymous,, 2024. Submitted to The Thirteenth International Conference on Learning Representations.

46. **Derivative-free guidance in continuous and discrete diffusion models with soft value-based decoding**
Li, X., Zhao, Y., Wang, C., Scalia, G., Eraslan, G., Nair, S., Biancalani, T., Ji, S., Regev, A., Levine, S. and others,, 2024. arXiv preprint arXiv:2408.08252.

For attribution in academic contexts, please cite this work as

```
Anonymous, "Posterior Inference and Soft Value Guidance in Sequential Models", ICLR Blogposts, 2025.
```

BibTeX citation

```
@inproceedings{anonymous2025posteriorinferenceand,
  author = {Anonymous, },
  title = {Posterior Inference and Soft Value Guidance in Sequential Models},
  abstract = {Fine-tuning, controlled generation, and sampling in sequential models has attracted a flurry of recent attention in a variety of settings, particularly with the growing availability of powerful open-source pretrained models.  For language modeling in discrete spaces, we would often like to align responses with human preferences or generate correct responses to complex reasoning questions.  For diffusion models, we may be interested in steering generation to produce samples belonging a certain class, images which score highly on metrics such as realism, preference alignment, or text-to-image consistency, and proteins or molecules with desired properties such as synthesizability.  Diffusion-based methods have also been applied for sampling from arbitrary target probability densities such as Boltzmann distributions, where we can only assume access to a unnormalized density or energy function.  <br> <br>  In this blog post, we provide overview of these sampling or controlled generation tasks from a probabilistic perspective, which incorporates notions from soft reinforcement learning, stochastic optimal control, and Sequential Monte Carlo.  A key role will be played by the soft value function, which yields both importance sampling weights and gradient guidance for diffusion processes.  This perspective gives a single conceptual framework for guidance in discrete and continuous spaces, and highlights how methodologies can be shared across problem settings.},
  booktitle = {ICLR Blogposts 2025},
  year = {2025},
  date = {April 28, 2025},
  note = {http://0.0.0.0:8080/2025/blog/soft-value-guidance/#target-distributions},
  url  = {http://0.0.0.0:8080/2025/blog/soft-value-guidance/#target-distributions}
}
```