A Robust Approach to Sequential Information Theoretic Planning

Sue Zheng [1]   **Jason Pacheco** [1]   **John W. Fisher, III** [1]

Abstract

In many sequential planning applications a natural approach to generating high quality plans is to maximize an information reward such as mutual information (MI). Unfortunately, MI lacks a closed form in all but trivial models, and so must be estimated. In applications where the cost of plan execution is expensive, one desires planning estimates which admit theoretical guarantees. Through the use of robust M-estimators in the planning phase we obtain bounds on absolute deviation of estimated MI. Moreover, we propose a sequential algorithm which integrates inference and planning by maximally reusing particles in each stage. We validate the utility of using robust estimators in the sequential approach on a Gaussian Markov Random Field wherein information measures have a closed form. Lastly, we demonstrate the benefits of our integrated approach in the context of sequential experiment design for inferring causal regulatory networks from gene expression levels. Our method shows improvements over a recent method which selects intervention experiments based on the same MI objective.

1. Introduction

In many applications of Bayesian inference one is faced with the following challenges – (1) exact inference is intractable, and (2) the cost of measurement far exceeds the cost of inference. The former challenge is well studied, however the latter so-called *planning* problem has received comparatively little attention within machine learning. We consider a formulation analogous to sequential Bayesian experiment design, whereby observations lead to actions that maximize a reward. Observations are chosen to maximize information gain over some latent variables, known as *information theoretic planning*.

The Bayesian perspective on information planning largely arises from statistics with the classic work of Blackwell ([1950](#)), Lindley ([1956](#)) and later Bernardo ([1979](#)). In these works, information rewards are considered to quantify the expected reduction in posterior uncertainty (Ali & Silvey, [1966](#)). Subsequent analyses (Basseville, [1989](#); Bartlett et al., [2003](#); Nguyen et al., [2009](#)) link such measures to bounds on risk. More recently their use has been considered for sensor planning (Ertin et al., [2003](#); Kreucher et al., [2005](#); Williams et al., [2007](#)), experiment design (Drovandi et al., [2014](#)), and active learning (MacKay, [1992](#); Settles, [2012](#)).

When performing closed-loop planning over T-stages, where observed values for an action are incorporated before choosing the next action, the policy with highest *expected* reward can be determined using a dynamic programming approach (Bertsekas, [1995](#)) which leverages backwards recursion to reduce computational complexity. This approach has exponential complexity in planning horizon T and combinatorial complexity in the number of action-observation pairs. Consequently, tractable greedy heuristics, which have computational complexity linear in the number of actions and in the planning horizon, are often applied instead.

While the greedy heuristic greatly reduces computation, consideration must also be given to the cost of evaluating the information reward. In many problems exact evaluation of MI lacks a closed-form expression and so must be estimated. Moreover, the ability to accurately estimate rewards, and characterize estimation error, is crucial to making high quality decisions. For this reason, we propose a sample-based approach that employs M-estimators (Catoni, [2012](#)) to yield MI estimates that are robust to outliers and facilitates both asymptotic and finite-sample analysis of estimator properties. In a comprehensive analysis we: formulate the estimator bias, establish a central limit theorem and consistency, show probabilistic bounds of estimator deviation for finite samples, and present a probabilistic statement of plan quality. Finally, to reduce sample complexity we extend our method to a practical sequential importance sampling, which reduces com-

---
[1]Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Boston, MA, USA. Correspondence to: Sue Zheng <zhengs@csail.mit.edu>.

putation by reusing samples drawn during inference for efficient planning.

Combining these elements we demonstrate the proposed approach for planning in two models. We first consider measurement selection in a tree-structured Gaussian MRF. The ability to perform exact posterior inference in this model allows us to numerically validate properties of the estimator with respect to non-robust empirical estimation. Next we consider the inference of causal regulatory networks from gene expression data, a challenging problem where exact inference is infeasible. We demonstrate that our general method compares favorably to a recent baseline (Cho et al., 2016), specifically designed for the current application, and in particular shows large relative gains in early planning stages.

2. Robust Information Planning

In the most basic form information planning sequentially chooses actions that maximize information gain. In the general case actions typically consist of measurement choices or a combination of measurements and interventions, as in the regulatory network example in Sec. 5. We begin this section with a formulation of planning with the mutual information reward, discuss its computational challenges, and present a robust estimation procedure for information theoretic planning.

2.1. Information-Theoretic Planning

To simplify discussion we consider a model comprising latent variables x and conditionally independent observations $\mathcal{Y}_T = \{y_1, \ldots, y_T\}$. At each time t a discrete *action* $a_t \in \{1, \ldots, A\}$ parameterizes the likelihood, denoted $p_{a_t}(y_t \mid x)$. Given observations $\mathcal{Y}_T$ and actions $\mathcal{A}_T = \{a_1, \ldots, a_T\}$ the posterior is:

$$p(x \mid \mathcal{Y}_T; \mathcal{A}_T) \propto p(x) \prod_{t=1}^{T} p_{a_t}(y_t \mid x) \qquad (1)$$

Planning can be performed in an *open loop* manner, entirely offline, or *closed loop* where observations are incorporated into the plan as they are observed (Williams, 2007). We focus on the more efficient closed loop greedy setting, where actions are chosen sequentially and observed values are incorporated before choosing the next action. Specifically, at stage t we choose the action to maximize the posterior mutual information (MI):

$$a_t^* = \arg\max_a \ I_a(X; Y_t \mid \mathcal{Y}_{t-1}) \qquad (2)$$
$$= \arg\max_a \ H_a(Y_t \mid \mathcal{Y}_{t-1}) - H_a(Y_t \mid X_t, \mathcal{Y}_{t-1})$$

where $H_a(\cdot)$ denotes differential entropy under the hypothesized action a. MI can be expressed in other ways; we

choose the above form only for clarity. Similarly, our choice of conditionally independent observations in the joint Eqn. (1) is for simplification but is easily extended to the case where nuisance variables must be integrated out.

2.2. Sample-Based Estimate of Information

The MI expression (2) typically lacks a closed form. In particular, entropy requires a posterior expectation and the marginal entropy $H(Y_t \mid \mathcal{Y}_{t-1})$ requires evaluation of the log-posterior predictive distribution $\log p(y_t \mid \mathcal{Y}_{t-1})$. At time t we have posterior samples $\{x^i, y_t^i\}_{i=1}^N \sim p_a(x, y \mid \mathcal{Y}_{t-1})$. We can estimate MI at each action $a = 1, \ldots, A$ using the empirical plug-in estimator:

$$\hat{I}_a = \frac{1}{N} \sum_{i=1}^{N} \log \frac{p_a(y_t^i \mid x)}{\frac{1}{M} \sum_{j=1}^{M} p_a(y_t^i \mid x^{ij})} \qquad (3)$$

We draw independent samples $\{x^{ij}\}_{j=1}^M \sim p(x \mid \mathcal{Y}_{t-1})$ for each action and each observation sample y^i to ensure estimates are independent. These estimates parallelize across actions and marginal estimates (7) parallelize across samples $\{y^i\}$. However, naive empirical mean estimation is sensitive to outliers, particularly in the small sample regime (Catoni, 2012). In the next section we present a robust sample-based approach to information theoretic planning which avoids large estimator deviations and facilitates theoretical analysis of estimator properties in subsequent sections.

2.3. Robust Estimation of Mutual Information

We define a pair of M-estimators of the posterior moments in Eqn. (2). The class of M-estimators (Huber, 2011) is characterized by the extrema of an *influence function* which modulates the impact of outlier samples. We use the M-estimator due to (Catoni, 2012) as it was developed to challenge empirical estimation in the finite sample setting. Given a collection of i.i.d. samples $\{\theta^i\}_{i=1}^N$ the estimator $\hat{\theta} \approx \mathbb{E}[\theta]$ is given by the solution to the root equation $\sum_{i=1}^{N} \psi\left(\alpha(\theta^i - \hat{\theta})\right) = 0$ where,

$$\psi(x) = \begin{cases} \log(1 + x + x^2/2), & x \geq 0 \\ -\log(1 - x + x^2/2), & x < 0. \end{cases} \qquad (4)$$

Here α is a free parameter controlling sensitivity to outlier samples. The function $\psi(x)$ is monotonically increasing in x and has a unique root, thus making root finding efficient with standard methods.

At time t we estimate MI for each potential action using posterior samples $\{x^i, y^i\}_{i=1}^N \sim p_{a_t}(x, y_t \mid \mathcal{Y}_{t-1})$. Given an estimate of the posterior predictive distribution,

$\hat{p}_{a_t}^i \approx p_{a_t}(y^i \mid \mathcal{Y}_{t-1})$ the M-estimate of MI is:

$$\theta^i = \log \frac{p_{a_t}(y^i \mid x)}{\hat{p}_{a_t}^i}, \quad \text{for } i = 1, \ldots, N \quad (5)$$

$$\hat{I}_{a_t} \Leftrightarrow \sum_{i=1}^{N} \psi(\alpha(\theta^i - \hat{I}_{a_t})) = 0 \quad (6)$$

Here we use a plug-in M-estimate of the posterior predictive distribution. As before we ensure that estimators are independent via M iid posterior samples. Our estimate of the posterior predictive is then:

$$\theta^{ij} = p_{a_t}(y_t^i \mid x^{ij}), \quad j = 1, \ldots, M \quad (7)$$

$$\hat{p}_{a_t}^i \Leftrightarrow \sum_{j=1}^{M} \psi(\alpha(\theta^{ij} - \hat{p}_{a_t}^i)) = 0 \quad (8)$$

For planning we select the action $a_t^* = \arg\max_a \hat{I}_a$ maximizing MI. By exploiting properties of the M-estimator we are able to characterize both asymptotic and finite-sample behavior, which we address next.

3. Estimator Properties

Beginning with an asymptotic analysis we ensure the estimator is consistent and Gaussian distributed in the infinite sample limit. Despite being consistent we establish a bias for finite samples and show the rate of bias decay. While asymptotic analysis is desireable, our primary interest lies in understanding finite sample behavior and the tradeoffs with sample size. For the finite sample regime we establish probabilistic bounds on absolute estimation error and characterize the role of estimator bias. Importantly, we conclude with by formulating the probability of selecting an optimal action in a single time instance.

Bias arises in both empirical and robust estimators from the marginal entropy. More generally, an unbiased plug-in estimator of $p(y)$ will give rise to a biased estimate of the $\log p(y)$. We begin by establishing a bias decay rate using a Taylor expansion of the estimator mean:

Proposition 1. *Let $\chi^2(p(x)||q(x))$ denote the chi-square divergence of p from q. The bias of the empirical estimator depends on N, M as follows:*

$$\mathbb{E}[\hat{I}_{NM}] - I = \frac{\chi^2(p(y,x)||p(y)p(x))}{2M} + \mathcal{O}(M^{-2}).$$

M-estimators obey consistency and asymptotic normality under fairly mild assumptions placed on the influence function (continuity, monotonicity, and the existence of a unique root), for a nice summary see (DasGupta, 2008). Having established the rate of bias decay we conclude that $M = \omega(\sqrt{N})$ avoids systematic bias in the limiting distribution, leading to a consistent estimator:

Proposition 2. *Let $\{x^i, y^i\}_{i=1}^N \sim p(x, y \mid \mathcal{Y})$ and for each y^i let $\{x^{ij}\}_{j=1}^M \sim p(x \mid \mathcal{Y})$ be independent samples. As $N \to \infty$ and $M = \omega(\sqrt{N})$, the estimator $\hat{I}_{NM}$ is asymptotically normal and consistent:*

$$\sqrt{N}(\hat{I}_{NM} - I) \to \mathcal{N}(0, \sigma_{\hat{I}}^2)$$

with variance $\sigma_{\hat{I}}^2 = \sigma^2 \left(\log \frac{p(y|x)}{p(y|\mathcal{Y})} \right)$.

We have chosen, as a matter of convenience, to set the free parameter $\alpha = \sqrt{2/(N\sigma^2)}$, and similarly for the posterior predictive estimator. In this way $\alpha \to 0$ as sample size increases and the robust estimator converges to the same limiting distribution as the empirical mean. It is possible that the limiting distribution under robust estimation has tighter variance, but we did not analyze this as our interest lies in the finite sample regime.

Following the analysis of (Catoni, 2012) we establish probabilistic bounds on the deviation of the robust and empirical estimators for finite samples. Unlike in the general setting considered by Catoni, the existence of estimator bias (Prop. 1) leads to systematic overestimates of the mutual information. For any confidence level $\epsilon > 0$ we have that with probability at least $1 - 2\epsilon$ the absolute error is bounded as follows:

Proposition 3. *Let $N > 2 + 2\log(\epsilon^{-1})$ and denote the posterior predictive estimator as $\hat{p}(y; \boldsymbol{x})$ to make explicit its dependence on the samples $\boldsymbol{x} \triangleq \{x^j\}_{j=1}^M$. Then with probability $p \geq 1 - 2\epsilon$,*

$$b - c \leq \hat{I}_{NM} - I \leq b + c$$

$$\text{where,} \quad c = \begin{cases} \frac{2(1+\log\epsilon^{-1})\sqrt{\frac{\sigma_{\hat{I}_{NM}}^2}{2N}}}{1 + \sqrt{1 - 2(1+\log\epsilon^{-1})/N}}, & \text{Robust} \\ \\ \sqrt{\frac{\sigma_{\hat{I}_{NM}}^2}{2N\epsilon}}, & \text{Empirical} \end{cases},$$

and $\sigma_{\hat{I}_{NM}}^2 = \sigma^2(\log \frac{p(y|x)}{\hat{p}(y;\boldsymbol{x})})$ is the sample variance and $b = \mathbb{E}_{\boldsymbol{x}}[KL(p(y)||\hat{p}(y; \boldsymbol{x}))]$ the estimator bias.

With high probability, the deviation is bounded in the interval $[-c, c]$ that is then shifted by b. While the bias takes identical form under both estimators, it is an expected deviation, so the value of bias will differ depending on the quality of the posterior predictive estimate. As $M \to \infty$ this bias vanishes and the deviation becomes symmetric about the true value of MI.

Evaluating the true value of MI is useful for quantifying reward, and is complicated due to bias, but it is not the primary goal of planning. Indeed, the value of information is secondary to the primary focus of choosing the correct maximizing action, or more broadly the correct ordering of

actions. The later goal can be accomplished in the presence of bias. In fact, the probability that the correct action is selected can be expressed as follows:

Proposition 4. *Without loss of generality, let $I_1 \geq I_2 \geq \ldots \geq I_A$. For $N \gg 1$,*

$$\mathbb{P}(a^* = 1) \approx \int_{-\infty}^{+\infty} \mathcal{N}(\hat{I}_1; I_1, \sigma_1^2) \prod_{a=2}^{A} \Phi\left(\frac{\hat{I}_1 - I_a}{\sigma_a}\right) d\hat{I}_1$$

where $\sigma_a^2 = \frac{1}{N}\sigma^2\left(\log\frac{p_a(y|x)}{p_a(y|\mathcal{Y})}\right)$ and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

The probability in Prop. 4 is not observable but does confirm an intuitive trade-off between estimator quality and information gain which is illustrated in Fig. 1. The CDF factors indicate that low quality estimates suffice when the actual information of the ath action differs significantly from the estimate, for example where $(\hat{I}_1 - I_a)/\sigma_a$ is large. Yet, when information gain is similar across actions, $(\hat{I}_1 - I_a)/\sigma_a$ small, suboptimal decisions incur little penalty due to similar information rewards across actions. It is in the intermediate regime, where information rewards differ moderately, that estimator quality translates to higher information gain.
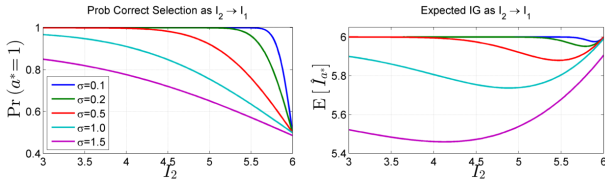

Figure 1. **Asymptotic Ranking.** *Left:* Probability of correct ranking according to Prop. 4. We vary I_2 from I_3 to I_1 where $I_1, I_3, I_4, I_5 = 6, 3, 2, 1$ respectively. As I_2 approaches I_1, the probability of selecting the correct optimizer $\mathbb{P}(a^* = 1)$, which depends on $\left(\frac{I_1 - I_2}{\sigma}\right)$, drops as there is increasing chance of selecting $a^* = 2$. *Right:* Expected information gain. While the probability of selecting the correct maximizer decreases as $I_2 \to I_1$, these actions have increasingly similar IG and the net reduction in expected IG for choosing the incorrect optimizer is small.

4. Sequential Inference and Planning

The algorithm described in Sec. 2 requires N posterior samples at each stage t and for each action $1, \ldots, A$ an additional M posterior samples are needed. Sample complexity is thus on the order $\mathcal{O}(TNAM)$. As a practical alternative we propose a sequential importance sampling approach which encourages sample reuse thereby reducing computation.

Our sequential algorithm is depicted in Fig. 2 and is motivated by sequential importance sampling for static models introduced by Chopin (2002); a special case of *resample-and-move* for dynamical systems (Gilks & Berzuini, 2001).

Drovandi et al. (2014) propose a similar approach for experimental design in model selection, though they only consider discrete observations and rely on the standard SIS estimate of the model evidence.

M-Estimator The estimator presented in Sec. 2.3 can be extended to importance weighted samples $\{\theta^i, w^i\}_{i=1}^N$ by observing the importance weighted expectation is $\mathbb{E}[\theta] \approx 1/N \sum_i N w^i \theta^i$. The importance weighted M-estimator is then given by the root equation:

$$\hat{\theta} \Leftrightarrow \sum_{i=1}^{N} \psi(\alpha(N w^i \theta^i - \hat{\theta})) = 0 \qquad (9)$$

When samples are drawn from the target distribution then $w^i = 1/N$ and Eqn. (9) reduces to the unweighted M-estimator.

Algorithm Summary At time t given samples $\{x^i, w^i\}_{i=1}^N$ we begin by sampling measurements for each hypothesized action $a = 1, \ldots, A$:

$$\{y_t^i\}_{i=1}^N \sim p_a(\cdot \mid x^i) \qquad (10)$$

As before we assume the measurement likelihood is easily sampled. We next estimate the posterior predictive distribution using an independent set of M importance weighted samples $\{x^j, w^j\}_{j=1}^M$. For each measurement sample $i = 1, \ldots, N$ the posterior predictive estimate is,

$$\hat{p}_a^i \Leftrightarrow \sum_{j=1}^{M} \psi\left(\alpha(M w^j p_a(y_t^i \mid x^j) - \hat{p}_a^i\right) = 0. \qquad (11)$$

The plug-in $\hat{p}_a^i \approx p_a(y_t^i \mid \mathcal{Y}_{t-1})$ value is used to estimate mutual information at each hypothesized action:

$$\hat{I}_a \Leftrightarrow \sum_{i=1}^{N} \psi\left(\alpha\left(N w^i \log\frac{p_a(y_t^i \mid x^i)}{\hat{p}_a^i} - \hat{I}_a\right)\right) = 0 \quad (12)$$

Next, select the maximally informative action $a_t^* = \arg\max_a \hat{I}_a$ and observe the corresponding model $y_t \sim p_{a_t}(y \mid x)$. Importance weights are updated via the likelihood to incorporate the new observation:

$$\widetilde{w}_t^i = w_t^i p_{a_t^*}(y_t \mid x^i), \quad w_t^i = \frac{\widetilde{w}_t^i}{\sum_k \widetilde{w}_t^k} \qquad (13)$$

Compute the effective sample size (ESS) as the reciprocal sum of squared weights $E = 1/\sum_i (w^i)^2$. If $E \geq \tau$, for some fixed tolerance τ, then samples remain fixed, resulting in computational savings. Conversely, if $E < \tau$ then posterior samples are *moved* via an MCMC kernel with the posterior target distribution and we set uniform weights $w^i = 1/N$.

Our implementation maintains separate weights and performs resampling independently for both sets of posterior
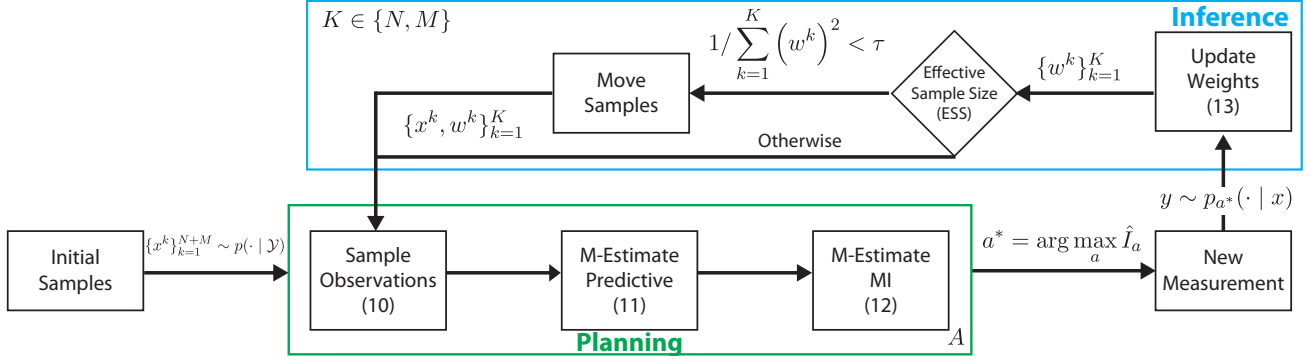
Figure 2. **Sequential Algorithm.** Our algorithm performs closed-loop greedy planning by alternating inference and planning stages. Planning: samples $y^i \sim p_a$ are drawn for each potential action $a = 1, \ldots, A$, followed by robust importance-weighted MI estimates. Note that planning estimates can be computed in parallel for all actions. Inference: After selecting action $a^* = \arg\max_a \hat{I}_a$ an observation is drawn from the corresponding model $y \sim p_{a^*}(y \mid x)$. Importance weights are then updated and new posterior samples are drawn if the ESS drops below a specified threshold τ. Numbers reference corresponding equations in the main text.

samples: those used to estimate the predictive distribution $\{x^j, w^j\}_{j=1}^M$ and for MI $\{x^i, w^i\}_{i=1}^N$. We find that this approach leads to additional computational savings by avoiding frequent resampling of the entire particle set when only a subset of weights become degenerate.

5. Experimental Results

We analyze performance of the proposed algorithm in two scenarios, beginning with a tree-structured Gaussian MRF. Posterior inference can be performed exactly and efficiently under this model, allowing us to validate our theoretical claims from Sec. 3. We further demonstrate that robust planning exhibits superior quality plans in the sequential setting, where deviation bounds to not explicitly hold. Finally, we consider sequential experiment design for gene regulatory network inference. Here, we demonstrate superior estimation of regulatory network structure prediction, with fewer interventions, compared to previous work (Cho et al., 2016). The supplementary material also contains a comparison to a related method for model selection (Drovandi et al., 2014).

5.1. Gaussian MRF Measurement Selection

Consider a tree-structured Gaussian MRF $\mathcal{G} = (\mathcal{E}, \mathcal{V})$ with edges $\mathcal{E}$, nodes $\mathcal{V}$, and joint probability $p(x, y) =$

$$\prod_{s \in \mathcal{V}} N(y_s \mid C_{a_s} x_s, \sigma^2) \prod_{(s,t) \in \mathcal{E}} N\left((x_s, x_t)^T \mid m_{st}, V_{st}\right).$$

Latent nodes x_s are 2D Gaussian random variables and observations y_s are scalar. The likelihood model at each node is defined over a set of random linear projections with parameters $\{C_a\}_{a=1}^A$. At each stage of information planning the algorithm must choose the projection maximizing infor-

mation gain $\arg\max_a I_a(X_s; Y_s)$ at the current node in a predetermined sequence. The node sequence is chosen randomly a priori and only the current node in the sequence is revealed to the planner.

In our experiments we generate random trees with $|\mathcal{V}| = 30$ nodes, each node having $A = 15$ randomly generated candidate projection operators. We draw a random sequence of $T = 25$ nodes to be observed, thus not all nodes receive measurements and the set of available projections varies with each stage. We compare performance of the sequential robust algorithm against an empirical estimator and naive random selection. Our results are summarized in Fig. 3.

Robust planning yields higher information gain. Compared to empirical estimation, robust planning consistently shows higher median information gain (IG). Fig. 3 (*left*) shows median and quartile IG over 100 random trials with $M, N = 50$ particles. By the final planning epoch quartiles of cumulative information gain are nearly non-overlapping for the two estimation methods. Overall, approximate information planning incurs a penalty of roughly 10% in realized IG under this model. Unsurprisingly, random planning incurs a much higher penalty. Moreover, the observed improvement remains as we vary sample size, Fig. 3 (*center*).

Less resampling yields more improvement. The accuracy of empirical planning degrades more rapidly than robust planning as the resampling threshold τ is reduced, resulting in lower realized IG, Fig. 3 (*right*). By setting $\alpha = \sqrt{2/(N\sigma^2)}$ in terms of the sample variance σ^2 the M-estimator better modulates the impact of outliers as importance weights degrade.

Validation of deviation bound The deviation bound in Prop. 3 is guaranteed to hold with probability at least $1 - 2\epsilon$ only for the Robust estimator with i.i.d. samples from the
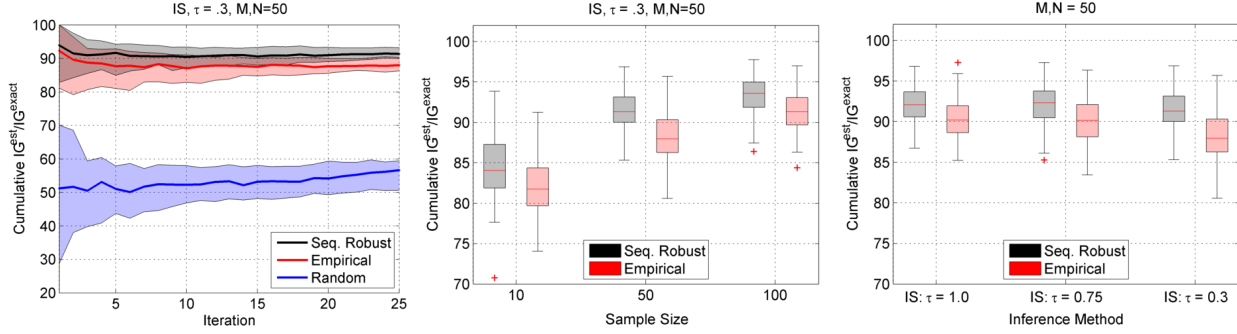
Figure 3. **Gaussian Sequential Planning.** Analysis of cumulative information gain (relative to optimal) for each planning algorithm. *Left:* Median and quartiles (shaded) of cumulative IG for 100 random trials with 50 particles. The performance difference between Robust and Empirical concentrates over iterations with nearly non-overlapping quartiles by the final iteration. *Middle:* Cumulative IG at iteration 25 for various particle counts (limits are quartiles, whiskers extremal points, and + outliers). Both estimators yield higher IG with additional particles yet improvements from using robust estimation persists across sample sizes. *Right:* Various resampling thresholds τ. IG in the non-robust estimator degrades more rapidly as the frequency of resampling decreases.
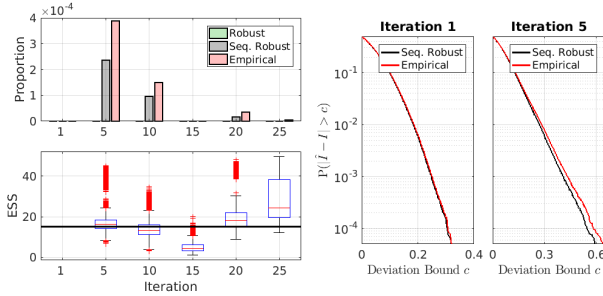

Figure 4. **Gaussian Estimator Deviations.** *Left-Top:* Deviation counts from 100K trials and $M, N = 50$ samples. Deviation bounds are evaluated for a fixed MRF at confidence level $\epsilon = 10^{-2}$ (Prop. 3) using 50M samples. Both estimators exceed the bounds at a rate less than 2ϵ, though empirical estimation does so more often than robust. *Left-Bottom:* Distribution of ESS at each iteration. Particles are resampled when ESS falls below threshold (black line). Notably, the deviation bound is never violated in instances of frequent resampling (iterations 1 and 15). *Right:* Empirical probability of deviations greater than the bound in Prop. 3. When the ESS is high (iteration 1), both algorithms perform similarly but when the ESS is low (iteration 5), Seq. Robust has lower probability of exceeding a specified deviation.

posterior. However, we verify this bound empirically for the sequential algorithm by evaluating the proportion of samples which exceed it at confidence level $\epsilon = 10^{-2}$. We estimate the bound using 50M independent posterior samples and find the bias component $b \sim 0.01$ is much less than the deviation component $c \sim 0.6$ of the bound. Both algorithms violate the bound less frequently than $1 - 2\epsilon$ suggesting Prop. 3 is conservative. Results are summarized in Fig. 4.

Empirical violates deviation bounds more frequently
While extreme deviations are rare under both estimators, suggesting that Prop. 3 is a conservative bound, we see

that empirical estimation consistently violates the deviation bound more frequently than M-estimator based planning, Fig. 4 (*top-left*). Moreover, low ESS values trigger resampling, leading to less frequent estimator deviations as Prop. 3 holds in this scenario. Fresh posterior samples are always drawn on the first iteration and we do not observe any bound violations over 100K trials as a result, Fig. 4 (*top-left*). Weights typically degenerate by iteration 15 causing most runs to draw new samples (Fig. 4 (*bottom-left*)) and the deviation bound is never violated, Fig. 4 (*top-left*). Deviations of higher magnitude have outsized probability under the empirical estimator when samples are reused, but magnitudes are similar when particles are freshly sampled, Fig. 4 (*right*).

5.2. Causal Gene Regulatory Networks

We switch focus to sequential Bayesian experiment design for estimating causal networks of gene interaction. In this setting we observe expression levels of interacting genes but do not know the causal structure of interaction. Using only network observations the underlying graph structure is only identifiable up to Markov equivalence classes (Pearl, 2003). To recover causal relationships we perform *knock-out* interventions in which a gene is removed from the network by clamping its value to zero. The number of interventions that can be performed is limited due to cell degradation; this motivates the use of an approach which correctly identifies the important interventions in the fewest planning stages. Following (Cho et al., 2016) we model the expression level of N genes, denoted $X \in \mathbb{R}^N$, as a Gaussian Bayes network G,

$$G \sim \text{Uniform-DAG},$$
$$\theta_j \mid G \sim \text{Normal-Inv-Gamma}(\alpha_j, \beta_j, \mu_j, \Lambda_j)$$
$$x_j \mid x_{\text{Pa}(j)}, \theta_j, G \sim N(m_j + w_j^T x_{\text{Pa}(j)}, \sigma_j^2).$$

Where $\mathrm{Pa}(j) = \{p_1, \ldots, p_d\}$ are the parents of node j and the network parameters $\theta_j = \{m_j, w_{jp_1}, \ldots, w_{jp_d}, \sigma_j^2\}$ describe gene interactions. At stage t let $\mathcal{I}$ denote the set of past interventions and similarly let $\mathcal{X}$ be the set of past expression levels. An intervention $I \in \{\emptyset, 1, \ldots, N\}$ clamps node $X_I = 0$ and has no effect for $I = \emptyset$. Given an intervention, expression levels are simulated from distribution given by the product of all non-clamped node likelihoods. We choose intervention at stage $t + 1$ to maximize mutual information,

$$I_{t+1} = \arg\max_{I} \mathbb{E}\left[\log \frac{p(X \mid G, \mathcal{X}; I, \mathcal{I})}{p(X \mid \mathcal{X}; I, \mathcal{I})} \right]. \quad (14)$$

Parameters θ can be explicitly marginalized out due to the use of a conjugate prior, leaving a closed form for the conditional likelihood $p(X \mid G, \mathcal{X}; I, \mathcal{I})$. However, the data evidence in the denominator of the MI objective Eqn. 14 is intractable as it requires a marginalization over graph structures–a super-exponential operation (Siracusa & Fisher III, 2009; Friedman & Koller, 2003). We estimate the evidence using a discriminance estimator and similarly approximate MI with a robust M-estimator.

In Fig. 5 we compare our approach to the active planning approach of (Cho et al., 2016) on the same model. We achieve significantly better performance in early iterations on all three criteria - MSE, area under precision recall curve (AUPRC), and area under receiver operating characteristic curve (AUROC). With more iterations, all methods, including random, perform comparably as all collect sufficiently many varied interventions to arrive at similar graph posteriors with additional interventions yielding little performance gain. The Sequential Robust approach is especially compatible with this setup as it naturally resamples graphs at initial iterations when good posterior samples are critical and the posterior is changing rapidly between iterations and saves computation at less critical later stages.

To develop some understanding about the selection process for this particular application, we show the intervention chosen by each algorithm under 50 trials over 20 iterations and the true graph with edge weights in Fig. 6. We note a strong dependence between the intervention sequence with the magnitude of the mean of each node in Fig. 6; *e.g.* Robust tends to select 6 first, followed by 4, 1, and 9. The intuition behind this is as follows: knocking out a normally highly expressive gene induces large changes in expression levels of any children nodes provided that there is a strong edge weight. Similarly, nodes 2, 3, and 8 have means near zero and are rarely selected since observations from clamping such a node would be close to observations when no nodes are clamped.

Comparing the selection differences between the two algorithms, we observe that our method consistently chooses the same initial interventions whereas the method of Cho et

al. exhibits much more variability across the 50 trials. Little is known about the graph structure in the early stages and choosing the optimal intervention can yield significantly greater performance gains over others; this is seen in the rightmost plot which shows average *realized* performance gain of each candidate intervention in the first iteration. Our method consistently selects $I = 6$, which yields twice the performance gain of the next best choice whereas Cho *et al*'s method chooses $I \in \{1, 4, 9\}$ and often waits until $t = 4$ to clamp node 6. MSE drops significantly immediately after incorporation of the observation from clamping node 6, seen in Fig. 5 for our method at $t = 2$ and for Cho et al. at $t = 5$.

The concentration of the graph posterior with increasing interventions is further illustrated in Fig. 7. Edge probabilities were calculated from 100 graph samples across 50 trials; probabilities ≤ 0.4 were filtered to declutter the graphs. The edge threshold was set based on the edge probabilities from the graph prior which take values ~ 0.3; deviations from that level reflect changes in the graph distribution. The edge probabilities change drastically following the initial interventions while little change is seen from the last 10 interventions. Using the same number of interventions, our method is able to fill in more of the graph structure. In the regime where interventions are especially costly, our method would be able to achieve the same performance using much fewer interventions.

6. Conclusion

We have presented and analyzed a robust sample-based approach to information based planning. In our analysis we have characterized both asymptotic and finite-sample behavior, thereby providing assurances of estimator quality and, consequently, guarantees on the correctness of action choices for sequential planning. Our extension to sequential importance sampling not only reduces sample complexity, but is motivated by our analysis of the robust estimator. Indeed, as importance samples degrade they tend to fail in the tails of the target distribution, where we expect the robust estimator to outperform the empirical.

On tractable Gaussian MRF models we observe consistent benefits of robust planning over empirical. Moreover, robust planning is more resilient to reduced sampling rates. On the more challenging problem of estimating gene interactions, our approach outperforms that of (Cho et al., 2016) when optimizing identical MI reward. This finding is despite the previous method being targeted to the application in question. The performance difference is greatest at early iterations which is when the benefits of planning are often greatest.
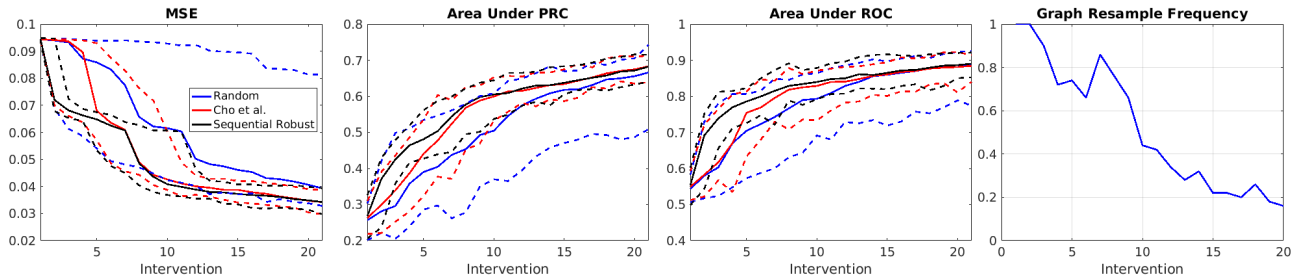
Figure 5. **Regulatory network inference evaluation.** Sequential Robust approach shows more rapid improvement in MSE of estimated edge weights (left), edge prediction AUPRC (center-left) and ROC (center-right) compared to (Cho et al., 2016). Plots show median (solid) and best/worst (dashed) runs out of 50 random trials. Sequential Robust frequently resamples graphs at early iterations when obtaining good posterior samples is critical and saves computation at later stages when information gain is negligeable (right).
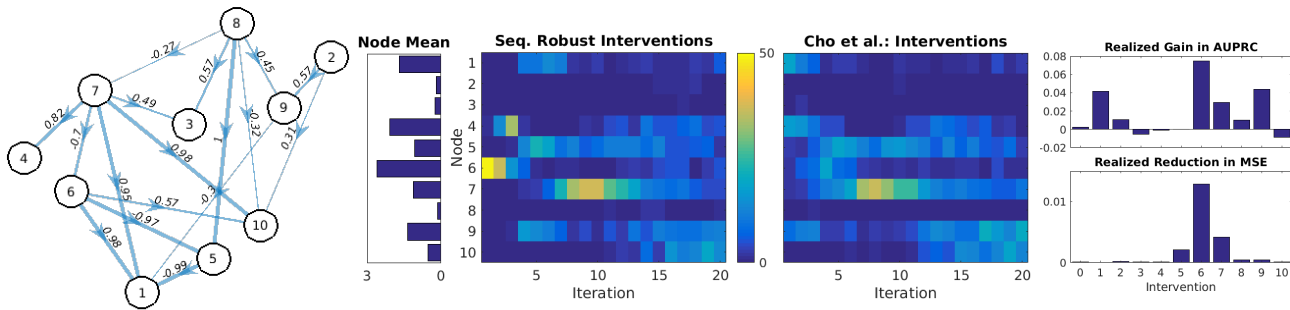

Figure 6. **Regulatory network planning.** The true network structure shows edges labeled by weights w_{ij} (left). The magnitude of the node mean in true network in the absence of any intervention (center-left) correlates strongly with selection sequence by Robust (center) since zeroing the expression level of a highly expressive gene is expected to induce large changes. Interventions chosen under Robust and Cho *et al*(center-right) from 50 trials. At early iterations Robust selects interventions consistently across trials whereas Cho *et al*'s method exhibits greater variability in the selection process. In both, greater variance is seen at later iterations when the optimal choice diverges across trials due to differences in selection history and in realized observations. The average performance gain realized after performing the specified intervention at $t = 1$ (right) indicates that clamping node 6 is by far the optimal choice.
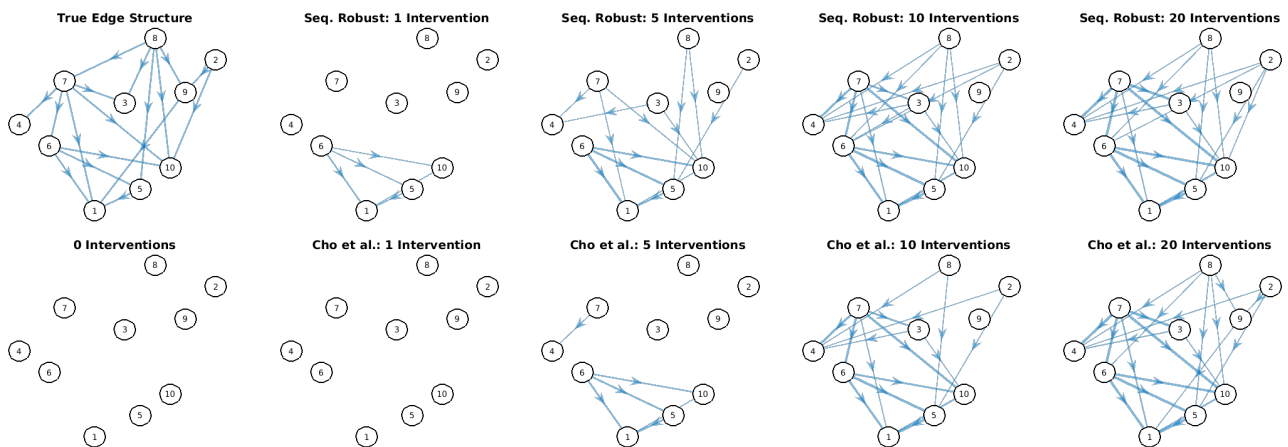

Figure 7. **Edge probabilities.** *Top-Left:* Edge structure of true graph. *Bottom-Left:* Graph prior. *Top Row:* Sequential Robust. *Bottom Row:* Cho *et al*. Edge probabilities ≥ 0.4 indicated by arrow thickness were calculated from 100 graph samples across 50 trials. Graph prior has no edge probabilities above the threshold. First five interventions identify many edges with strong weights; next five interventions refine the graph posterior by identifying additional edges; last 10 interventions do little to update the graph posterior.

References

(1):131–142, 1966. ISSN 0035-9246.

Ali, S. M. and Silvey, S. D. A general class of coefficients of divergence of one distribution from another. *JRSS*, 28

Bartlett, P. L., Jordan, M. J., and Mcauliffe, J. D. Convexity, classification, and risk bounds. *JASA*, 2003.

Basseville, M. Distance Measures for Signal Processing and Pattern Recognition. *Sig. Proc.*, 18(4):349–369, Dec 1989.

Bernardo, J. M. Expected Information as Expected Utility. *Ann. Stat.*, 7(3):686–690, May 1979.

Bertsekas, D. P. *Dynamic programming and optimal control*. Athena Scientific Belmont, MA, 1995.

Blackwell, D. Comparison of experiments. In Neyman, Jerzy (ed.), *2nd BSMSP*, pp. 93–102, Berkeley, CA, August 1950. UC Berkeley.

Catoni, O. Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. H. Poincar Probab. Statist.*, 48(4):1148–1185, 11 2012. doi: 10.1214/11-AIHP454.

Cho, H., Berger, B., and Peng, J. Reconstructing causal biological networks through active learning. *PloS One*, 11(3):e0150611, 2016.

Chopin, N. A sequential particle filter method for static models. *Biometrika*, 89(3):539–552, 2002.

DasGupta, A. *Asymptotic Theory of Statistics and Probability*. Springer Texts in Statistics. Springer New York, 2008. ISBN 9780387759715. URL https://books.google.com/books?id=sX4_AAAAQBAJ.

Drovandi, C. C., McGree, J. M., and Pettitt, A. N. A sequential monte carlo algorithm to incorporate model uncertainty in bayesian sequential design. *JCGS*, 23(1): 3–24, 2014.

Ertin, E., Fisher III, J. W., and Potter, L. C. Maximum Mutual Information Principle for Dynamic Sensor Query Problems. In *IPSN*, pp. 405–416, February 2003.

Friedman, N. and Koller, D. Being bayesian about network structure. a bayesian approach to structure discovery in bayesian networks. *JML*, 50(1-2):95–125, 2003.

Gilks, W. R. and Berzuini, C. Following a moving targetmonte carlo inference for dynamic bayesian models. *JRSS(B)*, 63(1):127–146, 2001.

Huber, P. J. *Robust statistics*. Springer, 2011.

Kreucher, C., Kastella, K., and Hero, A. Sensor management using an active sensing approach. *Signal Processing*, 85(3):607–624, 2005.

Lindley, D. V. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, December 1956. ISSN 0003-4851.

Liu, Q., Peng, J., Ihler, A., and Fisher III, J. Estimating the partition function by discriminance sampling. In *UAI*, 2015.

MacKay, D. J. C. Information-based objective functions for active data selection. *NeuralComp.*, 4(4):590–604, 1992.

Nguyen, X., Wainwright, M. J., and Jordan, M. I. On surrogate loss functions and f-divergences. *Annals of Statistics*, 2009.

Pearl, J. Causality: models, reasoning and inference. *Econometric Theory*, 19(675-685):46, 2003.

Settles, Burr. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1):1–114, 2012.

Siracusa, M. and Fisher III, J. Tractable bayesian inference of time-series dependence structure. In *AISTATS*, pp. 528–535, 2009.

Williams, J. L. *Information Theoretic Sensor Management*. PhD thesis, MIT, Cambridge, MA, USA, 2007.

Williams, J. L., Fisher III, J. W., and Willsky, A. S. Approximate Dynamic Programming for Communication-Constrained Sensor Network Management. *Trans. Sig. Proc.*, 55(8):3995–4003, August 2007.