Estimating a scalar log-concave random variable, using a silence set based probabilistic sampling

Maben Rabi, Junfeng Wu, Vyoma Singh, Karl Henrik Johansson

***Abstract*— We study the probabilistic sampling of a random variable, in which the variable is sampled only if it falls outside a given set, which is called the silence set. This helps us to understand optimal event-based sampling for the special case of IID random processes, and also to understand the design of a sub-optimal scheme for other cases. We consider the design of this probabilistic sampling for a scalar, log-concave random variable, to minimize either the mean square estimation error, or the mean absolute estimation error. We show that the optimal silence interval: (i) is essentially unique, and (ii) is the limit of an iterative procedure of centering. Further we show through numerical experiments that super-level intervals seem to be remarkably near-optimal for mean square estimation. Finally we use the Gauss inequality for scalar unimodal densities, to show that probabilistic sampling gives a mean square distortion that is less than a third of the distortion incurred by periodic sampling, if the average sampling rate is between 0.3 and 0.9 samples per tick.**

I. Introduction

Consider a sensor and a tracking station, that are connected by an ideal, analog communication link from the sensor to the tracking station. The sensor takes perfect observations of a discrete-time random process. At times of its choice, the sensor sends its current samples. The sensor is allowed to choose these times on the fly, based upon the causal record of its transmission decisions and transmitted messages up till the previous time instant. At times when the sensor does not send samples, it sends a special SILENCE symbol. This scheme is called *Event-based sampling [1].*

A. Previous results

Broadly speaking, two approaches can be seen in the literature - a deterministic one [2], [3], and a stochastic one [4], [5]. In the deterministic approach, silence set design is posed as a problem of ensuring that a candidate Lyapunov function decreases over time, or a suitable objective function of the state and control signals is minimized. In the stochastic approach, silence set design is posed as a networked sequential design problem [6], [7], with two or more decision agents. Within the stochastic approach, a variety of communication limitations have been captured: limited packet rate [8], packet losses [7] and delays [9]. Both approaches are computationally demanding, as they require the solution of Linear or Bilinear matrix inequalities, or a Dynamic programming problem.

M. Rabi is with Østfold University College, Norway. e-mail: firstname.lastname@hiof.no. J. Wu is with Chinese University of Hong Kong, Shenzen, China. V. Singh is with Koneru Lakshmaiah Education Foundation, Guntur, India. K. H. Johansson is with the Royal Institute of technology (KTH), Sweden.

If: (i) the state signal is scalar, (ii) the dynamics is linear, and (iii) the process and sensor noise densities are Gaussian, then the estimation error variance is minimized by symmetric silence intervals around the Kalman predictor [10], [11], [12], [7]. With this simplified structure for the optimal silence sets, the calculation of their sizes requires the numerical solution of an one-agent Dynamic programming problem. Andrén et al. [13] have shown that optimal silence sets can be non-convex.

Molin and Hirche [14], [15] give two-person iterative algorithm that they show to globally converge to a sequence of symmetric intervals around the Kalman predictor, if the initial state and noise are unimodal and symmetric, as assumed in the works mentioned in the previous paragraph.

A log-concave probability density [16] is one whose logarithm is a concave function over its support. Henningsson and Åström [17] showed that if the noise densities are log-concave, then optimal observer design for linear systems is simpler than without log-concavity, because the density of the estimation error is log-concave even after conditioning on the signal being inside the silence set. For s special class of log-concave densities, called strongly log-concave, in which tails decay at least as fast as for Gaussian densities, they give an upper bound on the variance of estimation error.

B. Ultra-myopic choice of silence sets

To sidestep the computational burden of calculating optimal silence sets, we use the suboptimal strategy of designing each individual silence set without regard for the effects of this choice on the costs and constraints at later times instants. Given a causal specification of a sampling rate budget, our strategy is to compute a suboptimal silence set for each time such that it minimizes only the estimation distortion at that instant, with no concern for the distortion at later time instants.

C. Our results

We pursue the question of how best to synthesize this suboptimal strategy, when the observed signal is a scalar, and is driven by noises that have log-concave densities.

In Section II we give the centering algorithm of [14], [15] for improving a given silence set. We are able to prove in Section III that for any scalar log-concave density, iterations of the centering algorithm converge to a unique optimal silence interval. This is similar to every scalar log-concave density has a unique optimal quantizer[18], [19], [20] .

When a scalar log-concave density is symmetric, then clearly it is optimal to use a symmetric silence interval around the mean cum mode. When such a density is not symmetric, we posit that super-level sets give excellent performance. We give some empirical evidence in Section IV to support this claim.

Finally, in Section V we use the Gauss inequality for unimodal densities to bound the rate-distortion trade-off for silence intervals that are symmetric around the mode.

II. Centering reduces distortion

We start with a general formulation of our problem, and specialize to the scalar case in Section III.

Problem 1. *For the random vector $X \in \mathbb{R}^n$, the probability density exists, and is given. The probabilistic sampling problem is to pick a measurable silence set $\mathcal{A}$, by an optimization in which: (a) the following chance constraint is met:*

$$\mathbb{P}\left[X \in \mathcal{A}\right] \geq \eta, \text{ for some prescribed } \eta \in [0,1], \quad (1)$$

and (b) the set $\mathcal{A}$ minimizes the average distortion:

$$\mathbb{E}\left[\delta\Big(X - \widehat{X}_{\mathcal{A}}\Big) | X \in \mathcal{A}\right],$$

where $\widehat{X}_A$ denotes the estimate under silence, and,

$$\widehat{X}_{\mathcal{A}} \triangleq \arg\min_{Y \in \mathbb{R}^n} \mathbb{E}\left[\delta\left(X - Y\right) | X \in \mathcal{A}\right], \quad (2)$$

and where the distortion function $\delta : \mathbb{R}^n \to \mathbb{R}$ is non-negative, and monotonically increasing as we move radially away from the origin along an arbitrary vector.

In other words, for any vector $x \in \mathbb{R}^n$, the value $\delta\left(\alpha x\right)$ is an increasing function of the positive real number α. Hence, given two different positive levels, the sub-level set corresponding to the smaller level is guaranteed to be entirely within the interior of the sub-level set corresponding to the higher level. Examples of valid distortion functions are:

$$|x|,\ |x|^2, \text{ or } x^T M x \text{ with the matrix } M > 0.$$

A. Centering is necessary for optimality

Definition 1 (centering of a set)**.** *Suppose we are given the random variable X, the silence set $\mathcal{S}$, the probability value η, the distortion function $\delta(\cdot)$, and the best estimate $\widehat{X}_{\mathcal{S}}$ as definied by Equation 2. Then the centering of the set $\mathcal{S}$ is denoted by $\mathcal{S}_{cen.}$, and is defined as the smallest sub-level set of the function $\delta\left(\cdot - \widehat{X}_{\mathcal{S}}\right)$, such that the set has a probability mass no less than η. In other words, if we let the nonegative number*

$$\alpha_\eta \triangleq \inf\left\{\alpha : \mathbb{P}\left[X \middle| \delta\left(X - \widehat{X}_{\mathcal{S}}\right) \leq \alpha\right] \geq \eta\right\}, \text{ then}$$
$$\mathcal{S}_{cen.} = \left\{x \in \mathbb{R}^n : \delta\left(x - \widehat{X}_{\mathcal{S}}\right) \leq \alpha_\eta\right\}. \quad (3)$$

For any two sets $\mathcal{A}, \mathcal{B}$, we denote by $\mathcal{A}/\mathcal{B}$ the set $\mathcal{A} \cap \mathcal{B}^C$, which is the set of all those elements of $\mathcal{A}$ not found in $\mathcal{B}$.

A *centered set* is any set that either equals its centering, or differs from it by a set of probability zero. This means:

$$\mathbb{P}\left[X \in \mathcal{S}/\mathcal{S}_{\text{cen.}}\right] = 0, \quad \text{if and only if } \mathcal{S} \text{ is a centered set.}$$

Algorithm 1 Centering algorithm

```
function CENTERING( p_X(·), S, δ(·), η)
    X̂_S = arg min_{Y∈R^n} E[δ(X − Y) | X ∈ S]
    α_η = inf{ α ∈ R : P[ X | δ(x − X̂) ≤ α ] ≥ η }
    S = { x ∈ R^n : δ(x − X̂) ≤ α_η }        ▷ S ⟵ S_cen.
    return S
```

For example consider sets on the real line, and let $\delta\left(\cdot\right)$ be the squared error distortion. If an interval has its conditional mean equal to its midpoint, then this interval is centered.

Lemma 1. *Let the random vector X have a regular probability density function. Given the probability value η, and the distortion function $\delta\left(\cdot\right)$, consider Problem 1. If an optimum silence set exists, then a centered one exists that is optimal.*

Proof: We shall show that if a candidate optimal set is such that it differs from its centering by a set of non-zero probability, then centering lowers the distortion.

Let the random vector $X \in \mathbb{R}^n$. Suppose that the positive number η is such that $\eta \in (0,1)$. Then consider any candidate optimal set $\mathcal{A}^*$ that is not centered, and satisfies:

$$\mathbb{P}\left[\mathcal{A}^*\right] \geq \eta.$$

Let $\widehat{X}_{\mathcal{A}^*}$ denote the best estimate under the silence set $\mathcal{A}^*$.

Without loss of generality, we can assume that the candidate optimal set has a probability mass of exactly η. If this were not the case, then we can shrink the given set, to derive a new silence set that has a probability mass of exactly η, but with a lower average distortion. In specific, given the candidate set, and a shrinkage factor σ, with $0 < \sigma < 1$, the shrunken set is:

$$\left\{x \in \mathbb{R}^n : \frac{1}{\sigma} \cdot \left(x - \widehat{X}_{\mathcal{A}^*}\right) \in \mathcal{A}^*\right\}.$$

Because the density is regular, the shrinkage factor σ can be chosen to achieve a probability mass of exactly η. Hence, we shall assume that the given candidate optimal silence set $\mathcal{A}^*$ satisfies the exact chance equality:

$$\mathbb{P}\left[\mathcal{A}^*\right] = \eta. \quad (4)$$

Then consider the centering of the given set $\mathcal{A}^*$:

$$C_\eta \triangleq \left\{x \in \mathbb{R}^n : \delta\Big(x - \widehat{X}_{\mathcal{A}^*}\Big) \leq r_\eta\right\},$$

where the 'radius' r_η is chosen such that $\mathbb{P}\left[C_\eta\right] = \eta$.

A suitable value for $_\eta$ can be chosen, because the density is regular. Since η is strictly less than 1, the radius of C_η must be finite. The set $\mathcal{A}^*$ is not centered. Hence:

$$\mathbb{P}\left[\mathcal{A}^\star / C_\eta\right] > 0.$$

Since both $\mathcal{A}^\star$, and C_η have the same probability mass,

$$\mathbb{P}\left[\mathcal{A}^\star / C_\eta\right] = \mathbb{P}\left[C_\eta / \mathcal{A}^\star\right].$$

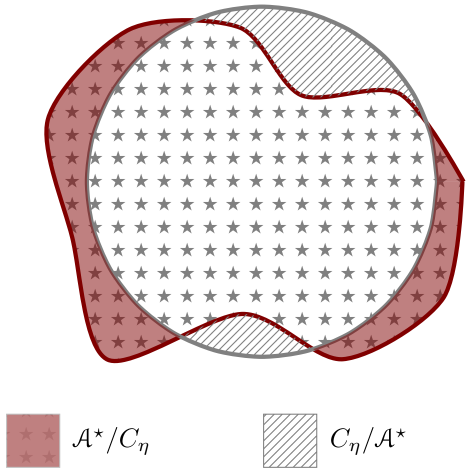


Fig. 1: The sets used in the proof of Lemma 1.

Note that on every point x of the set $\mathcal{A}^\star/C_\eta$, the distortion function $\delta\left(x-\widehat{X}_{\mathcal{A}^\star}\right)$ takes values that are bigger than values of the function at points of the set $C_\eta/\mathcal{A}^\star$. Therefore,

$$\mathbb{E}\left[\delta\left(X-\widehat{X}_{\mathcal{A}^\star}\right)\middle| X\in\mathcal{A}^\star/C_\eta\right] > \mathbb{E}\left[\delta\left(X-\widehat{X}_{\mathcal{A}^\star}\right)\middle| X\in C_\eta/\mathcal{A}^\star\right].$$

Using the above inequalities, we can write as below (at each step, the text in red signifies changes from the previous step):

$$\begin{aligned}
&\mathbb{E}\left[\delta\left(X-\widehat{X}_{\mathcal{A}^\star}\right)\middle|\mathcal{A}^\star\right]\cdot\mathbb{P}\left[\mathcal{A}^\star\right]\\
&=\int_{\mathcal{A}^\star}\delta\left(x-\widehat{X}_{\mathcal{A}^\star}\right)p_X\left(x\right)dx\\
&=\int_{\mathcal{A}^\star/C_\eta}\delta\left(x-\widehat{X}_{\mathcal{A}^\star}\right)p_X\left(x\right)dx\\
&\quad+\int_{\mathcal{A}^\star\cap C_\eta}\delta\left(x-\widehat{X}_{\mathcal{A}^\star}\right)p_X\left(x\right)dx,\\
&=\mathbb{E}\left[\delta\left(X-\widehat{X}_{\mathcal{A}^\star}\right)\middle|\mathcal{A}^\star/C_\eta\right]\cdot\mathbb{P}\left[\mathcal{A}^\star/C_\eta\right]\\
&\quad+\mathbb{E}\left[\delta\left(X-\widehat{X}_{\mathcal{A}^\star}\right)\middle|\mathcal{A}^\star\cap C_\eta\right]\cdot\mathbb{P}\left[\mathcal{A}^\star\cap C_\eta\right]\\
&>\mathbb{E}\left[\delta\left(X-\widehat{X}_{\mathcal{A}^\star}\right)\middle|C_\eta/\mathcal{A}^\star\right]\cdot\mathbb{P}\left[\mathcal{A}^\star/C_\eta\right]\\
&\quad+\mathbb{E}\left[\delta\left(X-\widehat{X}_{\mathcal{A}^\star}\right)\middle|\mathcal{A}^\star\cap C_\eta\right]\cdot\mathbb{P}\left[\mathcal{A}^\star\cap C_\eta\right],\\
&=\mathbb{E}\left[\delta\left(X-\widehat{X}_{\mathcal{A}^\star}\right)\middle|C_\eta/\mathcal{A}^\star\right]\cdot\mathbb{P}\left[C_\eta/\mathcal{A}^\star\right]\\
&\quad+\mathbb{E}\left[\delta\left(X-\widehat{X}_{\mathcal{A}^\star}\right)\middle|\mathcal{A}^\star\cap C_\eta\right]\cdot\mathbb{P}\left[\mathcal{A}^\star\cap C_\eta\right],\\
&=\mathbb{E}\left[\delta\left(X-\widehat{X}_{\mathcal{A}^\star}\right)\middle|C_\eta\right]\cdot\mathbb{P}\left[C_\eta\right],\\
&\geq\mathbb{E}\left[\delta\left(X-\widehat{X}_{C_\eta}\right)\middle|C_\eta\right]\cdot\mathbb{P}\left[C_\eta\right].
\end{aligned}$$

Since both the sets $\mathcal{A}^\star, C_\eta$ have the same probability mass,

$$\mathbb{E}\left[\delta\left(X-\widehat{X}_{\mathcal{A}^\star}\right)\middle|\mathcal{A}^\star\right] > \mathbb{E}\left[\delta\left(X-\widehat{X}_{C_\eta}\right)\middle|C_\eta\right].$$

Hence centering cannot increase the average distortion. □

III. Uniqueness of the optimal silence interval

Each run of the centering algorithm either lowers or preserves the average distortion. Consider the infinite sequence of average distortions obtained by repeatedly applying the centering algorithm. Such a sequence is non-increasing, and is also bounded below by zero. Hence this sequence converges. However, it could be that the limiting average distortion is a local minimum.

We show that log-concavity of the density implies that a local minimum must be a global minimum, in the special case where the random variable X is scalar, and the distortion function $\delta(\cdot)$ is the usual square error, or the absolute error.

The technique of our proofs is based on the generic family of all intervals with a prescribed probability mass. We span this family with a sliding interval, that has a variable length but a fixed probability mass. We shall compare the speed at which the midpoint of this interval moves, when compared to speeds of the conditional mean and the median of the interval. To save space, we give the proof of our paper's two theorems in the Arxiv version [21].

A. Centering minimizes the mean squared error

Theorem 1. *Let X be a scalar random variable with a density $p_X(\cdot)$ that is logarithmically concave. Then given any probability value $\eta\in[0,1]$, either:*

- *there is a unique interval $\mathcal{S}^* \triangleq \left[a^*_\eta, a^*_\eta+l^*_\eta\right]$ with probability mass of η, and minimizing the conditional variance $\mathbb{E}\left[\left(X-\mathbb{E}\left[X\,|X\in S\right]\right)^2|X\in S\right]$, over all silence sets $\mathcal{S}$ having probability mass at least η, or*
- *there is a sliding family containing an infinity of optimal intervals all of the same length. In specific, there is a unique positive length l^*_η, a unique lower bound $\underline{a}_\eta$, and a unique upper bound $\overline{a}_\eta$ such that, for every left end a within these bounds, the interval $\mathcal{S}^*_a \triangleq \left[a, a+l^*_\eta\right]$ has a probability mass of η, and minimizes the conditional variance over all silence sets $\mathcal{S}$ having a probability mass of at least η.*

Corollary 1.1. *A centered silence set minimizes the mean squared estimation error, if the density of the sampled random variable is logarithmically concave.*

Proof: Any non-increasing, and positive sequence of numbers must converge. Therefore the process of repeated iteration of the centering operation must lead to a limiting value for the conditional distortion.

An iteration of the centering operation cannot increase the distortion. If such an iteration results in the same distortion as before, then the original and newer sets can differ only by a set of measure zero. In which case the newer set must be a centered one. The Corollary follows from the guarantee of Theorem 1 that centered sets are essentially unique. □

B. Centering minimizes the mean absolute error

Theorem 2. *Let X be a scalar random variable with a density $p_X(\cdot)$ that is logarithmically concave. Then given any probability value $\eta\in[0,1]$, either:*

- *there is a unique interval* $\mathcal{S}^* \triangleq \left[a_\eta^*, a_\eta^* + l_\eta^*\right]$ *with probability mass of* η*, and minimizing the conditional absolute error* $\mathbb{E}\left[\left|X - mu_{[a,b]}\right| \middle| X \in S\right]$, *over all silence sets* $\mathcal{S}$ *having a probability mass of at least* η*, or*
- *there is a sliding family containing an infinity of optimal intervals all of the same length. In specific, there is a unique positive length* l_η^**, a unique lower bound* $\underline{a}_\eta$*, and a unique upper bound* $\overline{a}_\eta$ *such that, for every left end* a *within these bounds, the interval* $\mathcal{S}_a^* \triangleq \left[a, a + l_\eta^*\right]$ *has a probability mass of* η*, and minimizes the conditional absolute error over all silence sets* $\mathcal{S}$ *having a probability mass of at least* η*.*

The proof is given in the Arxiv version [21].

Corollary 2.1. *A centered silence set minimizes the mean absolute estimation error, if the density of the sampled random variable is logarithmically concave.*

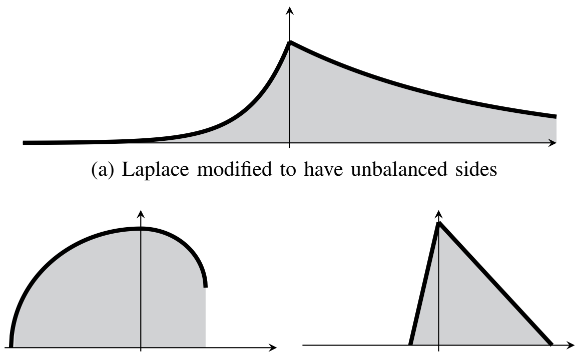
The proof is similar to that for Corollary 1.1.

IV. Super-level intervals are nearly optimal

We shall calculate the conditional variance of the following families of silence intervals. These famlilies result from heuristic attempts to keep the interval lengths as small as possible, while collecting the required probability mass η:

- **Super level sets:** A super level interval is defined as:
$$\mathcal{S}_\eta^{\text{super-level}} = \left\{x \in \mathbb{R}^n : p_X(x - \mu) \geq \alpha_\eta^{\text{super-level}}\right\},$$
where the level $\alpha_\eta^{super-level}$ is chosen to be the smallest level guaranteeing that the above interval has a probability mass of at least η.
- **Equal sides around mode:** An interval with equal sides around the mean is defined as:
$$\mathcal{S}_\eta^{\text{equal-sides}} = \left\{x \in \mathbb{R}^n : |x - \mu| \leq \alpha_\eta^{\text{equal-sides}}\right\},$$
where the half interval width $\alpha_\eta^{equal-sides}$ is chosen to be the smallest one guaranteeing that the above interval has a probability mass of at least η.
- **Equal areas around mode:** An interval with equal areas around the mean is defined as: $\mathcal{S}_\eta^{\text{equal-areas}} = \left[\alpha_\eta^{\text{equal-areas}}, \beta_\eta^{\text{equal-areass}}\right]$, where the limits are such that
$$\int_{\alpha_\eta^{\text{equal-areas}}}^{\mu} p_X(x)dx \;=\; \int_{\mu}^{\beta_\eta^{\text{equal-areas}}} p_X(x)dx \;=\; \frac{\eta}{2}.$$
- **Mode-as-conditional mean:** An interval with the conditional mean as its mode is defined as: $\mathcal{S}_\eta^{\text{mode-as-mean}} = \left[\alpha_\eta^{\text{mode-as-mean}}, \beta_\eta^{\text{mode-as-mean}}\right]$, where the limits are carefully chosen around the mean such that
$$\mathbb{E}\left[X \,\middle|\, X \in \left[\alpha_\eta^{\text{mode-as-mean}}, \beta_\eta^{\text{mode-as-mean}}\right]\right] = \mu.$$

The conditional variances of these families of intervals were calculated for three log-concave densities shown in Figure 2. These are the unbalanced Laplace density, a density patched up from two circular arcs, and a triangular density. We took these three to be representative of log-concave density classes induced by the concavity or convexity of the waveform pieces making up the densities.



(a) Laplace modified to have unbalanced sides

(b) Patchwork of circular arcs (c) Triangular density

Fig. 2: Log-concave densities used in our empirical study.

The variances incurred by the above families of silence intervals are depicted in Figure 3. Clearly super level interval are remarkably close to being optimal. Hence we can expect to get quite good approximations to the optimal silence interval by applying a couple of iterations of the centering algorithm, initializing it with a super-level interval.

V. Bound on Rate distortion trade-offs for scalar unimodal densities, via Gauss inequality

We now study the performance of probabilistic sampling for random variables that have symmetric, unimodal densities. Although log-concavity is not directly required for our result, that property is useful to preserve unimodality for the statistics of any random process with additive noise that is independent of the past and current states [22].

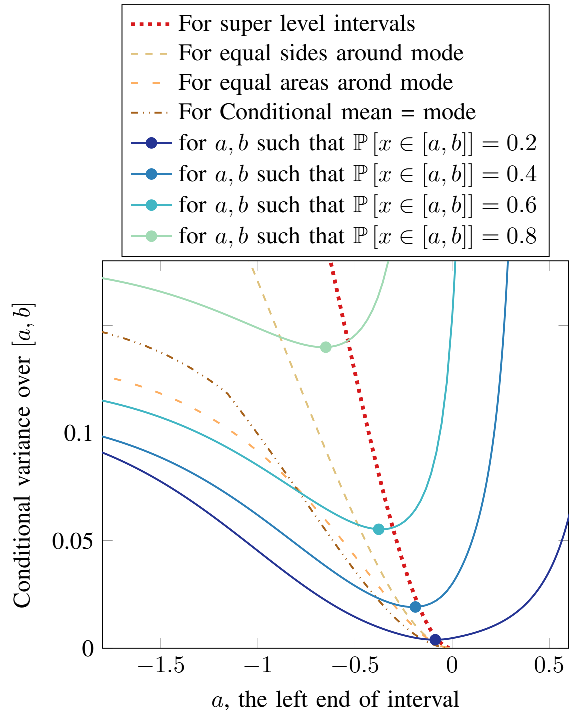
We use the Gauss inequality for the tails of scalar unimodal densities, to bound the rate-distortion curve of probabilistic sampling. We consider silence intervals that are symmetric about a mode of the density. Recall that the mode is a point such that the cumulative distribution function is convex everywhere to the left of the point, and is concave everywhere to the right of this point. This point may be non-unique; nevertheless we pick a mode and denote it by μ. We denote the silence set by $\mathcal{S} = [\mu - k, \mu + k]$. We give upper bounds for: (i) the sampling rate, and (ii) the conditional variance given that the random variable falls within the silence interval.

A. Upper bound on the sampling rate

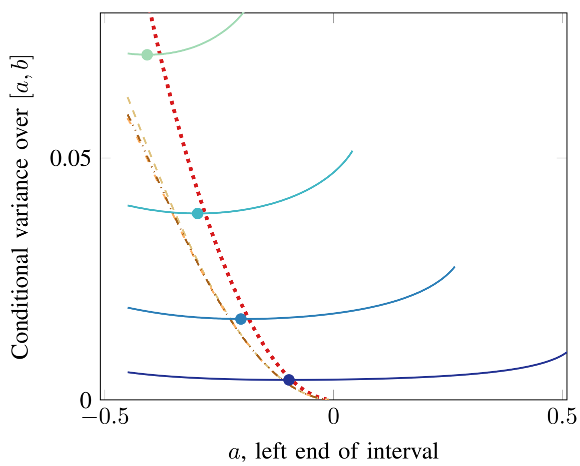
For a unimodal density $p_X(\cdot)$, let τ be defined as:
$$\tau^2 \triangleq (\text{mean} - \text{mode})^2 + \text{variance}.$$

The Gauss inequality (Section 1 in [23]) on the symmetric interval $\mathcal{S}$ around the mode, gives us:
$$\mathbb{P}\left[|X - \mu| > k\right] \;\leq\; \begin{cases} \frac{4}{9}\frac{\tau^2}{k^2}, & \text{if } k \geq \frac{2}{\sqrt{3}}\tau, \\ 1 - \frac{1}{\sqrt{3}}\frac{k}{\tau}, & \text{if } 0 \leq k \leq \frac{2}{\sqrt{3}}\tau. \end{cases} \tag{5}$$

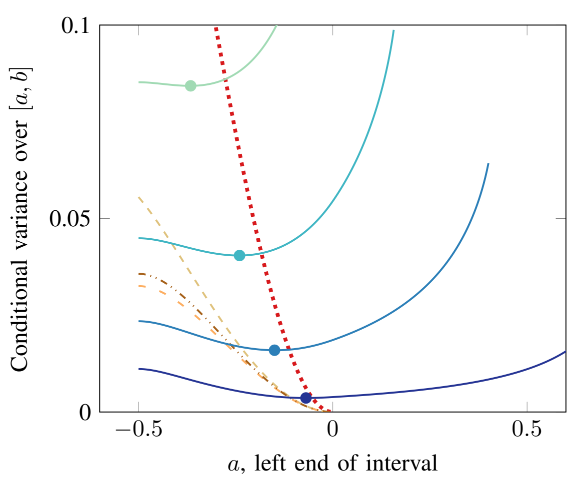
This gives us an upper bound on the rate at which the state process X falls outside the silence interval, which is exactly the rate at which samples are generated.



(a) for double sided exponential density



(b) for density with two circular arcs



(c) for triangular density

Fig. 3: The near-optimal performance of super-level intervals

B. *Upper bound on the mean-square error*

The mean-cum-mode μ is the least squares estimate of X given that it falls within the symmetric interval S. Because the density of X is unimodal, its conditional error variance of X over the interval $\mathcal{S}$ can be bounded above by the variance of the uniform distribution over the same interval, as shown below.

Let $\mathcal{A}$ be a measurable subset of $\mathcal{S}$, and let $\mathbb{P}_{U|\mathcal{S}}[A]$ denote the probability mass of the set $\mathcal{A}$ as per the uniform distribution over the interval $\mathcal{S}$. Consider the following two probabilities as functions of the positive real number t:

$$\Theta_{X|\mathcal{S}}(t) \triangleq \mathbb{P}_{X|\mathcal{S}}[|X-\mu| \le t],$$
$$\Lambda_{U|\mathcal{S}}(t) \triangleq \mathbb{P}_{U|\mathcal{S}}[|X-\mu| \le t].$$

Clearly these two functions coincide at the extremes:

$$\begin{aligned} \Theta_{X|\mathcal{S}}(0) &= \Lambda_{U|\mathcal{S}}(0) &= 0, \quad \text{and,} \\ \Theta_{X|\mathcal{S}}(t) &= \Lambda_{U|\mathcal{S}}(t) &= 1, \quad \text{if } t \ge k. \end{aligned}$$

On the interval $[0, k]$ the function $\Lambda_{U|\mathcal{S}}(t)$ is linear and increasing. On the same interval the function $\Theta_{X|\mathcal{S}}(t)$ is increasing. It is also concave. This is because

$$\Theta_{X|\mathcal{S}}(t) = \mathbb{P}_{X|\mathcal{S}}[\mu \le X \le \mu + t] + \mathbb{P}_{X|\mathcal{S}}[\mu - t \le X \le \mu],$$

and both the terms on the right hand side are concave functions of t, since their derivatives with respective to t are non-increasing functions of t (because the density p_X is unimodal about the mode μ).

On the interval $[0, k]$ the graph of $\Lambda_{U|\mathcal{S}}(t)$ is a straight line, that intersects the graph of $\Theta_{X|\mathcal{S}}(t)$ at the end points 0 and k. Hence the graph of $\Lambda_{U|\mathcal{S}}(t)$ forms a chord for the graph of the concave function $\Theta_{X|\mathcal{S}}(t)$. Hence, conditioned on being in the interval $\mathcal{S}$ the density $p_X(\cdot)$ is more peaked than the uniform density over the same interval, as in:

$$\Theta_{X|\mathcal{S}}(t) \ge \Lambda_{U|\mathcal{S}}(t), \text{ for } 0\ t \le k. \tag{6}$$

Hence the corresponding variances obey:

$$\mathbb{E}_{X|\mathcal{S}}\left[(X-\mu)^2 | X \in \mathcal{S}\right] \le \mathbb{E}_{U|\mathcal{S}}\left[(X-\mu)^2 | X \in \mathcal{S}\right] = \frac{1}{3}k^2.$$

This upper bound can be quite large if k is large. Therefore we cap it at the variance σ^2 to get the refined upper bound:

$$\mathbb{E}_{X|\mathcal{S}}\left[(X-\mu)^2 | X \in \mathcal{S}\right] \le \min\left\{\sigma^2, \frac{1}{3}k^2\right\}. \tag{7}$$

The mean squared error in estimating X, under probabilistic sampling that uses the silence set $\mathcal{S}$ is given by:

$$0 \times (1 - \mathbb{P}_X[\mathcal{S}]) + \mathbb{E}_{X|\mathcal{S}}\left[(X-\mu)^2 | X \in \mathcal{S}\right] \times \mathbb{P}_X[\mathcal{S}].$$

Using Equation (7), we can bound this quantity above with:

$$\begin{cases} \frac{4}{9}\frac{\tau^2}{k^2} \times \min\left\{\sigma^2, \frac{1}{3}k^2\right\}, & \text{if } k \ge \frac{2}{\sqrt{3}}\tau, \\ \left(1 - \frac{1}{\sqrt{3}}\frac{k}{\tau}\right) \times \min\left\{\sigma^2, \frac{1}{3}k^2\right\}, & \text{if } 0 \le k \le \frac{2}{\sqrt{3}}\tau. \end{cases} \tag{8}$$

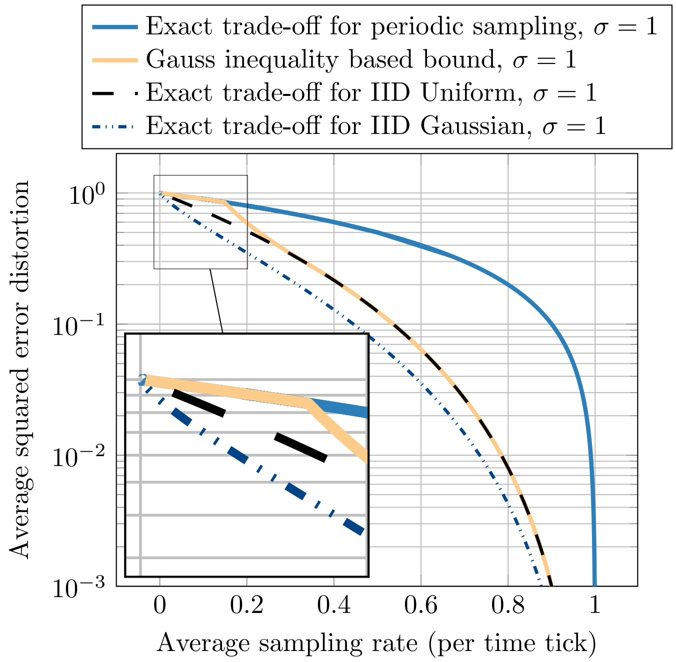


Fig. 4: Sampling rate versus Estimation error variance trade-off, for scalar IID processes with unimodal, symmetric densities.

C. A tight bound on the rate-distortion trade-off

For any scalar unimodal density with a bounded second moment, and for a silence interval symmetric about the mode, we have shown that (1) the sampling rate is bounded above by the RHS of (5), and (2) the squared error distortion is bounded above by the expression in (8).

We shall now apply these to the special case where the density is also symmetric about the mode. In this case, the quantity τ^2 found on the RHS of the Gauss inequality (5) simply equals the variance σ^2. In Figure 4 we plot our rate-distortion bound, and also the exact performances of periodic sampling, of optimal sampling when the density is uniform with $\sigma = 1$, and of optimal sampling when the density is Gaussian with $\sigma = 1$.

VI. Concluding remarks

A. Fast convergence of iterative centering

For log-concave densities, the centering algorithm converges quite fast to optimum or near-optimum silence intervals, as illustrated in Figure 5. This figure concerns the exponential density, and is a copy of Figure **??**, but now showing the how quickly the conditional mean changes under iterated centering. The fast convergence happens basically because the conditional mean of the sliding interval increases rather slowly as a function of the left end of the sliding interval. While each iteration produces only a relatively modest change in the left end of the interval, it nevertheless produces rather big changes in the value of the right end of the interval. This in turn produces big drops in the size of the interval, and also in its average distortion.

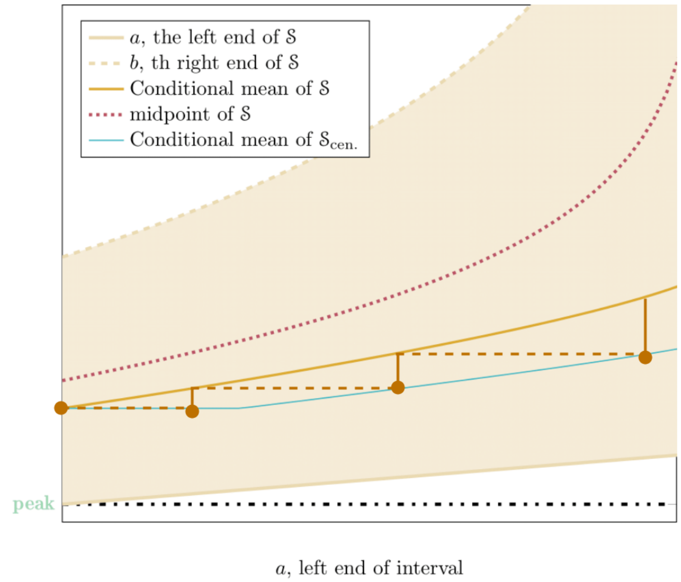


Fig. 5: Fast convergence of iterated centering

We conjecture that such fast convergence holds for every log-concave density. We think this because any tail of a log-concave density must be bounded above by an exponential decay. This shall ensure that the conditional mean of our sliding interval changes relatively slowly, even as the midpoint and the length change relatively fast.

B. Application to event-triggered sampling

For even-triggered sampling of a random process, a sub-optimal scheme follows from our formulation for individual random variables. But this needs to consider forecasting and making communication decisions over time. Firstly the sampled random process shall be correlated over time, and secondly each sample is to be generated from within a horizon that shall be longer than just one tick. This entangles the design of the silence sets at any time tick to those at other time ticks.

For example consider a single sample to be generated from within a time horizon. We have to design a probability mass function (PMF) of this sample being generated at any instant from within this horizon. This PMF shall be non-uniorm, perhaps with small values at the start of the horizon, and a peak at some suitable time instant such as the middle of horizon. At each instant, we can propagate the nonlinear filter for density of the process at that instant, given that the sample has not been generated yet. We can then apply our centering algorithm on that density.

C. Utility of the Gauss bound

The Gauss inequality is useful when the average sampling rate is in the range of approximately 0.3 to 0.9 samples per tick. It is in this range that we get the most benefits of interval based probabilistic sampling. As Figure 4 shows, there is almost no advantage to interval based sampling, if the average sampling rate is either close to zero or close to one tick per tick. We can also see an order of magnitude drop in the average distortion, if the sampling rate is between 0.6 and 0.9 samples per tick.

Acknowledgments

M. R. and V. S. thank Varun Bansal for help with programming in Python. M. R. also thanks the Automatic Control lab at KTH for hosting him during July 2016; this stay produced the main results recorded in this paper. The authors thank Adam Molin for useful discussions on Example 1. All authors thank Chithrupa Ramesh, Takashi Tanaka, and Alexandre Proutiere for valuable discussions, and the anonymous reviewers of the 2023 American Control Conference, for identifying errors in an earlier version.

References

[1] M. Miskowicz, Ed., *Event-Based Control and Signal Processing*. CRC press, 2015.

[2] M. Lemmon, “Event-triggered feedback in control, estimation, and optimization,” in *Networked Control Systems*, A. Bemporad, M. Heemels, and M. Johansson, Eds. Springer London, 2010, vol. 406, pp. 293–358.

[3] W. P. M. H. Heemels, K. H. Johansson, and P. Tabuada, “An introduction to event-triggered and self-triggered control,” in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, Dec 2012, pp. 3270–3285.

[4] K. J. Åström and B. Bernhardsson, “Systems with Lebesgue sampling,” in *Directions in mathematical systems theory and optimization*, 0.4emBerlin: Springer, 2003, vol. 286, pp. 1–13.

[5] D. Shi, L. Shi, and T. Chen, *Event-Based State Estimation*. Springer Cham, 2015.

[6] M. Rabi, C. Ramesh, and K. H. Johansson, “Separated design of encoder and controller for networked linear quadratic optimal control,” *SIAM Journal on Control and Optimization*, vol. 54, no. 2, pp. 662–689, 2016.

[7] J. Chakravorty and A. Mahajan, “Remote estimation over a packet-drop channel with markovian state,” *IEEE Transactions on Automatic Control*, vol. 65, no. 5, pp. 2016–2031, 2020.

[8] J. Chakravorty and A. Mahajan, “Fundamental limits of remote estimation of autoregressive markov processes under communication constraints,” *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1109–1124, 2017.

[9] M. J. Khojasteh, P. Tallapragada, J. Cortés, and M. Franceschetti, “The value of timing information in event-triggered control,” *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 925–940, 2020.

[10] B. Hajek, K. Mitzel, and S. Yang, “Paging and registration in cellular networks: jointly optimal policies and an iterative algorithm,” *IEEE Trans. Information Theory*, vol. 54, no. 2, pp. 608–622, 2008. [Online]. Available: http://dx.doi.org/10.1109/TIT.2007.913566

[11] G. Lipsa and N. Martins, “Remote state estimation with communication costs for first-order lti systems,” *IEEE Trans. Automatic Control*, vol. 56, no. 9, pp. 2013–2025, 2011.

[12] A. Nayyar, T. Başar, D. Teneketzis, and V. Veeravalli, “Optimal strategies for communication and remote estimation with an energy harvesting sensor,” *IEEE Trans. Automatic Control*, vol. 58, no. 9, pp. 2246–2260, 2013.

[13] M. T. Andrén, B. Bernhardsson, A. Cervin, and K. Soltesz, “On event-based sampling for lqg-optimal control,” in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, Dec 2017, pp. 5438–5444.

[14] A. Molin and S. Hirche, “An iterative algorithm for optimal event-triggered estimation,” *IFAC Proceedings Volumes*, vol. 45, no. 9, pp. 64 – 69, 2012.

[15] A. Molin and S. Hirche, “Event-triggered state estimation: An iterative algorithm and optimality properties,” *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5939–5946, 2017.

[16] M. Y. An, “Log-concave Probability Distributions: Theory and Statistical Testing,” Duke University, Tech. Rep., 1995. [Online]. Available: https://econwpa.ub.uni-muenchen.de/econ-wp/game/papers/9611/9611002.pdf

[17] T. Henningsson and K. J. Åström, “Log-concave observers,” in *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems*, Kyoto, Japan, Jul. 2006.

[18] A. V. Trushkin, “Sufficient conditions for Uniqueness of a Locally Optimal Quantizer for a class of Convex error weighting functions,” *IEEE Trans. Information theory*, vol. IT-28, no. 2, pp. 187–198, March 1982.

[19] ——, “Monotony of Lloyd’s method II for log-concave density and convex error weighting function (corresp.),” *IEEE Trans. Information Theory*, vol. 30, no. 2, pp. 380–383, 1984.

[20] J. Kieffer, “Uniqueness of locally optimal quantizer for log-concave density and convex error weighting function,” *IEEE Transactions on Information Theory*, vol. 29, no. 1, pp. 42–47, 1983.

[21] M. Rabi, J. Wu, V. Singh and K. H. Johansson, “Estimating a scalar log-concave random variable, using a silence set based probabilistic sampling,” *preprint.*, http://arxiv.org/abs/2303.04584 March 2023.

[22] I. Ibragimov, “On the composition of unimodal distributions,” *Theor. Probability Appl.*, vol. 1, no. 2, pp. 255–260, 1956.

[23] P. J. Bickel and A. M. Krieger, “Extensions of Chebychev’s inequality with applications,” *Probab. Math. Statist.*, vol. 13, no. 2, pp. 293–310 (1993), 1992.