On the Trackability of Stochastic Processes
Based on Causal Information
Baran Tan Bacinoglu∗, Yin Sun ‡, Elif Uysal∗
∗METU, Ankara, Turkey, ‡Auburn University, AL, USA,
E-mail: barantan@metu.edu.tr, yzs0078@auburn.edu, uelif@metu.edu.tr

Abstract—We consider the problem of tracking an unstable
stochastic process \( X_t \) by using causal knowledge of another
stochastic process \( Y_t \). We obtain necessary conditions and suf-
ficient conditions for maintaining a finite tracking error. We
provide necessary conditions as well as sufficient conditions for
the success of this estimation, which is defined as order \( m \)
moment trackability. By-products of this study are connections
between statistics such as Rényi entropy, Gallager’s reliability
function, and the concept of anytime capacity.

Index Terms—tracking conditions; causal information; Hölder
inequality; Rényi entropy; Gallager’s reliability function;
Gartner-Ellis limit; anytime capacity; causal estimation

I. INTRODUCTION

The tracking of unstable processes from noisy, delayed or
infrequent samples is a fundamental problem that naturally
arises in networked control systems. Non-stationary stochastic
processes such as random walks and their scaling limits such as
the Wiener process and the Ornstein-Uhlenbeck process are
examples of unstable processes that are often used to model
uncontrolled systems. Tracking of such processes may also
arise in situations that do not require closed-loop control, and
may have applications beyond networked control systems. For
example, the Wiener process is used to model the physical
diffusion process known as Brownian motion [1], option
pricing in financial analysis [2], phase noise in communication
channels [3] and forms a basis for analysis tools such as
Feynman-Kac formula [4].

The problem of communicating the state of unstable sources
has been considered in [5], which introduced the notions of
anytime reliability and anytime capacity. It was shown that
anytime capacity provides the necessary and sufficient condi-
tion on the rate of an unstable scalar Markov source that
is tractable in the finite mean-squared error sense. Accord-
ingly, it was claimed that anytime capacity, which is
upper bounded by Shannon capacity, is the correct figure of
merit to measure the quality of a channel on the purpose of
tracking an unstable source and also controlling through an
unreliable channel [6]. However, while anytime capacity is
known to be strictly positive for some channels, a closed-
form expression for anytime capacity has not been shown as
opposed to Shannon capacity which can be expressed as an
optimization of mutual information. On the other hand, any-
time capacity of particular channels such as erasure channels
with feedback [7] and Markov channels [6] have been derived.

Aside from the theory developed in [5], the problem of
stabilizing a system with limited communication has been
extensively studied from stochastic control [8]–[12], rate-
distortion theory [13]–[17] and joint source-channel coding
[18] perspectives for linear systems and from the perspectives
of metric and topological entropy for non-linear dynamical
systems [19]. In our study, the state estimation side of this
problem is investigated from a perspective that is centered on
a definition of reliable estimation which we refer as order \( m \)
moment trackability in accordance with \( m \)-th moment stability.

Based on this definition, we study the estimation of integer-
valued \(^1\) stochastic processes which may represent linear or
non-linear discrete-time systems.

Our contributions are as follows:
• We show two moment-entropy inequalities for integer-
valued random variables inspired from the inequality for the
moments of guessing random variables in [20]. One
of these bounds is for bounded integer-valued random
variables (see Lemma 1) while the other (see Lemma 2)
is valid for integer-valued random variables that do not
necessarily have finite support.
• We provide necessary conditions (see Theorem 1 and
Theorem 2) for tracking integer-valued sources using
causal information. Corollaries of Theorem 1 are upper
bounds on anytime capacity based on Gallager’s reliabil-
ity function and the Gartner-Ellis limit of the information
density between channel inputs and outputs.
• We provide sufficient conditions for tracking integer-
valued sources using causal information in Theorem 3
and Theorem 4 where the former is based on an upper
bound for the estimation error of maximum a posteriori
(MAP) estimators (Lemma 3) and the latter is based on
estimators we suggest.

II. SYSTEM MODEL

Consider the problem of tracking a scalar discrete-time
and discrete-valued stochastic process \( \{X_t\}_{t=1,2,\ldots} \) based
on causal knowledge of another stochastic process \( \{Y_t\}_{t=1,2,\ldots} \).
At any time \( t \), the estimator generates a guess \( \hat{X}_t = \tilde{f}_t(Y_{1:t}) \)
of the current value \( X_t \), where \( \tilde{f}_t(\cdot) \) is a function and
\( Y_{1:t} = (Y_1, Y_2, \ldots, Y_t) \) is the information that is available at time \( t \).

Definition 1. For any \( m > 0 \), \( \{X_t\}_{t=1,2,\ldots} \) is said to be order
\( m \) moment trackable based on \( \{Y_t\}_{t=1,2,\ldots} \) if there exists a

\(^1\)Our results are for integer-valued sources, however, note that this is not
restrictive for digital systems where data is represented using integers.
family of functions \( \{f_t(\cdot)\}_{t=1,2,...} \) such that \( \hat{X}_t = f_t(Y_{1:t}) \) and
\[
\sup_{t>0} \mathbb{E} \left[ |X_t - \hat{X}_t|^m \right] < \infty. \tag{1}
\]

The first goal of the present paper is to find necessary conditions and sufficient conditions for the \( m \)-th moment trackability of process \( \{X_t\}_{t=1,2,...} \), based on the side information process \( \{Y_t\}_{t=1,2,...} \). In [5], the anytime capacity of a noisy channel was shown to be a necessary and sufficient quality measure of a channel to allow order \( m \) moment trackability of a Markov source \( \{S_t\}_{t=1,2,...} \) based on the channel output \( \{Y_t\}_{t=1,2,...} \). The second goal of the paper is to find new bounds of the anytime capacity, based on the trackability results.

### III. Main Results

#### A. Necessary Conditions for Trackability

We provide two necessary conditions for order \( m \) moment trackability, which are expressed in terms of Rényi entropy and information density. The Rényi entropy of order \( \alpha \), where \( \alpha \geq 0 \) and \( \alpha \neq 1 \), is defined as [21]
\[
H_\alpha(X) = \frac{1}{1-\alpha} \log \mathbb{E} \left[ P_X(X)^\alpha \right] - 1
\]
\[
= \frac{1}{1-\alpha} \log \left[ \sum_{x \in \mathcal{X}} P_X(x)^\alpha \right]. \tag{2}
\]

Given joint distribution \( P_{XY} \), the information density function is defined as [22]
\[
i(x;y) = \log \left[ \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} \right]. \tag{3}
\]

The first necessary condition that we present is as follows:

**Theorem 1.** If \( \{X_t\}_{t=1,2,...} \) is an integer-valued stochastic process that satisfies
\[
|X_t| \leq c_t, \tag{5}
\]
\[
\lim_{t \to \infty} \frac{1}{t} \log(\log(c_t)) = 0, \tag{6}
\]
then \( \{X_t\}_{t=1,2,...} \) is order \( m \) moment trackable based on \( \{Y_t\}_{t=1,2,...} \), where \( Y_t \in \mathcal{Y} \) and \( |\mathcal{Y}| < \infty \), only if the following inequality holds, for all \( \rho \in (0,m] \) and \( q > \rho + 1 \),
\[
\lim_{t \to \infty} \frac{1}{pt} \log \mathbb{E} \left[ e^{-\frac{1}{\rho}i(X_t,Y_t)} | Y_t^{q} \right] \geq \lim_{t \to \infty} \frac{1}{t} H_{\frac{1}{\rho+1}}(X_t). \tag{7}
\]

**Proof.** See Appendix A.

The proof of Theorem 1 uses the following moment-entropy inequality for the Rényi entropy, which is inspired by Theorem 1 in [20].

**Lemma 1.** If \( X \) is an integer-valued random variable taking values from the set \( \mathcal{X} = \{-M_-,\ldots,-1,0,1,\ldots,M_+\} \) where \( M_- \) and \( M_+ \) are positive integers, then for all \( \rho \geq 0 \)
\[
\mathbb{E}[|X|^\rho] + 1 \geq [3 + \log(M_- M_+)]^{-\rho} e^{\rho H_{\frac{1}{\rho+1}}(X)}. \tag{8}
\]

**Proof.** See Appendix B.

Lemma 1 requires that \( M_- \) and \( M_+ \) are finite. As a result, Theorem 1 only applies to stochastic processes that satisfy (5) and (6). Next, we will provide a necessary condition for the trackability of unbounded stochastic processes in Theorem 2, which is based on the following moment-entropy inequality.

**Lemma 2.** If \( X \) is an integer-valued random variable, then for all \( \rho \in (0,m) \)
\[
\mathbb{E}[|X|^\rho] + 1 \geq \left[ 1 + 2\zeta\left(\frac{m}{\rho}\right) \right]^{-\rho} e^{\rho H_{\frac{1}{\rho+1}}(X)}, \tag{9}
\]
where \( \zeta(\cdot) \) is the Riemann zeta function
\[
\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}. \tag{10}
\]

**Proof.** See Appendix C.

**Theorem 2.** An integer-valued stochastic process \( \{X_t\}_{t=1,2,...} \) is order \( m \) moment trackable based on \( \{Y_t\}_{t=1,2,...} \), where \( Y_t \in \mathcal{Y} \) and \( |\mathcal{Y}| < \infty \), only if (7) holds for all \( \rho \in (0,m) \) and \( q > \rho + 1 \).

**Proof.** The proof is identical to the proof of Theorem 1, except that it uses Lemma 2 instead of Lemma 1. Note that \( \zeta\left(\frac{m}{\rho}\right) \) is finite for all \( \rho \in (0,m) \).

**B. Upper Bounds of Anytime Capacity**

Now, we show that (7) implies two inequalities that provide upper bounds on anytime capacity. First one can be expressed in terms of Gallager’s reliability function which is defined as [23]
\[
E_0(\rho, P_{Y|X}, P_X) = -\log \sum_{y \in \mathcal{Y}} \left( \sum_{x \in \mathcal{X}} P_X(x) P_{Y|X}(y|x) \right)^{1+\rho}. \tag{11}
\]

In [24], an alternative expression for Gallager’s reliability function was used as follows
\[
E_0(\rho, P_{Y|X}, P_X) = -\log \mathbb{E} \left[ e^{-\frac{1}{\rho}i(X;Y)} | Y \right]^{1+\rho}, \tag{12}
\]
where \( P_{XY} = P_X(x) P_{Y|X}(y|x) P_X(x) \) is the joint density for \( X, Y \) and \( \bar{X} \).

In this paper, we find the following expression of Gallager’s reliability function convenient, due to its connection with the LHS of (7).
\[
E_0(\rho, P_{Y|X}, P_X) = -\log \mathbb{E} \left[ e^{-\frac{1}{\rho}i(X;Y)} | Y \right]^{1+\rho}. \tag{13}
\]

Using (13), one can observe that the LHS of (7) becomes the Gallager’s reliability function as \( q \) reduces to \( \rho + 1 \). Based on this observation, we derive the following corollary of Theorem
Corollary 1. Suppose that \( S_t \rightarrow X_{1:t} \rightarrow Y_{1:t} \) is a Markov chain for each \( t \). If \( \{ S_t \}_{t=1,2,...} \) is an integer-valued stochastic process that satisfies
\[
|S_t| \leq c_t, \quad (14)
\]
\[
\lim_{t \to \infty} \frac{1}{t} \log(\log(c_t)) = 0, \quad (15)
\]
then \( \{ S_t \}_{t=1,2,...} \) is order \( m \) moment trackable based on \( \{ Y_t \}_{t=1,2,...} \) where \( Y_t \in \mathcal{Y} \) and \( |\mathcal{Y}| < \infty \), only if
\[
\lim_{t \to \infty} \frac{1}{mt} E_0(m, P_{Y_{1:t}|X_{1:t}}, P_{X_{1:t}}) \geq \limsup_{t \to \infty} \frac{1}{t} H_\infty(S_t). \quad (16)
\]

Proof. Apply Theorem 1 for \( S_t \) considering \( \rho = m \) and the limit that \( q \) reduces to \( \rho + 1 \) yields
\[
\lim_{t \to \infty} \frac{1}{mt} E_0(m, P_{Y_{1:t}|S_t}, P_{S_t}) \geq \limsup_{t \to \infty} \frac{1}{t} H_\infty(S_t). \quad (17)
\]
Observe that (17) implies (16) as \( E_0(m, P_{Y_{1:t}|S_t}, P_{S_t}) \) is upper bounded by \( E_0(m, P_{Y_{1:t}|X_{1:t}}, P_{X_{1:t}}) \) due to data-processing inequality for Rényi divergence (see [24, Theorem 5]). □

Corollary 1 can be related to the \( \alpha \)-anytime capacity of a channel (see [25, Definition 3.2]) when we consider the following communication system. Let \( Y_{1:t} \) be the outputs of a channel given by \( P_{Y_{1:t}|X_{1:t}}(y_{1:t}|x_{1:t}) \) with \( X_{1:t} \) being inputs that encode a source \( \{ S_t \} \). As the outputs of the channel depend on the source process only through the channel inputs, the system follows \( S_t \rightarrow X_{1:t} \rightarrow Y_{1:t} \). For ease of analysis, we will consider the type of source representing a stream of bits with fixed rate as follows:

Definition 2. For \( R \) being a positive integer, a discrete-time process \( \{ S_t \} \) is said to be a rate-\( R \) source if it obeys:
\[
S_{t+1} = 2^R S_t + W_t, \quad (18)
\]
where \( \{ W_t \} \) is an i.i.d. process such that \( W_t \) is uniformly chosen from the set \( \{ 0, 1, ..., 2^R - 1 \} \), and \( X_0 = 0 \).

Note that a rate-\( R \) source satisfies \( |S_t| \leq 2^Rt \) almost surely and \( H_\infty(S_t) = Rt \log(2) \) as it has a uniform distribution for all \( t \). Accordingly, we can apply Corollary 1 to a rate-\( R \) source and show the following

Corollary 2. If \( C_{any}(\alpha) \) is the \( \alpha \)-anytime capacity of a discrete memoryless channel (DMC) without feedback, \( R \) is a positive integer, \( m > 0 \) is an arbitrary positive number, and
\[
R \log(2) \leq C_{any}(mR), \quad (19)
\]
then
\[
R \log(2) \leq \frac{E_0(m)}{m}, \quad (20)
\]
where \( E_0(m) = \sup_{P_X} E_0(m, P_{Y|X}, P_X) \) for given transition probabilities \( P_{Y|X} \) of the channel.

Proof. First suppose that (19) holds which means a rate-\( R \) source is order \( m \) moment trackable though a DMC with anytime capacity \( C_{any}(\alpha) \) (see [25, Theorem 3.3]). On the other hand, if a rate-\( R \) source is order \( m \) moment trackable through a DMC, the following should also hold:
\[
\liminf_{t \to \infty} \frac{1}{mt} E_0(m, P_{Y_{1:t}|X_{1:t}}, P_{X_{1:t}}) \geq R \log(2), \quad (21)
\]
which follows from Corollary 1. Moreover, this implies (20) or \( E_0(m, P_{Y_{1:t}|X_{1:t}}, P_{X_{1:t}}) \leq t E_0(m) \) (see [23, Theorem 5]) for DMCs without feedback.

A result that is similar to Corollary 2 was shown (see [7, Theorem 3.3.2]) for symmetric DMCs with feedback based on sphere packing exponent. On the other hand, Corollary 2 holds both for asymmetric and symmetric DMCs without feedback.

The second inequality that we provide can be obtained from (16) while considering a rate-\( R \) source for \( S_t \). Accordingly, when \( Y_{1:t} \) are the outputs of a channel with inputs \( X_{1:t} \) that encode a rate-\( R \) source, the source is order \( m \) trackable based on \( Y_{1:t} \) only if:
\[
\liminf_{t \to \infty} \frac{1}{pt} \log E_p[\rho(X_{1:t}|Y_{1:t})] \geq R \log(2). \quad (22)
\]

In fact, the LHS of (22) is the Gartner-Ellis limit of \( i(X_{1:t}|Y_{1:t}) \) which provides another upper bound for anytime capacity if we use (22) instead of (21) in the proof of Corollary 2. Also, observe that both (21) and (22) can be applied for channels other than DMCs without feedback.

C. Sufficient Conditions for Trackability

Next, we provide two sufficient conditions for order \( m \) moment trackability. The first one is based on MAP estimators.

Definition 3. An estimator \( \hat{X}_t(MAP) \) is said to be a maximum a posteriori (MAP) estimator if
\[
\hat{X}_t(MAP) = \arg \max_{x \in X} P_{X_t|Y_{1:t}}(x|Y_{1:t}), \quad (23)
\]
with ties in the maximization broken arbitrarily.

We will use the following lemma to derive a sufficient condition for order \( m \) moment trackability based on MAP estimators:

Lemma 3. For an integer-valued stochastic process \( \{ X_t \} \), a discrete-valued stochastic process \( \{ Y_t \} \) and \( d(\cdot, \cdot) \) being a distance metric such that \( d : X \times X \to \mathbb{Z} \geq 0 \) we have the following for arbitrary real numbers \( \rho > 0 \) and \( s > 1 \):
\[
\mathbb{E} \left[ d(X_t, \hat{X}_t(MAP)) \right] \leq \zeta(s) \sum_{y_{1:t}} P_{Y_{1:t}}(y_{1:t}) \left( \sum_{x} \left[ P_{X_t|Y_{1:t}}(x|y_{1:t}) \right]^{\frac{1}{s}} d(x, x')^{s} \right)^{\rho}, \quad (24)
\]

Proof. See [26]. □

A sufficient condition for order \( m \) moment trackability using Lemma 3 is as follows:

Theorem 3. Let
\[
\tau(x, y_{1:t}) = \mathbb{E} \left[ P_{X_t|Y_{1:t}}(X_t|Y_{1:t}) \right] d \left( x, x' \right)^{s} \left| X_t = x \right. \left| Y_{1:t} = y_{1:t} \right), \quad (25)
\]

[^2]: See [26] for the proof.
where \( s > 1 \) is an arbitrary real number and \( m \) is an integer. Then, the integer-valued stochastic process \( \{X_t\}_{t=1,2,...} \) is order \( m \) moment trackable using \( \{Y_t\}_{t=1,2,...} \) if
\[
\sup_{t>0} \mathbb{E} \left[ \tau(X_t, Y_{1:t})^m P_{X_t|Y_{1:t}}(X_t|Y_{1:t})^{\frac{m}{m+1}} \right] < \infty. \tag{26}
\]

**Proof.** Apply Lemma 3 for \( d(x, x') = |x - x'|^m \) and \( \rho = m \), then observe that
\[
\mathbb{E} \left[ |X_t - \hat{X}_t|^{\text{MAP}}_m \right] \leq \nabla(s) \mathbb{E} \left[ \tau(X_t, Y_{1:t})^m P_{X_t|Y_{1:t}}(X_t|Y_{1:t})^{-\frac{m}{m+1}} \right]. \tag{27}
\]

In addition to MAP estimators, we consider another type of estimators which are defined below:

**Definition 4.** For \( \rho > 0 \) being an arbitrary real number, let \( \{\hat{X}_t^{(\rho)}(Y_{1:t})\} \) be a family of estimators such that \( \hat{X}_t^{(\rho)}(Y_{1:t}) \) is uniformly chosen from the set \( A_t(\rho, y_{1:t}, J_t(\rho, y_{1:t})) \) where
\[
A_t(\rho, y_{1:t}, c) = \left\{ x : \frac{P_{X_t|Y_{1:t}}(x|y_{1:t})}{P_{X_t|Y_{1:t}}(x'|y_{1:t})} \geq c |x - x'|^{\rho}, \forall x' \right\}, \tag{28}
\]
and
\[
J_t(\rho, y_{1:t}) = \sup\{ c \geq 0 : A_t(\rho, y_{1:t}, c) \neq \emptyset \}. \tag{29}
\]

Observe that, as opposed to MAP estimators, the estimator \( \hat{X}_t^{(\rho)} \) has a notion of distance and it requires that a possible value to be less likely proportional with its distance to the estimate. This requirement is natural as more likely values cluster around the estimate value. Accordingly, considering the family of estimators \( \{\hat{X}_t^{(\rho)}\} \) yields

**Theorem 4.** If \( p > 1 \) and \( s > 1 \) are arbitrary real numbers, and \( m \) is a positive integer, then, the integer-valued stochastic process \( \{X_t\}_{t=1,2,...} \) is order \( m \) moment trackable on \( \{Y_t\}_{t=1,2,...} \) if
\[
\sup_{t>0} \mathbb{E} \left[ \mathbb{E} \left[ P_{X_t|Y_{1:t}}(X_t|Y_{1:t})^{-\frac{m}{m+1}} \mid Y_{1:t} \right]^{p(m+1)} \right]^{\frac{1}{p}} \times \mathbb{E} \left[ J_t(sm(m + 1), Y_{1:t})^{\frac{s-1}{p}} \right] < \infty, \tag{30}
\]
where \( J_t(\rho, y_{1:t}) \) is as defined in (29).

**Proof.** See [26]. \( \square \)

Note that the first term in (30) can be expressed in terms of conditional Rényi 3 entropy when \( p = 1 \) while \( J_t \) function in the second term can be considered as a measure of the shape of the conditional distribution \( P_{X_t|Y_{1:t}}(x|Y_{1:t}) \).

### IV. Conclusion

We considered necessary and sufficient conditions for tracking a random source. Our results may provide insights to the design of causal information (via real-time coding) for systems that rely on the tracking of random sources.

\(^{3}\)We consider the definition of conditional Rényi entropy that fits our case.

### V. Acknowledgements

The authors would like to thank Serdar Yuksel and Kemal Leblebicioglu for their valuable comments.

### Appendix

#### A. The Proof of Theorem 1

Consider arbitrary estimators \( \{\hat{X}_t\} \) such that \( |\hat{X}_t| \leq c_t \) for \( t > 0 \). Let us define estimators \( \{\hat{X}_t^{(c)}\} \) such that \( \hat{X}_t^{(c)} = \lceil \hat{X}_t \rceil \) where \( \lceil \cdot \rceil \) is the ceiling function. If \( m \in (1, \infty) \),
\[
\mathbb{E} \left[ |X_t - \hat{X}_t^{(c)}|_m \right] \leq \mathbb{E} \left[ |X_t - \hat{X}_t|_m \right] + 1, \tag{31}
\]
where the inequality follows from Minkowski’s inequality and that \( \mathbb{E} \left[ |\hat{X}_t - \hat{X}_t^{(c)}|_m \right] \leq 1 \). If \( m \in (0, 1) \),
\[
\mathbb{E} \left[ |X_t - \hat{X}_t^{(c)}|_m \right] \leq \mathbb{E} \left[ |X_t - \hat{X}_t|_m \right] + \mathbb{E} \left[ |\hat{X}_t - \hat{X}_t^{(c)}|_m \right], \tag{32}
\]
where the first inequality is due to triangle inequality, the second inequality follows from the inequality that \( (a + b)^m \leq a^m + b^m \) for \( a, b \geq 0 \) when \( m \in (0, 1) \), and the third inequality is due to \( \mathbb{E} \left[ |\hat{X}_t - \hat{X}_t^{(c)}|_m \right] \leq 1 \). Hence, combining (31) and (32), we conclude that:
\[
\sup_{t>0} \mathbb{E} \left[ |X_t - \hat{X}_t^{(c)}|_m \right] < \infty \tag{33}
\]
holds only if
\[
\sup_{t>0} \mathbb{E} \left[ |X_t - \hat{X}_t^{(c)}|_m \right] < \infty. \tag{34}
\]

Accordingly, (34) is a necessary condition to satisfy (33).

Now, we find a necessary condition for (34). Let \( E_t := X_t - \hat{X}_t^{(c)} \) be estimation error for estimators \( \{\hat{X}_t^{(c)}\} \). As \( |X_t| \leq c_t \) for \( t > 0 \) and \( \hat{X}_t^{(c)} \) is integer-valued, \( E_t \) is an integer valued random variable taking values in \([-2c_t, 2c_t]\).

Using Lemma 1 for \( E_t \) being conditioned on \( Y_{1:t} \), we have:
\[
\mathbb{E} [E_t|_m | Y_{1:t} = y_{1:t}] + 1 \geq \mathbb{E} [E_t|_m | Y_{1:t} = y_{1:t}] + 1 \geq (3 + 2 \log(2c_t))^{-\rho} \times \mathbb{E} \left[ P_{E_t|Y_{1:t}}(E_t|Y_{1:t})^{-\frac{m}{m+1}} | Y_{1:t} = y_{1:t} \right]^{p+1} \tag{35}
\]
where \( P_{E_t|Y_{1:t}} \) is the conditional distribution for \( E_t \) conditioned on \( Y_{1:t} \) and the first inequality is due to that \( E_t \) is integer-valued and the second inequality is due to Lemma 1.

As \( (E_t, Y_{1:t}) \rightarrow (X_t, Y_{1:t}) \) is a bijective transformation when both \( X_t \) and \( \hat{X}_t^{(c)} \) are integer-valued, (35) becomes:
\[
\mathbb{E} [E_t|_m | Y_{1:t} = y_{1:t}] + 1 \geq (3 + 2 \log(2c_t))^{-\rho} \times \mathbb{E} \left[ P_{X_t|Y_{1:t}}(X_t|Y_{1:t})^{-\frac{m}{m+1}} | Y_{1:t} = y_{1:t} \right]^{p+1}. \tag{36}
\]

Clearly, any estimator \( \hat{X}_t \) which can take values that are outside of \([-c_t, c_t]\) is suboptimal for minimizing \( \mathbb{E} \left[ |X_t - \hat{X}_t|^m \right] \).
Taking expectations over \( Y_{1:t} \) on both sides in (36) gives:
\[
E[|E_t|^m] + 1 \geq (3 + 2 \log(2c_t))^{-\rho} \times \ldots
\]

waves
corrupted by phase noise,” IEEE Transactions on Information Theory ,
vol. 34, no. 6, pp. 1437–1448, Nov 1988.

Now, consider
\[
E\left[ P_{X_t|Y_{1:t}}(X_t|Y_{1:t})^\frac{\rho}{p(\rho+1)} \right] \geq E\left[ e^{-\frac{\rho}{p(\rho+1)}i(X_t;Y_{1:t})} \right]^{p(\rho+1)}
\times E\left[ P_{X_t|Y_{1:t}}(X_t|Y_{1:t})^\frac{\rho}{p(\rho+1)} \right]^{1-p(\rho+1)},
\]

where the inequality follows from the reverse Hölder inequality
for \( p \in (1, \infty) \) and \( i(X_t;Y_{1:t}) \) is the information density
for \( P_{X_t|Y_{1:t}} \). Then, we can get:
\[
\frac{1}{\rho} \log E\left[ P_{X_t|Y_{1:t}}(X_t|Y_{1:t})^\frac{\rho}{p(\rho+1)} \right] \geq \frac{1}{\rho} \log E\left[ e^{-\frac{\rho}{p(\rho+1)}i(X_t;Y_{1:t})} \right]^{p(\rho+1)}
+ H_\alpha(X_t),
\]

where \( \alpha = (p(\rho+1) - 1)/((p-1)(\rho+1)) \).

Combining (37) and (39):
\[
\frac{1}{\rho} \log E[|E_t|^m] + 1 \geq \frac{1}{\rho} \log E\left[ e^{-\frac{\rho}{p(\rho+1)}i(X_t;Y_{1:t})} \right]^{p(\rho+1)}
+ H_\alpha(X_t) - \log(3 + 2 \log(c_t)).
\]

As \( \lim_{t \to \infty} \log(\log(c_t))/t = 0 \),
\[
\limsup_{t \to \infty} \frac{-1}{t} \log(3 + 2 \log(2c_t)) = 0.
\]

Therefore, combining (40) and (41) implies that:
\[
\limsup_{t \to \infty} \frac{1}{\rho t} \log E[|E_t|^m] + 1 < \infty
\]
holds only if
\[
\liminf_{t \to \infty} \frac{-1}{\rho t} \log E\left[ e^{-\frac{\rho}{p(\rho+1)}i(X_t;Y_{1:t})} \right]^{p(\rho+1)}
\geq \limsup_{t \to \infty} \frac{1}{\rho t} H_\alpha(X_t).
\]

In addition, if (34) holds then (42) holds. Hence (43) is a
necessary condition for (34). Therefore, (43) is a necessary
condition for (33), i.e., \( \{X_t\} \) being order \( m \) moment trackable
though process \( \{Y_t\} \). As \( p > 1 \) is arbitrary, \( p(\rho+1) \) can be
replaced with an arbitrary \( q \) such that \( q > \rho + 1 \).

B. The Proof of Lemma 1

We have two methods to prove Lemma 1. The first method
follows the proof of Theorem 1 in [20], with the guessing
function replaced by \( A(x) \) defined in (44) below and some
other necessary changes. In the sequel, we provide a second

\[
A(x) = \begin{cases} |x| & \text{if } x \neq 0, \\ \epsilon & \text{if } x = 0, \end{cases}
\]

where \( \epsilon \) is an arbitrary positive real number. Accordingly, observe that
\[
E[|X|^p] + \epsilon^p P_X(0)
= \sum_{x \in \mathcal{X}} P_X(x)A(x)^p
\geq \left[ \sum_{x \in \mathcal{X}} P_X(x) \right]^\frac{1}{p} \left[ \sum_{x \in \mathcal{X}} A(x) \right]^{-\frac{p}{p-1}}.
\]

where the inequality is due to the reverse Hölder inequality
for \( p \in (1, \infty) \). Considering \( p = 1 + \rho \) in (45), we get
\[
E[|X|^p] + \epsilon^p P_X(0)
\geq \left[ \sum_{x \in \mathcal{X}} P_X(x) \right]^\frac{1}{1+p} \left[ \sum_{x \in \mathcal{X}} A(x) \right]^{-\rho}
\geq \left[ \sum_{x \in \mathcal{X}} P_X(x) \right]^\frac{1}{1+p} \left[ 2 + \frac{1}{\epsilon} + \log(M_{-M+}) \right]^{-\rho},
\]

where the second inequality is due to
\[
\sum_{x=-M}^{M} A(x)^{-1} = \epsilon^{-1} + \sum_{i=1}^{M_{-i}} -1 + \sum_{j=1}^{M_{+j}} -1
\leq 2 + \epsilon^{-1} + \log(M_{-M+}).
\]

Letting \( \epsilon = 1 \), combining (46) with \( P_X(0) \leq 1 \) and
\[
e\rho^H(X) = \sum_{x \in \mathcal{X}} P_X(x) \left[ \sum_{x \in \mathcal{X}} A(x) \right]^{-\frac{p}{p-1}}.
\]
we obtain (8).

C. The Proof of Lemma 2

The proof is similar to the proof of Lemma 1 where \( A(x) \)
is defined as in (44). We have
\[
E[|X|^m] + 1
\geq \left[ \sum_{x=-\infty}^{\infty} P_X(x) \right]^{1+p} \left[ \sum_{x=-\infty}^{\infty} A(x) \right]^{-\frac{p}{p-1}}
\geq \left[ \sum_{x=-\infty}^{\infty} P_X(x) \right]^{1+p} \left[ 1 + 2 \sum_{n=1}^{\infty} \frac{1}{n^p} \right]^{-p},
\]

where the first inequality follows from reverse Hölder inequality
and the second inequality follows from the choice of \( \epsilon = 1 \).

REFERENCES

[1] I. Karatzas and S. E. Shreve, Brownian Motion and Stochastic Calculus
(Graduate Texts in Mathematics). New York, NY, USA: Springer, 1991.
[2] ———, Methods of Mathematical Finance. Springer, 1998, vol. 39.
[3] G. J. Foschini and G. Vannucci, “Characterizing filtered light waves
corrupted by phase noise,” IEEE Transactions on Information Theory,
vol. 34, no. 6, pp. 1437–1448, Nov 1988.
