Multiple-Output Channel Simulation and Lossy Compression of Probability Distributions

Chak Fung Choi and Cheuk Ting Li
The Chinese University of Hong Kong
Email: cfchoi@link.cuhk.edu.hk, ctli@ie.cuhk.edu.hk

Abstract—We consider a variant of the channel simulation problem with a single input and multiple outputs, where Alice observes a probability distribution \( P \) from a set of prescribed probability distributions \( \mathcal{P} \), and sends a prefix-free codeword \( W \) to Bob to allow him to generate \( n \) i.i.d. random variables \( X_1, X_2, \ldots, X_n \), which follow the distribution \( P \). This can also be regarded as a lossy compression setting for probability distributions. This paper describes encoding schemes for three cases of \( P \): \( P \) is a distribution over positive integers, \( P \) is a continuous distribution over \([0,1]\) with a non-increasing pdf, and \( P \) is a continuous distribution over \([0,\infty)\) with a non-increasing pdf. We show that the growth rate of the expected codeword length is sub-linear in \( n \) when a power law bound or exponential tail bound is satisfied. An application of multiple-outputs channel simulation is the compression of probability distributions.

I. INTRODUCTION

The asymptotic channel simulation problem \([1], [2]\) is described as follows. Let \( \mathcal{P} = \{ P_\theta : \theta \in \mathcal{A} \} \) be a set of probability distributions indexed by \( \theta \). The encoder observes \( \theta_1, \ldots, \theta_n \) and sends a message \( M \) to the decoder. The decoder then outputs \( X_1, \ldots, X_n \). The encoder and decoder may also share common randomness. The goal is to have the conditional distribution of \( X_1, \ldots, X_n \) given \( \theta_1, \ldots, \theta_n \) to be approximately \( P_{\theta_1} \times \cdots \times P_{\theta_n} \), while minimizing the rate of the message \( M \) as \( n \to \infty \). It was shown by Bennett et al. \([1]\) that for the case with unlimited common randomness, the optimal rate is given by \( C = \max_{P_\theta} I(\theta; X) \) (where \( X \) follows the conditional distribution \( P_\theta \) given \( \theta \) i.e., the capacity of the channel \( \theta \to X \)). For the case where \( \theta \) is known to follow the distribution \( p(\theta) \), Winter \([2]\) showed that a rate of \( I(\theta; X) \) for the message, and a rate of \( H(X|\theta) \) for the common randomness suffices. Cuff \([3]\) characterized the optimal trade-off between the communication rate and the common randomness rate.

The channel simulation problem is also studied in a one-shot setting \( (n = 1) \), where the encoder observes \( \theta \) and sends a prefix-free codeword \( W \) to the decoder, which then outputs \( X \). The goal is to have \( X \) follow the conditional distribution \( P_\theta \) given \( \theta \) exactly, while minimizing the expected length \( \mathbb{E}(L(W)) \) of \( W \). Harsha et al. \([4]\) studied the case with unlimited common randomness, and showed that \( \mathbb{E}(L(W)) \leq C + (1 + \epsilon) \log(C + 1) + O(1) \) bits of codeword and \( O(\log |X| + \log |Y|) \) bits of common randomness suffices for one-shot setting (where \( C \) is the capacity of the channel \( \theta \to X \), and \( |X| \) denotes the cardinality of \( X \)). Braverman and Garg \([5]\) improved the result by eliminating the multiplicative factor \((1 + \epsilon) \). Li and El Gamal \([6]\) strengthened the result by showing that \( C + \log(C + 1) + 5 \) bits of codeword and \( \log(|X|)|Y| - 1 + 2 \) bits of common randomness suffice. The case without common randomness is studied in \([7], [8]\).

A universal setting where \( \mathcal{P} \) is the class of continuous distributions over \( \mathbb{R} \) was studied by Li and El Gamal \([9]\). In this case, it is more natural to omit the index \( \theta \) and assume the encoder observes a distribution \( P \in \mathcal{P} \), and the expected length would depend on \( P \).

This paper studies an extension of the one-shot universal channel simulation setting, called the multiple-output channel simulation setting, described as follows. The encoder observes \( P \in \mathcal{P} \) and sends a codeword \( W \in \{0,1\}^* \) from an agreed-upon prefix-free code to the decoder. The decoder then outputs \( X_1, \ldots, X_n \). There is no common randomness shared between the encoder and the decoder. The goal is to have \( X_1, \ldots, X_n \) i.i.d. following \( P \) exactly, while minimizing the expected length \( \mathbb{E}(L(W)) \). This setting is depicted in Figure 1. A straightforward approach is to apply the scheme in \([9]\) \( n \) times, resulting in an expected length that grows linearly in \( n \). In this paper, we are interested in schemes where the expected length that grows sublinearly in \( n \).

\[
P \in \mathcal{P} 
\xrightarrow{\text{Encoder}} W \in \{0,1\}^* 
\xrightarrow{\text{Decoder}} X_1, X_2, \ldots, X_n \sim P
\]

Figure 1. Multiple-output channel simulation setting.

Another approach is to have the encoder generate \( X_1, \ldots, X_n \) i.i.d. following \( P \) and encode them into \( W \), so the decoder can decode \( X_1, \ldots, X_n \). Since the decoder can perform a random shuffle on its output, the ordering of \( X_1, \ldots, X_n \) does not matter, that is, the encoder only need to encode the multiset \( \{X_1, \ldots, X_n\} \). The problem of encoding multisets was studied by Varshney and Goyal \([10], [11], [12]\), who showed that for the case where the alphabet \( X \) is finite, \( |X| \log(n+1) \) bits suffice to encode the multiset \( \{X_1, \ldots, X_n\} \). Nevertheless, this approach is inapplicable for the case where the space \( X \) is continuous, since it is impossible to encode a real number into a finite number of bits.

In this paper, we study three cases of the class of distributions \( \mathcal{P} \), where an expected length that grows sublinearly in \( n \) is possible. In Section \( \[\] \) we present a scheme for the case where \( \mathcal{P} \) is the class of distributions over positive integers. This scheme is also applicable to the problem of encoding multisets \([10], [11], [12]\). In Section \( \[\] \) we present a scheme
For the case where $P$ is the class of continuous distribution over $[0,1]$ with a non-increasing pdf. Our scheme is based on the dyadic decomposition construction in [9]. In Section IV we present a scheme for the case where $P$ is the class of continuous distribution over $[0, \infty)$ with a non-increasing pdf, which combines the two aforementioned schemes.

For an application of multiple-output channel simulation, consider the setting of lossy compression of a probability distribution. The encoder encodes $P \in \mathcal{P}$ into $W \in \{0,1\}^s$. The decoder decodes $W$ into $\hat{P}$. For the case where $P$ is the class of continuous distributions over real numbers, one method is to approximate the pdf of $P$ by a piecewise linear function $\hat{P}$ with vertices that have rational coordinates, and compress those coordinates into $W$. There are two shortcomings of this method. First, the main use of a probability distribution is to simulate random variables from it, and it is impossible to obtain samples following $P$ exactly using $W$ or $\hat{P}$ (we can only sample from $\hat{P}$ which is inexact). Second, this method generally produces $\hat{P}$ that is a biased estimate of $P$, i.e., $\mathbb{E}[\hat{P}(A)] \neq P(A)$ for some $A \subseteq \mathbb{R}$. More sophisticated kernel interpolation techniques may be used to approximate $P$, but the same problems persist.

Using multiple-output channel simulation, we can allow the decoder to obtain i.i.d. samples $X_1, \ldots, X_n$ following $P$, and produce the estimate as the empirical distribution $\hat{P}(A) = n^{-1} \sum_{i=1}^n \mathbb{1}\{X_i \in A\}$. This overcomes the two aforementioned problems. First, the decoder can obtain exact i.i.d. samples of $P$ as long as the number of samples is not greater than $n$. Second, $\hat{P}$ is an unbiased estimate of $P$, i.e., $\mathbb{E}[\hat{P}(A)] = P(A)$ for $A \subseteq \mathbb{R}$. Our scheme allows the decoder to obtain i.i.d. samples and perform statistical tests on $\hat{P}$ without the need of transmitting all information about $P$.

Throughout this paper, we assume that log is base 2. Log in base $e$ is written as $\ln$. We use the notation: $[a:b] = [a,b] \cap \mathbb{Z}$. For two bit sequences $A, B \in \{0,1\}^s$, denote their concatenation as $A||B$, and the length of $A$ as $L(A)$.

II. $P$ is a distribution over positive integers

This section develops a coding scheme for the case where $P$ is a distribution over positive integers, called difference run-length encoding scheme. We then show that when $P$ satisfies the bound $P(X > x) \leq c e^{-\lambda x}$ or $P(X > x) \leq c e^{-\lambda x}$, where $c, \lambda > 1$, the expected codeword length for encoding is $o(n)$.

We first review the Elias gamma code [13].

**Definition 1** (Elias gamma code [13]). Let $Z$ be a positive integer. The codeword $g(Z)$ is defined as

$$g(Z) = 0^n || 1 || a_{n-1} a_{n-2} \ldots a_0,$$

where $a_0 a_{n-1} \ldots a_0$ is the binary representation of $Z$. The length of the codeword $g(Z)$ is $L(g(Z)) = [2 \log Z + 1]$.

We then define the difference run-length encoding scheme.

**Definition 2** (Difference run-length encoding scheme). The encoder and decoder are described as follows:

1) The operations of the encoder are:
   a) Generate i.i.d. $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n \sim P$.
   b) Sort $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$ in ascending order such that $\tilde{X}(1) \leq \tilde{X}(2) \leq \cdots \leq \tilde{X}(n)$.
   c) Let $D_1 = \tilde{X}(1)$. For $i \in [2 : n]$, let $D_i = \tilde{X}(i) - \tilde{X}(i-1)$.
   d) Let $w_1 = g(D_1)$. For $i \in [2 : n]$, if $D_i > 0$, then let $w_i = 1||g(D_i)$, where the first bit “1” indicates that $D_i > 0$. If $D_i = 0$ and $D_{i-1} \neq 0$, the encoder finds the smallest positive integer $j_i$ such that $D_{i+j} \neq 0$ (i.e., $j_i$ is the number of consecutive zeros starting at index $i$; assume $D_{n+1} = 1$), then let $w_i = 1||g(j_1)$, where the first bit “0” indicates that $D_i = 0$ and $D_{i-1} \neq 0$. If $D_i = 0$ and $D_{i-1} = 0$, then let $w_i = 0$ (the empty sequence). Then, the encoder sends a codeword $W$, which is a series of concatenated $w_i$, i.e., $W = w_1 || w_2 || \ldots || w_n$, to the decoder.

2) The operations of the decoder are:
   a) Upon receiving $w$, the decoder uses the Elias gamma code to decode $\tilde{X}(1) = D_1$, and discard the decoded bits. Initialize $i = 2$.
   b) If the next undecoded bit is 1, the decoder discard that bit, decode $D_2$ (and discard the decoded bits), compute $\tilde{X}(i) = \tilde{X}(i-1) + D_i$, and increment $i$. If the next undecoded bit is 0, the decoder discard that bit, decode $j_i$ (and discard the decoded bits), compute $\tilde{X}(i), \tilde{X}(i+1), \ldots, \tilde{X}(i+j_{i-1}) = \tilde{X}(i-1)$, and increment $i$ by $j_i$. Repeat this step until there is no more undecoded bit.
   c) Lastly, it shuffles $\tilde{X}(1), \tilde{X}(2), \ldots, \tilde{X}(n)$ randomly and outputs them as $X_1, X_2, \ldots, X_n$.

**Example 3.** Suppose $P$ is the geometric distribution $\text{Geom}(0.7)$. Alice generates 10,000 instances of $X_i \sim \text{Geom}(0.7)$, with frequencies summarized in Table I. The codeword $w$ will be in the form $g(1)||0||g(7039)||1||g(1)||\cdots||0||g(2)||1||g(1)$. Through direct computation, the codeword length is 139, which is significantly less than $n = 10,000$.

| $X_i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-------|---|---|---|---|---|---|---|---|---|
| Freq. | 7040 | 2056 | 641 | 184 | 55 | 13 | 9 | 5 | 1 |

Table I: Table of frequencies of $X_1$ and $X_2$ for Example 3.
Theorem 4. Let $P$ be a distribution over positive integers. If $P$ satisfies the bound $P(X > x) \leq cx^{-\lambda}$ for all integer $x \geq 0$, where $c > 1$ and $\lambda > 1$, then the expected codeword length of difference run-length coding scheme for $P$ is upper bounded as

$$E(L(W)) \leq \frac{50c\lambda n^{\frac{2}{\lambda}} \log(\sqrt{n} + 1)}{\lambda - 1}.$$  

Proof: We will separate the set of indices into two parts and calculate an upper bound of expected codeword length for encoding these two parts separately. Consider the sets $U = \{i : D_i > 0\}$, $V = \{i : D_i = 0 \text{ and } D_{i-1} \neq 0\}$. Note that $w_i \neq 0$ only if $i \in U \cup V$.

Consider $i \in U$. Note that $L(w_i) = \left\lfloor 2\log D_i + 1 \right\rfloor + 1 \leq 2\log D_i + 2 \leq 2\log(2D_i + 1)$. Define $\ell(x) = 2\log(2x + 1)$, which is a concave function. We have

$$\sum_{i \in U} L(w_i) \leq \sum_{i \in U} \ell(D_i) \Rightarrow E \left[ \sum_{i \in U} \ell(D_i) \right] \leq n \ell \left( \frac{\bar{X}(n)}{n} \right) \leq n \sum_{x=0}^{\infty} P\left( \bar{X}(n) > x \right) \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right),$$

where (a) is because if $i \notin U$, then $D_i = 0$ and $\ell(D_i) = 0$. For (b), it follows by Jensen’s inequality and $\sum_{i=1}^{n} D_i = X(n)$. For (c), the equality follows by the fact that

$$E \left[ \ell \left( \frac{\bar{X}(n)}{n} \right) \right] = E \left[ \sum_{x=0}^{\infty} \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right) \right]$$

$$= E \left[ \sum_{x=0}^{\infty} \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right) 1\{\bar{X}(n) > x\} \right]$$

Consider the term $P \left( \bar{X}(n) > x \right)$, we have

$$P \left( \bar{X}(n) > x \right) \leq \frac{\sum_{i=1}^{U} (\bar{X}_i > x)}{n} \leq c_{n}x^{-\lambda}.$$  

Hence,

$$n \sum_{x=0}^{\infty} P \left( \bar{X}(n) > x \right) \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right) \leq n \sum_{x=0}^{\infty} \min(c_{n}x^{-\lambda}, 1) \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right).$$

Letting $x_0 = \left\lfloor n^{1/\lambda} \right\rfloor$, we have

$$n \sum_{x=0}^{x_0-1} \min(c_{n}x^{-\lambda}, 1) \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right) \leq n \sum_{x=0}^{x_0-1} \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right) + n \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right).$$

Consider the term $n \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} (\ell((x+1)/n) - \ell(x/n))$ in (2).

Therefore,

$$n \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right) \leq n \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right)$$

$$= \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} \ell' \left( \frac{x}{n} \right).$$

Consider the term $n \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} (\ell((x+1)/n) - \ell(x/n))$ in (2).

Therefore,

$$n \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right) \leq n \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right)$$

$$= \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} \ell' \left( \frac{x}{n} \right).$$

Hence,

$$n \sum_{x=0}^{x_0-1} \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right) + n \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} \left( \ell \left( \frac{x+1}{n} \right) - \ell \left( \frac{x}{n} \right) \right)$$

$$\leq n \ell \left( \frac{x_0}{n} \right) + \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} \ell' \left( \frac{x}{n} \right)$$

$$= n \ell \left( \frac{x_0}{n} \right) + \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} \ell' \left( \frac{x}{n} \right)$$

$$= n \ell \left( \frac{x_0}{n} \right) + \sum_{x=x_0}^{\infty} c_{n}x^{-\lambda} \ell' \left( \frac{x}{n} \right) dx$$

$$= 2n \log \left( \frac{2x_0}{n} + 1 \right) + \frac{4cnx_0^{-\lambda} \log_e}{2n + 1} + \int_{x_0}^{\infty} x \beta \left( \frac{x}{n} \right) dx$$

where (d) follows by the fact that $\sum_{x=N}^{\infty} \ell(x) \leq \ell(N) + \int_{N}^{\infty} \ell(x)dx$ when $\ell(x)$ is a decreasing function.

Consider the term $2n \log \left( 2x_0/n + 1 \right)$ in (3), we have
\[ 2n \log \left( \frac{2x_0}{n} + 1 \right) \]
\[ \leq 2n \log \left( \frac{2n^\frac{\lambda}{2} + 2}{n} + 1 \right) \]
\[ \leq 2n \left( \frac{2n^\frac{\lambda}{2} + 2}{n} \right) \log e \]
\[ = 4n^\frac{\lambda}{2} \log e + 4 \log e \]
\[ \leq 8n^\frac{\lambda}{2} \log e. \]

Consider the term \( \int_{x_0}^{n} 4cnx^{-\lambda} \log e / \left( \frac{2x}{n} + 1 \right) \, dx \) in (3). Since \( \frac{2x}{n} + 1 > 1 \),
\[ \int_{x_0}^{n} \frac{4cnx^{-\lambda} \log e}{\left( \frac{2x}{n} + 1 \right)} \, dx \]
\[ \leq \int_{x_0}^{n} \frac{4cnx^{-\lambda} \log edx}{\left( \lambda - 1 \right)} \]
\[ \leq (4cn \log e) \frac{n^\frac{\lambda}{2} - n^{\lambda-1}}{\lambda - 1} \]
\[ \leq 4c n^\frac{\lambda}{2} \log e \frac{\log e - 4cn^{2-\lambda} \log e}{\lambda - 1} \]
\[ \leq \frac{4cn x_0^{-\lambda}}{\lambda - 1}. \]

where (e) follows by the fact that \( n^\frac{\lambda}{2} \leq x_0 \leq n^\frac{\lambda}{2} + 1 \) and \( x_0^{-\lambda} \leq n^{-\lambda} \).

Consider the term \( \int_{n}^{\infty} 4cnx^{-\lambda} \log e / \left( \frac{2n}{x} + 1 \right) \, dx \) in (3).
Since \( \frac{2n}{x} + 1 > \frac{2n}{x} \), we have
\[ \int_{n}^{\infty} \frac{4cnx^{-\lambda} \log e}{\left( \frac{2n}{x} + 1 \right)} \, dx \]
\[ \leq \int_{n}^{\infty} \frac{2cn^2x^{-\lambda-1} \log edx}{\lambda - 1} \]
\[ = \frac{2cn^2 \log e}{\lambda - 1}. \]

Hence (3) can be bounded as,
\[ 2n \log \left( \frac{2x_0}{n} + 1 \right) \]
\[ + \int_{x_0}^{n} \frac{4cnx^{-\lambda} \log e}{\left( \frac{2x}{n} + 1 \right)} \, dx + \int_{n}^{\infty} \frac{4cnx^{-\lambda} \log e}{\left( \frac{2n}{x} + 1 \right)} \, dx, \]
\[ \leq 8n^\frac{\lambda}{2} \log e + 4c \log e \]
\[ + \frac{4cn^\frac{\lambda}{2} \log e}{\lambda - 1} + \frac{2cn^{2-\lambda} \log e}{\lambda} \]
\[ \leq \frac{26cn^\frac{\lambda}{2}}{\min(\lambda - 1, 1)}. \]

Therefore, \( E \left[ \sum_{i \in U} L(g(D_i)) \right] \leq 26cn^{1/\lambda} / \min(\lambda - 1, 1) \).

Consider \( i \in V \). Note that \( j_i = \min\{n \in N : i + n \in U\} \)
and \( L(w_i) = [2 \log j_i + 1] + 1 \leq \ell(j_i) \). we have
\[ E \left[ \sum_{i \in V} L(w_i) \right] \leq E \left[ \sum_{i \in V} \ell(j_i) \right]. \]

Note that \( |V| \leq \bar{X}(n) \) and \( \sum_{j_i \in V} j_i \leq n \).

\[ E \left[ \sum_{j_i \in V} \ell(j_i) \right] \]
\[ = E \left[ \sum_{j_i \in V} \ell(j_i) + \sum_{k=0}^{\bar{X}(n) - |V|} \ell(0) \right] \]
\[ \leq E \left[ \bar{X}(n) \ell \left( n \frac{E}{\bar{X}(n)} \right) \right] \]
\[ \leq E \left( \bar{X}(n) \right) \ell \left( n \frac{E}{\bar{X}(n)} \right), \] (4)

where \( (f), (g) \) follows by Jensen’s inequality.

Since \( x \ell(n/x) \) is an increasing function when \( x \geq 0 \), (4) is bounded above when \( E(\bar{X}(n)) \) is bounded above. For any non-negative integer-valued random variables \( X \), we have \( E[X] = \sum_{x=0}^{\infty} P(X > x) \). We have
\[ E \left( \bar{X}(n) \right) \]
\[ = \sum_{x=0}^{\infty} P \left( \bar{X}(n) > x \right) \]
\[ \leq \sum_{x=0}^{\infty} 1 \leq \sum_{x=0}^{\infty} \min(cn x^{-\lambda}, 1) \]
\[ = 1 + \sum_{x=x_0}^{\infty} cn x^{-\lambda} \]
\[ \leq x_0 + cn x_0^{-\lambda} + \int_{x_0}^{\infty} cn x^{-\lambda} dx \]
\[ \leq n^{\frac{\lambda}{2}} + 1 + c + \frac{cn^{\frac{\lambda}{2}}}{\lambda - 1} \]
\[ \leq 4c n^{\frac{\lambda}{2}} + 2c \leq \frac{4c n^{\frac{\lambda}{2}}}{\lambda - 1}. \]

Hence,
\[ E \left( \bar{X}(n) \right) \ell \left( n \frac{E}{\bar{X}(n)} \right) \]
\[ \leq \frac{8c n^{\frac{\lambda}{2}}}{\lambda - 1} \log \left( \frac{2n(\lambda - 1)}{4c n^{\frac{\lambda}{2}}} + 1 \right) \]
\[ \leq \frac{8c n^{\frac{\lambda}{2}}}{\lambda - 1} \log \left( n^{1 - \frac{\lambda}{2}} + 1 \right) \]
\[ \leq \frac{8c n^{\frac{\lambda}{2}}}{\lambda - 1} \log \left( 2n^{1 - \frac{\lambda}{2}} \right) \]
\[ = \frac{8c n^{\frac{\lambda}{2}}}{\lambda - 1} + 16c n^{\frac{\lambda}{2}} \log \sqrt{n} \]
\[ \leq \frac{24c n^{\frac{\lambda}{2}} \log (\sqrt{n} + 1)}{\lambda - 1}. \]

Therefore, an upper bound for the \( E \left[ \sum_{i=1}^{n} L(w_i) \right] \) is
\[
E \left[ \sum_{i \in U} L(w_i) \right] + E \left[ \sum_{i \in V} L(w_i) \right] \\
\leq \frac{26c n^2}{\min(\lambda - 1, 1)} + \frac{24c \lambda n^2 \log(\sqrt{n} + 1)}{\lambda - 1} \\
\leq \frac{50c \lambda n^2 \log(\sqrt{n} + 1)}{\lambda - 1}.
\]

\[\text{Theorem 5.}\] Let \( P \) be a distribution over positive integers. If \( P \) satisfies the bound \( P(X > x) \leq ce^{-\lambda x} \) for all integer \( x \geq 0 \), where \( c > 1 \) and \( \lambda > 0 \), then the expected codeword length of difference run-length coding scheme for \( P \) is upper bounded as

\[
E(L(W)) \leq \frac{13(2\lambda + 1)c}{\lambda} \log^2 (n + 1).
\]

Proof: By replacing

\[
P \left( \tilde{X}(n) > x \right) \leq P \left( \bigcup_{i=1}^{n} \left( \tilde{X}_i > x \right) \right) \leq cnx^{-\lambda}
\]

in (1) with

\[
P \left( \tilde{X}(n) > x \right) \leq P \left( \bigcup_{i=1}^{n} \left( \tilde{X}_i > x \right) \right) \leq cne^{-\lambda x},
\]

and replacing \( x_0 = \lceil n^{1/\lambda} \rceil \) in (2) with \( x_0 = \lceil (\ln n)/\lambda \rceil \), we can rewrite (3) to

\[
2n \log \left( \frac{2x_0}{n} + 1 \right) + \frac{4c n e^{-\lambda x_0} \log e}{\frac{2x_0}{n} + 1} + \int_{x_0}^{\infty} \frac{4c n e^{-\lambda x} \log e}{\frac{2x}{\lambda} + 1} dx.
\]

Consider the term \( 2n \log \left( \frac{2x_0}{n} + 1 \right) \) in (5), we have

\[
2n \log \left( \frac{2x_0}{n} + 1 \right) \\
\leq 2n \log \left( \frac{2 \left( \frac{\ln n}{\lambda} + 1 \right)}{n} + 1 \right) \\
\leq 2n \left( \frac{2 \left( \frac{\ln n}{\lambda} + 1 \right)}{n} \right) \log e \\
= 4 \left( \frac{(\ln 2) \log n}{\lambda} + 1 \right) \log e.
\]

Consider the term \( 4c n e^{-\lambda x_0} \log e/(2x_0/n + 1) \) in (5). Since \( 2x_0/n + 1 \geq 1 \) and \( n e^{-\lambda x_0} \leq 1 \),

\[
\frac{4c n e^{-\lambda x_0} \log e}{\frac{2x_0}{n} + 1} \leq 4c \log e.
\]

Consider the term \( \int_{x_0}^{\infty} 4c n e^{-\lambda x} \log e/(2x/n + 1) dx \) in (5). Since \( 2x_0/n + 1 \geq 1 \),

\[
\int_{x_0}^{\infty} \frac{4c n e^{-\lambda x} \log e}{\frac{2x}{n} + 1} dx \\
\leq \int_{x_0}^{\infty} \frac{4c n e^{-\lambda x} \log e}{\frac{2x}{n} + 1} dx \\
\leq 4c \log e.
\]

Therefore, \( E \left[ \sum_{i \in U} L(w_i) \right] \leq \frac{13c}{\lambda} \log^2 (n + 1) \).

For finding an upper bound of \( E \left[ \sum_{i \in V} L(w_i) \right] \), we use same argument in Theorem (4). We start with finding an upper bound for \( E \left( \tilde{X}(n) \right) \).

\[
E \left( \tilde{X}(n) \right) \\
= \sum_{x=0}^{\infty} P \left( \tilde{X}(n) > x \right) \\
\leq \sum_{x=0}^{\min(c n e^{-\lambda x}, 1)} \left( \frac{2}{\lambda} + 1 \right) e^{-\lambda x} \\
= \sum_{x=0}^{x_0-1} \left( \frac{2}{\lambda} + 1 \right) e^{-\lambda x} + \sum_{x=x_0}^{\infty} c n e^{-\lambda x} \\
\leq x_0 + c n e^{-\lambda x_0} + \int_{x_0}^{\infty} c n e^{-\lambda x} dx \\
\leq \frac{\ln n + 1 + c}{\lambda} + (2 + \frac{1}{\lambda}) c.
\]

Hence,

\[
E \left( \tilde{X}(n) \right) \leq \frac{n}{E \left( \tilde{X}(n) \right)} \\
\leq 2 \left( \frac{(\ln 2) \log n}{\lambda} + (2 + \frac{1}{\lambda}) c \right) \log \left( \frac{2n\lambda}{(\ln 2) \log n + (2\lambda + 1)c} + 1 \right) \\
\leq 2 \left( \frac{(\ln 2) \log n}{\lambda} + (2 + \frac{1}{\lambda}) c \right) \log (n + 1) \\
\leq \left( \frac{2 \ln 2 + (2\lambda + 1)c}{\lambda} \right) \log^2 (n + 1).
\]
Therefore, an upper bound for the $E\left[\sum_{i=1}^{n} L(w_{i})\right]$ is

$$E\left[\sum_{i \in U} L(w_{i})\right] + E\left[\sum_{i \in V} L(w_{i})\right] \leq \frac{4 \log n}{\lambda} + 6c \left(2 + \frac{1}{\lambda}\right) \frac{\log^2 (n + 1)}{\lambda} \leq \frac{4 + 6(2\lambda + 1)c + 2 \log 2 + (2\lambda + 1)c}{\lambda} \log^2 (n + 1) \leq \frac{13(2\lambda + 1)c + 2 \log 2 + (2\lambda + 1)c}{\lambda} \log^2 (n + 1).$$

This encoding scheme also plays an essential role in establishing a coding scheme for the case where $P$ is a continuous distribution over $[0, \infty)$ with a non-increasing pdf, which will be discussed in Section IV.

III. $P$ is a continuous distribution over $[0, 1]$ with a non-increasing pdf

We develop another coding scheme for the case where $P$ is a continuous distribution over $[0, 1]$ with a non-increasing pdf, which is another building block for the case where $P$ is a distribution over $[0, \infty)$ in Section IV. Our scheme is based on the dyadic decomposition construction in [8], [9].

Definition 6. Let $f$ be the pdf of the distribution $P$, which is a non-increasing function over $[0, \infty)$. For $k \in \mathbb{Z}_{\geq 0}$ and $a \in [0 : \max(2^{k-1} - 1, 0)]$, define the rectangle

$$R(k, a) = \left[2^{-k+1}a, 2^{-k}(2a + 1)\right) \times \left[f\left(2^{-k+1}(a + 1)\right), f\left(2^{-k}(2a + 1)\right)\right) \subseteq \mathbb{R}^2.$$

Consider the positive part of the hypograph of $f$ defined as $\text{hyph} f^+ = \{(x, y) : x \in \mathbb{R}_{\geq 0}, 0 \leq y \leq f(x)\}$. Note that $\{R(k, a)\}$ is a partition of $\text{hyph} f^+$ (except a set of measure zero) into rectangles. Every point $x$ in the interior of $\text{hyph} f^+$ is contained in only one rectangle $R(k, a)$.

Note that $k, a$ can be 0. When encoding $k, a$, we will use shifted Elias gamma code defined as follows. Let $g_{s}(x) = g(x + 1)$, where $g$ is the Elias gamma encoding function.

We are now ready to define our coding scheme for the continuous distribution $P$ over $[0, 1]$ with a non-increasing pdf.

Definition 7. The coding scheme for the case where $P$ is a continuous distribution over $[0, 1]$ with a non-increasing pdf consists of:

1) Encoder:

a) After observing $P$, the encoder generates $n$ i.i.d. points $p_1, p_2, \ldots, p_n \in \text{hyph} f^+$ uniformly over $\text{hyph} f^+$.

b) Let $U = \{(k, a) : R(k, a) \cap \{p_1, p_2, \ldots, p_n\} \neq \emptyset\}$. Assume $U = \{(k_1, a_1), (k_2, a_2), \ldots, (k_{|U|}, a_{|U|})\}$, where $(k_i, a_i)$ are ordered in lexicographic order.

c) Let $w_i = g_{s}(k_i)) || g_{s}(a_i) || g(N_i)$ for $i \in [1 : |U|]$, where $N_i = |R(k_i, a_i) \cap \{p_1, p_2, \ldots, p_n\}|$. The encoder then sends the codeword $W$, which is the concatenation of $w_i$, i.e., $W = w_1 || w_2 || \cdots || w_{|U|}$, to the decoder.

2) Decoder:

a) Upon receiving $W$, the decoder recovers $k_i, a_i, N_i$ for $i \in [1 : |U|]$.

b) For each $i$, the decoder generates $N_i$ points uniformly over $R(k_i, a_i)$. It collects all the $x$-coordinate of the generated points, shuffle these numbers uniformly at random and outputs the shuffled sequence as $X_1, X_2, \ldots, X_n$.

We present the following theorem which shows that the codeword length grows sub-linearly in $n$ when $P$ follows a non-increasing pdf.

Theorem 8. The expected codeword length of the above coding scheme for the case where $P$ is a distribution over $[0, 1]$ with a non-increasing pdf is

$$E(L(W)) \leq 92\sqrt{n}f(0)(\log(\sqrt{n}f(0) + 1)).$$

Example 9. Consider the following pdf $f$ over $[0, 1]$,

$$f(x) = \begin{cases} 2 - 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Figure 2 depicts the decomposition of this pdf into rectangles. Figure 3 depicts a log-log plot of the expected codeword length (computed by listing all rectangles with width at least $2^{-8}$) versus $n$, compared to the bound in Theorem 8. Notice that the growth rate of the expected codeword length has a similar order as our bound, and they are both sublinear (which can be observed from their slopes in the log-log plot which are less than 1).

![Figure 2](image2.png)

Figure 2. Decomposition of the distribution in Example 8

We now prove Theorem 8.

**Proof:** Let $N_{k, a} = |R(k, a) \cap \{p_1, p_2, \ldots, p_n\}|$. Since $n$ points are generated randomly and independently on the $\text{hyph} f^+$, $N_{k, a}$ is a random variable and follows a
distribution Binomial\( (n, A(k, a)) \), where \( A(k, a) \) is the area of rectangle \( R(k, a) \). More specifically, \( A(k, a) = 2^{-k} \left( f(2^{-k}(2a + 1)) - f(2^{-k+1}(a + 1)) \right) \). Also, the probability to include the triple \((k, a, N_{k,a})\) in the encoding \( W \) is equal to \( P(N_{k,a} \geq 1) \). The expected codeword length \( E[L(W)] \) can be calculated by summing the expected codeword length for all possible triples.

\[
E[L(W)] = \sum_{k=0}^{\infty} \sum_{a=0}^{\text{max}(2k-1,0)} (L(g_s(k)) + (a)) P(N_{k,a} \geq 1) + \sum_{k=0}^{\infty} \sum_{a=0}^{\text{max}(2k-1,0)} E[L(g_s(N_{k,a}))]
\]

\[
= (L(g_s(0)) + L(g_s(0))) P(N_{0,0} \geq 1) + \sum_{k=1}^{\infty} \sum_{a=0}^{2k-1-1} (L(g_s(k))) (1 - (1 - A(k, a))^n) + \sum_{k=1}^{\infty} \sum_{a=0}^{2k-1-1} (L(g_s(a))) (1 - (1 - A(k, a))^n) + E[L(g(0,0,0))] + \sum_{k=1}^{\infty} \sum_{a=0}^{2k-1-1} E[L(g_s(N_{k,a}))]. \tag{6}
\]

Note that \( (L(g_s(0)) + L(g_s(0))) P(N_{0,0} \geq 1) \leq (1+1)(1) = 2 \).

Consider the term \( 1 - (1 - A(k, a))^n \) in (6). Let \( k_0 = \left\lfloor \log \left( \sqrt{n f(0)} + 1 \right) \right\rfloor \). If \( k \leq k_0 \), then bound \( 1 - (1 - A(k, a))^n \) above by 1. Otherwise, bound \( 1 - (1 - A(k, a))^n \) above by \( nA(k, a) \). Consider the term \( \sum_{k=1}^{\infty} \sum_{a=0}^{2k-1-1} (L(g_s(k)) + L(g_s(a))) (1 - (1 - A(k, a))^n) \), we have

\[
\sum_{k=1}^{\infty} \sum_{a=0}^{2k-1-1} (L(g_s(k)) + L(g_s(a))) (1 - (1 - A(k, a))^n) \leq \sum_{k=1}^{k_0} \sum_{a=0}^{2k-1-1} (L(g_s(k)) + L(g_s(a))) \]

\[
+ \sum_{k=k_0+1}^{\infty} \sum_{a=0}^{2k-1-1} (L(g_s(k)) + L(g_s(a))) nA(k, a).
\]

Consider the term \( \sum_{k=1}^{k_0} \sum_{a=0}^{2k-1-1} L(g_s(k)) \) in (6). We have

\[
\sum_{k=1}^{k_0} \sum_{a=0}^{2k-1-1} L(g_s(k)) \leq \sum_{k=1}^{k_0} 2^k \log(2k + 3) \]

\[
\leq \sum_{k=1}^{k_0} (k + 1)(2k + 1) \leq 2k_0 + 2k_0 \leq 4\left( \sqrt{n f(0)} + 1 \right) \log(\sqrt{n f(0)} + 1) \]

\[
\leq 8\sqrt{n f(0)} \log(\sqrt{n f(0)} + 1),
\]

where \( (a) \) follows from the fact that \( \log(2k + 3) \leq 2k + 2 \) when \( k \geq 1 \).

Consider the term \( \sum_{k=1}^{k_0} \sum_{a=0}^{2k-1-1} L(g_s(a)) \) in (6). We have

\[
\sum_{k=1}^{k_0} \sum_{a=0}^{2k-1-1} L(g_s(a)) \leq \sum_{k=1}^{k_0} 2^k \log(2k + 3) \leq \sum_{k=1}^{k_0} 2^k \log(2k + 2)
\]

\[
= \sum_{k=1}^{k_0} (k + 2)2^k
\]

\[
\leq 2(2^{k_0}k_0 + 2^{k_0} - 1) \leq 4\left( \sqrt{n f(0)} + 1 \right) \log(\sqrt{n f(0)} + 1) \leq 8\sqrt{n f(0)} \log(\sqrt{n f(0)} + 1).
\]

Consider the term \( \sum_{k=k_0+1}^{\infty} \sum_{a=0}^{2k-1-1} g_s(k) nA(k, a) \) in (6).

We have

\[
\sum_{k=k+1}^{\infty} \sum_{a=0}^{2k-1-1} g_s(k) nA(k, a) \leq n \sum_{k=k_0+1}^{\infty} \sum_{a=0}^{2k-1-1} \log(2k + 3) 2^{-k+1} \left( f(2^{-k}(2a + 1)) - f(2^{-k+1}(a + 1)) \right)
\]

\[
- f(2^{-k+1}(a + 1)).
\]
\[ (b) \quad n \sum_{k=k_0+1}^{\infty} \log(2k + 3) 2^{-k+1} f(0) \]
\[ \leq n \sum_{k=k_0+1}^{\infty} f(0) 2^{-k+1} (k_0 + 3) \]
\[ \leq n f(0) 2^{-k_0+3} k_0 \]
\[ \leq n f(0) 2^{4 \log(\sqrt{n f(0) + 1})} \log(\sqrt{n f(0) + 1}) \]
\[ = \frac{16 n f(0)}{n f(0) + 1} \log(\sqrt{n f(0) + 1}) \]
\[ \leq 16 \sqrt{n f(0)} \log(\sqrt{n f(0) + 1}), \]

where \((b)\) follow by the fact that \(\sum_{a=0}^{2^{k-1} - 1} f(2^{-k}(2a + 1)) - f(2^{-k+1}(a + 1)) \leq f(2^{-k}) \leq f(0)\). For \((c)\), \(k + 1 \geq \log(2k + 3)\), when \(k \geq 2\).

Consider the term \(\sum_{k=k_0+1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} g_a(k) a N(1, a)\) in \((6)\). We have

\[ \sum_{k=k_0+1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} g_a(k) a N(1, a) \]
\[ = n \sum_{k=k_0+1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} \log(2a + 3) 2^{1-k} \left( f(2^{-k}(2a + 1)) - f(2^{-k+1}(a + 1)) \right) \]
\[ \leq n \sum_{k=k_0+1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} \log(2a + 3) 2^{1-k} \left( f(2^{-k}(2a + 1)) - f(2^{-k+1}(a + 1)) \right) \]
\[ \leq 16 \sqrt{n f(0)} \log(\sqrt{n f(0) + 1}), \]

where \((d)\) follows from the fact that \(\log(2a + 3) \leq \log(2k + 1) \leq \log(2^{k+1})\) for \(a \in [0; 2^{k-1} - 1]\).

Consider the term \(E[L(g(N_0, 0))]\) in \((6)\). We have

\[ E[L(g(N_0, 0))] \]
\[ \leq E[2 \log(2N_{k, a} + 1)] \]
\[ \leq n \sum_{k=k_0+1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} E[2 \log(2N_{k, a} + 1)] \]
\[ \leq 2 \log(2E(N_{k, a} + 1)) \]
\[ = 2 \log(2n f(1) + 1) \]
\[ \leq 2 \log(3n f(0)) \]
\[ \leq 4 + 4 \log(\sqrt{n f(0) + 1}), \]

where \((e)\) follows from Jensen’s inequality.

Consider the term \(\sum_{k=1}^{k_1} 2^k \log\left(\frac{8n f(0)}{2^{2k}}\right)\) in \((6)\).

We have

\[ \sum_{k=1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} E[L(g(N_{k, a}))] \]
\[ \leq \sum_{k=1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} E[2 \log(2N_{k, a} + 1)] \]
\[ \leq \sum_{k=1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} 2 \log(2E(N_{k, a} + 1)) \]
\[ = \sum_{k=1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} 2 \log(2n f(1) + 1) \]
\[ \leq \sum_{k=1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} 2 \log(3n f(0)) \]
\[ \leq \sum_{k=1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} 4 + 4 \log(\sqrt{n f(0) + 1}), \]

where \((f)\) follows from the fact that \(E(N_{k, a}) \geq 1\) when \(k \leq k_1\) and \(2^k \log\left(\frac{4n f(0)}{2^{2k}} + 1\right)\) is decreasing when \(k \geq k_1 + 1\).

Consider the term \(\sum_{k=1}^{k_1} 2^k \log\left(\frac{8n f(0)}{2^{2k}}\right)\) in \((6)\).

We have

\[ \sum_{k=1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} E[L(g(N_{k, a}))] \]
\[ \leq \sum_{k=1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} E[2 \log(2N_{k, a} + 1)] \]
\[ \leq \sum_{k=1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} 2 \log(2E(N_{k, a} + 1)) \]
\[ = \sum_{k=1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} 2 \log(2n f(1) + 1) \]
\[ \leq \sum_{k=1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} 2 \log(3n f(0)) \]
\[ \leq \sum_{k=1}^{\infty} \sum_{a=0}^{2^{k-1} - 1} 4 + 4 \log(\sqrt{n f(0) + 1}), \]

where \((i)\) follows from the fact that \(\log(\sqrt{4n f(0)}) - 1 \leq k_1 \leq \log(\sqrt{4n f(0)})\), and thus \(\sqrt{n f(0)} \leq 2k_1 \leq 2 \sqrt{n f(0)}\).
We present the following theorem which shows that the codeword length grows sub-linearly in \( n \) when \( P \) follows a non-increasing pdf and satisfies a power law bound.

**Theorem 11.** The expected codeword length of the above coding scheme, for the case where \( P \) is a distribution over \([0, 1]\) with a non-increasing pdf, and \( P \) satisfies the bound 
\[
P(X > x) \leq c x^{-\lambda}
\]
for all \( x \in [0, 1] \), is bounded above as
\[
E(L(W)) \leq \frac{418c(\lambda + 1) \max(n^{1/\lambda}, \sqrt{n}) \max(\sqrt{f(0)}, 1)}{\min(\lambda - 1, 1)} \cdot \log(\sqrt{n} \max(f(0), 1) + 1).
\]

Before proving the theorem, we review the concept of majorization for non-increasing functions \([13]\).

**Definition 12** (Majorization). Let \( f, g \) be two continuous non-increasing functions over \([0, \infty)\). It is said that \( f \) is majorized by \( g \), denoted by \( f \prec g \), if \( \int_0^x f(t) dt \leq \int_0^x g(t) dt \) for any \( x \geq 0 \).

We state the following equivalent characterization of majorization, which is proved in \([14]\) Theorem 2.5.

**Lemma 13** \([14]\). Let \( f, g \) be two continuous functions. \( f \prec g \) if and only if
\[
\int_0^\infty \phi(f(t)) dt \geq \int_0^\infty \phi(g(t)) dt
\]
for all concave function \( \phi : \mathbb{R} \to \mathbb{R} \).

Before we prove Theorem 11, we show the following lemma.

**Lemma 14.** Let \( f \) be a non-increasing pdf over \([0, \infty)\) that satisfies the bound \( \int_0^x f(t) dt \leq c x^{-\lambda} \) for any \( x \in [0, \infty) \). Let \( f^* : [0, \infty) \to \mathbb{R} \) be a pdf defined as
\[
f^*(x) = \begin{cases} c \lambda x_0^{-\lambda-1}, & x \leq t_0 \\ c \lambda x^{-\lambda-1}, & x > t_0, \end{cases}
\]
where \( t_0 = (c(\lambda + 1))^{1/\lambda} \). Then \( f \prec f^* \).

**Proof:** Note that \( \int_0^\infty f^*(x) dx = 1 \). Further note that \( f(0) \geq f^*(0) \) otherwise \( \int_0^\infty f(x) dx < \int_0^\infty f^*(x) dx = 1 \). Suppose \( \int_0^a f(x) dx < \int_0^a f^*(x) dx \) for some \( a \). Suppose \( a \leq t_0 \). Note that \( f(a) < f^*(a) \) and thus \( f(x) < f^*(x) \) when \( x \geq a \). Thus, \( \int_0^\infty f(x) dx = \int_0^a f(x) dx + \int_a^\infty f(x) dx < \int_0^a f^*(x) dx + \int_a^\infty f^*(x) dx = 1 \). Contradiction arises. Suppose \( a > t_0 \). Note that \( \int_0^\infty f(x) dx \leq c a^{-\lambda} = \int_0^\infty f^*(x) dx \). Thus, \( \int_0^\infty f(x) dx = \int_0^a f(x) dx + \int_a^\infty f(x) dx < \int_0^a f^*(x) dx + \int_a^\infty f^*(x) dx = 1 \). Contradiction arises. Therefore, \( f \prec f^* \).

We now present the proof of Theorem 11.

**Proof of Theorem 11** Let \( f \) be the pdf of the distribution \( P \), and \( X \) be a random variable following the distribution of \( P \). Consider the distribution of \( |X| \). We have, for integer \( x \geq 0 \),
\[
P(|X| > x) = P(X > x) \leq c x^{-\lambda}.
\]
By Theorem 4 we have
\[ \mathbb{E}(L(W_{\text{int}})) \leq \frac{50c\lambda n^2 \log(\sqrt{n} + 1)}{\lambda - 1}. \]

Consider \( W_i \) for \( i \geq 1 \). By Theorem 8
\[
\begin{align*}
\mathbb{E}(L(W_i)) &\leq \mathbb{E}\left( \{n_i \geq 1\} \cdot 92\sqrt{n_i f_i(0)(\log(\sqrt{n_i f_i(0)} + 1))} \right) \\
&= \mathbb{E}\left( 92\sqrt{n_i f_i(0)(\log(\sqrt{n_i f_i(0)} + 1))} \right) \\
&\leq 92\sqrt{\mathbb{E}(n_i f_i(0)(\log(\mathbb{E}(n_i f_i(0)) + 1))} \\
&\leq 92\sqrt{n f(i-1)(\log(\sqrt{n f(i-1)} + 1)),}
\end{align*}
\]
where (a) is by Jensen’s inequality and the concavity of \( \sqrt{t} \log(\sqrt{t} + 1) \), and (b) is by
\[
\mathbb{E}(n_i f_i(0)) = \left\{ n \int_{t-1}^{t} f(t) dt \right\} \frac{f(i-1)}{\int_{t-1}^{t} f(t) dt} = n f(i-1).
\]

Hence, we have
\[
\sum_{i=1}^{\infty} \mathbb{E}(L(W_i)) \\
\leq \sum_{i=1}^{\infty} 92\sqrt{\mathbb{E}(n f(i-1)(\log(\sqrt{n f(i-1)} + 1)))} \\
\leq 92\sqrt{\mathbb{E}(n f(0)(\log(\sqrt{n f(0)} + 1))} \\
+ \int_{0}^{\infty} 92\sqrt{n f(x)} \left( \log(\sqrt{n f(x)} + 1) \right) dx.
\]

Consider the term \( \int_{0}^{\infty} 72\sqrt{n f(x)} \left( \log(\sqrt{n f(x)} + 1) \right) dx \). By Lemma 13 Lemma 14 and the concavity of \( \sqrt{t} \log(\sqrt{t} + 1) \),
\[
\int_{0}^{\infty} 92\sqrt{n f(x)} \left( \log(\sqrt{n f(x)} + 1) \right) dx \\
\leq \int_{0}^{\infty} 92\sqrt{n f^*(x)} \left( \log(\sqrt{n f^*(x)} + 1) \right) dx \\
= \int_{0}^{1} 92\sqrt{n f^*(x)} \left( \log(\sqrt{n f^*(x)} + 1) \right) dx \\
+ \int_{1}^{\infty} 92\sqrt{n f^*(x)} \left( \log(\sqrt{n f^*(x)} + 1) \right) dx \\
\leq 92t_0 \sqrt{n f^*(0)} \left( \log(\sqrt{n f^*(0)} + 1) \right) \\
+ 92 \int_{0}^{\infty} \sqrt{n c x^2 - \lambda x - 1} \left( \log(\sqrt{n f^*(0)} + 1) \right) dx \\
(\geq) \leq 92t_0 \sqrt{n} \left( \log(\sqrt{n} + 1) \right) \\
+ 92 \int_{0}^{\infty} \sqrt{n c x^2 - \lambda x - 1} \left( \log(\sqrt{n} + 1) \right) dx \\
= 92 (c + 1) \frac{\sqrt{n} \log(\sqrt{n} + 1)}{(\lambda - 1) (c + 1)} \\
+ 184 \sqrt{n c \lambda \log(\sqrt{n} + 1)} \\
\leq 92 (c + 1) \frac{\sqrt{n} \log(\sqrt{n} + 1)}{(\lambda - 1) (c + 1)}
\]
where (c) follows from \( f^*(0) = c (c + 1)^{-1 - \frac{1}{\lambda}} \leq 1 \). Hence,
\[
\sum_{i=1}^{\infty} \mathbb{E}(L(W_i)) \\
\leq 92\sqrt{n f(0)(\log(\sqrt{n f(0)} + 1))} \\
+ 276 (c + 1)^{1/\lambda} \sqrt{n} \log(\sqrt{n} + 1) \min\left(\lambda - 1, 1\right).
\]

Therefore, the expected codeword length grows sublinearly.

Our coding scheme described in Definition 10 uses the coding scheme in Definition 7 as a building block. Therefore, if we use Theorem 8 in our analysis, then even if we assume a stronger tail bound, such as exponential tail bound \( P(X > x) \leq e^{-\lambda x} \), the order of the growth of expected codeword length cannot be better than \( O((\sqrt{n f(0)})(\log(\sqrt{n f(0)} + 1))) \).

V. CONCLUSION AND DISCUSSION

In this paper, we introduced a new problem in channel simulation called the multiple-output channel simulation. We also describe encoding schemes for three classes of probability distributions and show that the growth rate of the expected codeword length is sub-linear in \( n \) when a power bound or exponential bound is satisfied. An application of multiple-outputs channel simulation is the compression of probability distributions.

We list some potential extensions of our result. First, it may be possible to generalize the result to more classes of probability distributions, such as unimodal distributions over the real line. Second, since this paper only focus on upper bounds of our codeword length, lower bounds may also be derived in the future in order to show tightness. Third, we may also consider the case where common randomness is available to the encoder and decoder.
VI. ACKNOWLEDGMENT

The authors acknowledge support from the Direct Grant for Research, The Chinese University of Hong Kong.

REFERENCES

[1] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, “Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem,” IEEE Transactions on Information Theory, vol. 48, no. 10, pp. 2637–2655, 2002.

[2] A. Winter, “Compression of sources of probability distributions and density operators,” arXiv preprint quant-ph/0208131, 2002.

[3] P. Cuff, “Distributed channel synthesis,” IEEE Transactions on Information Theory, vol. 59, no. 11, pp. 7071–7096, 2013.

[4] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, “The communication complexity of correlation,” in Twenty-Second Annual IEEE Conference on Computational Complexity (CCC’07), pp. 10–23, IEEE, 2007.

[5] M. Braverman and A. Garg, “Public vs private coin in bounded-round information,” in International Colloquium on Automata, Languages, and Programming, pp. 502–513, Springer, 2014.

[6] C. T. Li and A. El Gamal, “Strong functional representation lemma and applications to coding theorems,” IEEE Transactions on Information Theory, vol. 64, no. 11, pp. 6967–6978, 2018.

[7] G. R. Kumar, C. T. Li, and A. El Gamal, “Exact common information,” in 2014 IEEE International Symposium on Information Theory, pp. 161–165, IEEE, 2014.

[8] C. T. Li and A. El Gamal, “Distributed simulation of continuous random variables,” IEEE Transactions on Information Theory, vol. 63, no. 10, pp. 6329–6343, 2017.

[9] C. T. Li and A. El Gamal, “A universal coding scheme for remote generation of continuous random variables,” IEEE Transactions on Information Theory, vol. 64, no. 4, pp. 2583–2592, 2018.

[10] L. R. Varshney and V. K. Goyal, “Ordered and disordered source coding,” sort, vol. 1, p. 1, 2006.

[11] L. R. Varshney and V. K. Goyal, “Toward a source coding theory for sets,” in Data Compression Conference (DCC’06), pp. 13–22, IEEE, 2006.

[12] L. R. Varshney and V. K. Goyal, “Benefiting from disorder: Source coding for unordered data,” arXiv preprint arXiv:0708.2310, 2007.

[13] P. Elias, “Universal codeword sets and representations of the integers,” IEEE transactions on information theory, vol. 21, no. 2, pp. 194–203, 1975.

[14] K.-M. Chong, “Some extensions of a theorem of hardy, littlewood and polya and their applications,” Canadian Journal of Mathematics, vol. 26, no. 6, pp. 1321–1340, 1974.