Mutual Information for a Deletion Channel

Michael Drmota
Inst. Discrete Mathematics and Geometry
TU Wien
A-1040 Wien, Austria
Email: michael.drmota@tuwien.ac.at

Wojciech Szpankowski
Department of Computer Science
Purdue University
West Lafayette, IN, USA
Email: spa@cs.purdue.edu

Krishnamurthy Viswanathan
Hewlett-Packard Laboratories
Palo Alto, CA, USA
Email: krishnamurthy.viswanathan@hp.com

I. Introduction

A deletion channel with parameter d takes a binary sequence $x := x^n = x_1 \cdots x_n$ where $x_i \in \mathcal{A} = \{0,1\}$ as input and deletes each symbol in the sequence independently with probability d. The output of such a channel is them a *subsequence* $Y = Y(x) = x_{i_1}...x_{i_M}$ of $x = x^n$, where M follows the binomial distribution $\mathrm{Bi}(n,(1-d))$, and the indices $i_1,...,i_M$ correspond to the bits that are not deleted. Despite significant effort [2], [4], [8], [9], [10], [11], [13] the mutual information of the deletion channel and its capacity are still unknown. Our goal is to provide a more detailed charactetrization of the mutual information and the channel capacity for memoryless sources (however, extension to Markovian sources is possible) for two special cases: $d \to 1$ and $d \to 0$. The latter case was already discussed in [9], [8]. We accomplish it by relating the problem to the so called *hidden pattern matching* analyzed recently in [1], [6].

The channel capacity of the deletion channel is then

$$C(d) = \lim_{n\to\infty} \frac{1}{n} \sup_P I(X_1^n; Y(X_1^n)),$$

where the supremum is taken over all stationary ergodic processes, and $I(X_1^n; Y(X_1^n))$ is the mutual information of the deletion channel. It is well known [3] that this limit exists and there exist many bounds for the capacity; see the survey article by Mitzenmacher [10].

In this paper, we first represent the mutual information of a deletion channel in terms of the count $\Omega_X(\omega)$ that enumerates the number of occurrences of a word w as a *subsequence* (i.e., not consecutive symbols) of X.

Theorem 1. *For any source, the mutual information satisfies*

$$I(X_1^n;Y(X_1^n)) = \sum_w d^{n-|w|}(1-d)^{|w|} \left(\mathbb{E}[\Omega_{X_1^n}(w)\log\Omega_{X_1^n}(w)] - \mathbb{E}[\Omega_{X_1^n}(w)]\log\mathbb{E}[\Omega_{X_1^n}(w)]\right), \quad (1)$$

where the sum is taken over all words w of length $|w| \le n$.

Then, we focus on memoryless sources, however, most of our results are valid for larger classes of (Markovian) sources. By Theorem 1 we have $I(X_1^n;Y(X_1^n)) = S_1(X_1^n, Y(X_1^n)) - S_2(X_1^n, Y(X_1^n)) := S_1 - S_2$ where

$$S_1 = \sum_w d^{n-|w|}(1-d)^{|w|}\mathbb{E}[\Omega_{X_1^n}(w)\log\Omega_{X_1^n}(w)], \quad (2)$$

$$S_2 = \sum_w d^{n-|w|}(1-d)^{|w|}\mathbb{E}[\Omega_{X_1^n}(w)]\log\mathbb{E}[\Omega_{X_1^n}(w)]. \quad (3)$$

Theorem 2. *Suppose that X is generated by a memoryless source with parameter p. Then the limit $I(d,p) = \lim_{n\to\infty}\frac{1}{n}I(X_1^n;Y(X_1^n))$ as well as the (non-negative) limits*

$$\lambda(d,p) = \lim_{n\to\infty}\frac{1}{n}S_1(X_1^n, Y(X_1^n))$$

$$\lim_{n\to\infty}\frac{1}{n}S_2(X_1^n, Y(X_1^n)) = H(1-d) - (1-d)H(p)$$

exist, and $I(d,p) = \lambda(d,p) + (1-d)H(p) - H(1-d)$. Furthermore,

$$I(d,p) = \inf_{n\ge1}\frac{1}{n}I(X_1^n;Y(X_1^n)),\ \lambda(d,p) = \sup_{n\ge1}\frac{1}{n}S_1(X_1^n, Y(X_1^n)).$$

where $H(p) = -p\log p - (1-p)\log(1-p)$ is the entropy rate.

These representations lead to $I(d,p) \le I(X_1^1;Y(X_1^1)) = H(p)(1-d)$ which is asymptotically optimal for $d\to0$ but not for $d\to1$, as our next result (Theorem 3) shows. This also implies that $\lambda(d,p)\le H(1-d)$. Recall, that for $d\to1$ it is just known that $C(d) = \Theta(1-d)$ [10], [11], [4]. Our next result shows that – in contrast to the case $d\to0$ – memoryless sources are not (asymptotically) optimal as $d\to1$.

Theorem 3. *Suppose that X is generated by a memoryless source with parameter p. Then we have the upper bound*

$$I(d,p) \le C(1-d)^{4/3}\log\frac{1}{1-d}$$

as $d\to1$, where the constant $C>0$ is absolute.

Finally we demonstrate the strength of our method by reproving Kanoria and Montenari's expansion for $I(d,p)$ for $d\to0$ leading to $C(d) = I(d/1/2) + O(d^{3/2-\varepsilon}) == 1 + d\log d - Ad + O(d^{3/2-\varepsilon})$ (cf Theorem 4), where $A = \log(2e) - \sum_{\ell\ge1}2^{-\ell-1}\ell\log\ell$. Observe that the symmetric memoryless source is asymptotically optimal in this regime.

II. General Sources

In this section we first derive a novel formula for the mutual information and then present a simple proof of the upper bound $C(d)\le 1-d$ for general sources.

A. Proof of Theorem 1

Let $X = X_1^n$ be a random sequence (over the binary alphabet $\mathcal{A} = \{0, 1\}$) and $w = w_1 w_2 \ldots w_m$ a given word (over $\mathcal{A}$) of length $|w| = m \leq n$. We denote by $\Omega_X(\omega)$ the number of occurrences of w as a subsequence of X, that is,

$$\Omega_X(\omega) = \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} \mathbf{I}_{[X_{i_1} = w_1]} \mathbf{I}_{[X_{i_2} = w_2]} \cdots \mathbf{I}_{[X_{i_m} = w_m]}, \tag{4}$$

where $\mathbf{I}_A = 1$ if A occurs and zero otherwise. The problem of counting subsequences in a text is known as the *hidden pattern matching* problem and was studied in [1], [6].

The hidden pattern matching is related to the the deletion channel problem by the following observation

$$P(Y(X_1^n) = w | X_1^n) = \Omega_{X_1^n}(w) d^{n-|w|}(1-d)^{|w|}. \tag{5}$$

We first compute the expectation of $\Omega_X(w)$.

Lemma 1. *Suppose that $X = X_1^n$ is a general source. Then*

$$\mathbb{E}\left[\Omega_{X_1^n}(w)\right] = \binom{n}{|w|} \overline{P}_n(w),$$

where

$$\overline{P}_n(w) = \frac{1}{\binom{n}{|w|}} \sum_{i_1 < \cdots < i_m} P(X_{i_1} = w_1, X_{i_2} = w_2, X_{i_m} = w_m)$$

with $\sum_{|w|=m} \overline{P}_n(w) = 1$. In particular, if X is memoryless, then $\overline{P}_n(w) = P(w)$ where $P(w)$ is the probability that $X_1 X_2 ... X_{|w|} = w$ (see [1] for dynamic sources).

Proof: By (4) we have $\mathbb{E}\left[\Omega_X(w)\right] = \sum_{1 \leq i_1 < i_2 < \cdots \leq i_m \leq n} P(X_{i_1} = w_1, X_{i_2} = w_2, \cdots, X_{i_m} = w_m) = \binom{n}{|w|} \overline{P}_n(w)$. ■

Proof of Theorem 1. The starting point is relation (5). Then we use $I(X;Y) = H(Y) - H(Y|X)$. We first compute $H(Y)$. Observe that $P(Y = w) = \sum_{x \in \mathcal{A}^n} P(X = x) \Omega_x(w) d^{n-|w|}(1-d)^{|w|}$ which leads to

$$H(Y) = -\sum_w d^{n-|w|}(1-d)^{|w|} \Big(\mathbb{E}[\Omega_{X_1^n}(w)] \log \mathbb{E}[\Omega_{X_1^n}(w)] - \mathbb{E}[\Omega_{X_1^n}(w)] \log (d^{n-|w|}(1-d)^{|w|}) \Big).$$

Next, we estimate the conditional entropy $H(Y|X)$. Notice that for $x \in \mathcal{A}^n$ and $y \in \mathcal{A}^m$ we have $P(x, y) = P(x) \Omega_x(y) d^{n-m}(1-d)^m$. Combining with (5) we find

$$H(Y|X) = -\sum_w d^{n-|w|}(1-d)^{|w|} \Big(\mathbb{E}[\Omega_X(w) \log \Omega_X(w)] + \mathbb{E}[\Omega_X(w)] \log d^{n-|w|}(1-d)^{|w|} \Big).$$

Putting all together we prove (1).

B. Upper Bound for the Capacity

In this section we recall an upper bound for the capacity of the deletion channel that also follows from the capacity of the erasure channel (e.g., see [2]). We provide a direct proof.

Lemma 2. *For every binary source $X = X_1^n$ the mutual information of the deletion channel with deletion probability d is bounded by $I(X_1^n; Y(X_1^n)) \leq n(1-d)$.*

Proof: By Theorem 1, we can write $I(X_1^n; Y(X_1^n)) = S_1 - S_2$ where S_1 and S_2 are defined in (2)–(3). Since $\Omega_X(w) \leq \binom{n}{|w|}$ we first have

$$S_1 \leq \sum_w d^{n-|w|}(1-d)^{|w|} \log \binom{n}{|w|} \mathbb{E}[\Omega_X(w)].$$

Furthermore

$$I(X_1^n; Y(X_1^n)) \leq -\sum_w d^{n-|w|}(1-d)^{|w|} \binom{n}{|w|} \overline{P}_n(w) \log \overline{P}_n(w)$$

$$= \sum_{m=0}^n d^{n-m}(1-d)^m \binom{n}{m} \sum_{|w|=m} \overline{P}_n(w) \log(1/\overline{P}_n(w)).$$

Since $\overline{P}_n(w)$ is a probability distribution for word w of length m we have $\sum_{|w|=m} \overline{P}_n(w) \log(1/\overline{P}_n(w)) \leq \log 2^m = m$ and consequently

$$I(X_1^n; Y(X_1^n)) \leq \sum_{m=0}^n \sum_{|w|=m} d^{n-m}(1-d)^m \binom{n}{m} m = n \cdot (1-d).$$

This completes the proof, and in conclusion we establish the upper bound $C \leq 1 - d$ for the capacity. ■

III. Memoryless Sources

We now assume that the source is memoryless over a binary alphabet with p being the probability of emitting a "0". The main goal of this section is to prove Theorems 2 and 3.

A. Proof of Theorem 2

The next property follows from the definition $\Omega_X(w)$.

Lemma 3. *We have the following stochastic recurrence relation:*

$$\Omega_{X_1^{n+k}}(w) = \sum_{w_1 w_2 = w} \Omega_{X_1^n}(w_1) \Omega_{X_{n+1}^{n+k}}(w_2), \tag{6}$$

where the sum is taken over all pairs w_1, w_2 such that the concatenation $w_1 w_2$ equals w.

Lemma 4. *Let z_m and a_m, $1 \leq m \leq M$, be non-negative numbers. Then we have*

$$\sum_{m=1}^M z_m \log \frac{\sum_{m=1}^M z_m}{\sum_{m=1}^M a_m} \leq \sum_{m=1}^M z_m \log \frac{z_m}{a_m}. \tag{7}$$

Proof: Apply the inequality $\log x \leq x - 1$. ■

Lemma 5. *Suppose that X is generated by a memoryless source with parameter p. Then*

$$I(X_1^{n+k}; Y(X_1^{n+k})) \leq I(X_1^n; Y(X_1^n)) + I(X_1^k; Y(X_1^k)).$$

Proof: By applying (7) for the sum (6) we obtain

$$\Omega_{X_1^{n+k}}(w)\log\Omega_{X_1^{n+k}}(w)-\Omega_{X_1^{n+k}}(w)\log\mathbb{E}\left[\Omega_{X_1^{n+k}}(w)\right]$$
$$=\sum_{w_1w_2=w}\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2)$$
$$\times\left(\log\left(\sum_{w_1w_2=w}\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2)\right)\right.$$
$$\left.-\log\left(\sum_{w_1w_2=w}\mathbb{E}\left[\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2)\right]\right)\right)$$
$$\le\sum_{w_1w_2=w}\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2)\left(\log\left(\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2)\right)\right.$$
$$\left.-\log\mathbb{E}\left[\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2)\right]\right)$$
$$=\sum_{w_1w_2=w}\Omega_{X_1^n}(w_1)\left(\log\Omega_{X_1^n}(w_1)-\log\mathbb{E}\left[\Omega_{X_1^n}(w_1)\right]\right)\Omega_{X_{n+1}^{n+k}}(w_2)$$
$$+\sum_{w_1w_2=w}\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2)\left(\log\Omega_{X_{n+1}^{n+k}}(w_2)\right.$$
$$\left.-\log\mathbb{E}\left[\Omega_{X_{n+1}^{n+k}}(w_2)\right]\right).$$

Let now $c_n=I(X_1^n;Y(X_1^n))$. Then

$$c_{n+k}=\sum_w d^{n+k-|w|}(1-d)^{|w|}\left(\mathbb{E}\left[\Omega_{X_1^{n+k}}(w)\log\Omega_{X_1^{n+k}}(w)\right]\right.$$
$$\left.-\mathbb{E}\left[\Omega_{X_1^{n+k}}(w)\right]\log\mathbb{E}\left[\Omega_{X_1^{n+k}}(w)\right]\right).$$

Hence by taking expectations of the above and by using the relation

$$1=\sum_{w_1}d^{n-|w_1|}(1-d)^{|w_1|}\mathbb{E}[\Omega_{X_1^n}(w_1)]$$
$$=\sum_w d^{n-|w|}(1-d)^{|w|}\binom{n}{|w|}P_n(w)=\sum_{\ell=0}^{n}d^{n-\ell}(1-d)^{\ell}\binom{n}{\ell}$$

(and a similar relation for the sum over w_2) we immediately derive $c_{n+k}\le c_n+c_k$. Note that we have used the property that X_1^n and X_{n+1}^{n+k} are independent and that X_{n+1}^{n+k} have the same distribution as X_1^k. ■

By Fekete's lemma [12] the following corollary follows.

Corollary 1. *For memoryless sources we have* $I(d,p)=\inf_{n\ge1}\frac{1}{n}I(X_1^n;Y(X_1^n))$.

In particular we have $I(d,p)\le\frac{1}{n}I(X_1^n;Y(X_1^n))$ for all $n\ge1$. If we apply this for $n=1,2$ we find

$$I(d,p)\le(1-d)H(p),$$
$$I(d,p)\le(1-d)(H(p)+p^2+q^2-1).$$

For example, by looking at the second bound we observe that $\sup_{0\le p\le1}I(d,p)\le\frac{1-d}{2}$ which implies that memoryless sources do not meet the general upper bound $1-d$ for $d\to1$. Actually we will show that $\sup_{0\le p\le1}I(d,p)$ is much smaller as $d\to1$ (Theorem 3).

We now prove Theorem 2. As above, we write $I(X_1^n;Y(X_1^n))=S_1-S_2$; also $a_n\sim b_n$ if $a_n/b_n\to1$ as $n\to\infty$.

Lemma 6. *Suppose that X generated by a binary memoryless source with parameter p. Then we have $S_2\sim n\cdot(H(1-d)-(1-d)H(p))$ as $n\to\infty$.*

Proof: By Theorem 1 and Lemma 1, and by the trivial observation $\sum_{|w|=m}P(w)=1$ we have

$$S_2=\sum_w d^{n-|w|}(1-d)^{|w|}\binom{n}{|w|}P(w)\,\log\binom{n}{|w|}$$
$$+\sum_w d^{n-|w|}(1-d)^{|w|}\binom{n}{|w|}P(w)\log P(w)$$
$$=\sum_{m=0}^{n}d^{n-m}(1-d)^m\binom{n}{m}\log\binom{n}{m}$$
$$+\sum_{m=0}^{n}d^{n-m}(1-d)^m\binom{n}{m}\sum_{|w|=m}P(w)\log P(w).$$

The second term above can be computed directly. By the definition of the entropy we have $\sum_{|w|=m}P(w)\log P(w)=-mH(p)$. Consequently,

$$\sum_{m=0}^{n}d^{n-m}(1-d)^m\binom{n}{m}\sum_{|w|=m}P(w)\log P(w)$$
$$=-\sum_{m=0}^{n}d^{n-m}(1-d)^m\binom{n}{m}mH(p)=-n(1-d)H(p).$$

In order to evaluate the first term we apply the results of [5], [7] about the so called *binomial sums*. Notice that

$$\sum_{m=0}^{n}d^{n-m}(1-d)^m\binom{n}{m}\log\binom{n}{m}\sim\log\binom{n}{n(1-d)}\sim nH(1-d).$$

This completes the proof of the lemma. ■

The next step is to show a similar property for S_1, namely that $S_1\sim n\cdot\lambda(d,p)$, where $\lambda(d,p)$ is a non-negative constant. The problem is to obtain some information about $\lambda(d,p)$, but for this we would need precise information about the behavior of $\Omega_X(w)$.

Lemma 7. *Suppose that X generated by a binary memoryless source and set $a_n=S_1(X_1^n,Y(X_1^n))$. Then $a_{n+k}\ge a_n+a_k$..*

Proof: We have

$$\Omega_{X_1^{n+k}}(w)\log\Omega_{X_1^{n+k}}(w)=\left(\sum_{w_1w_2=w}\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2)\right)$$
$$\times\log\left(\sum_{\tilde w_1\tilde w_2=w}\Omega_{X_1^n}(\tilde w_1)\Omega_{X_{n+1}^{n+k}}(\tilde w_2)\right)$$
$$\ge\sum_{w_1w_2=w}\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2)\log\left(\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2)\right)$$
$$=\sum_{w_1w_2=w}\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2)\log\Omega_{X_1^n}(w_1)$$
$$+\sum_{w_1w_2=w}\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2)\log\Omega_{X_{n+1}^{n+k}}(w_2)$$

and consequently

$$a_{n+k} = \sum_w d^{n+k-|w|}(1-d)^{|w|}\, \mathbb{E}\left[\Omega_{X_1^{n+k}}(w) \log \Omega_{X_1^{n+k}}(w)\right]$$
$$\geq \sum_w \sum_{w_1 w_2 = w} d^{n+k-|w_1|-|w_2|}(1-d)^{|w_1|+|w_2|}$$
$$\mathbb{E}\left[\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2) \log \Omega_{X_1^n}(w_1)\right]$$
$$+ \sum_w \sum_{w_1 w_2 = w} d^{n+k-|w_1|-|w_2|}(1-d)^{|w_1|+|w_2|}$$
$$\mathbb{E}\left[\Omega_{X_1^n}(w_1)\Omega_{X_{n+1}^{n+k}}(w_2) \log \Omega_{X_{n+1}^{n+k}}(w_2)\right].$$

Hence, we obtain similarly as in Lemma 5 the relation $a_{n+k} \geq a_k + a_n$. ■

The superadditivity property of Lemma 7 provides the following convergence result.

Lemma 8. *Suppose that X is generated by a binary memoryless source. Then there exists a non-negative constant $\lambda(d,p) \leq H(1-d)$ such that $S_1 \sim n \cdot \lambda(d,p)$ as $n \to \infty$.*

Proof: Since $\Omega_X(w)$ is a non-negative integer we certainly have $S_1 \geq 0$. Furthermore, since $\Omega_X(w) \leq \binom{n}{|w|}$ it follows (as in the proof of Lemma 6) that

$$S_2 \leq \sum_w d^{n-|w|}(1-d)^{|w|}\, \mathbb{E}[\Omega_X(w)] \log \binom{n}{|w|} \sim nH(1-d).$$

Hence (by using the notation of Lemma 7)

$$0 \leq \lambda(d,p) := \sup_{n\geq 1} \frac{a_n}{n} \leq H(1-d).$$

By another application of Fekete's lemma [12] the sequence a_n/n has a limit that equals the supremum $\sup(a_n/n)$. Of course, we have used here the property that $a_{n+k} \geq a_n + a_k$. ■

The proof of Theorem 2 is a combination of Lemma 6 and Lemma 8. The lower bound of $\lambda(d,p)$ follows from the fact that $I(d,p) \geq 0$.

Remark. (Extension to Mixing Sources) We should point out that most results of this section hold for more general sources. For example, from the proof of Lemma 6 we conclude that

$$S_2 \sim n \cdot \left(H(1-d) - (1-d)H(\bar{P})\right)$$

where $\bar{P}$ is defined in Lemma 1, provided $\bar{P}_n$ has a limit $\bar{P}$. Define strongly mixing sources [12] as

$$c_1 P(X_1^m)P(X_{m+1}^n) \leq P(X_1^n) \leq c_2 P(X_1^m)P(X_{m+1}^n).$$

We conclude that Lemma 7 generalizes to $a_{n+k} \geq a_n + a_k + C$ for some constant C, hence Lemma 8 holds.

B. *Proof of Theorem 3: $d \to 1$*

We first note that the empty word does not contribute to the sum (1). Next we consider words of length 1. If $w = 0$ and if $X = X_1^n$ consists of m zeroes and $n-m$ ones then $\Omega_X(w) = m$. The situation is completely symmetric if $w = 1$. Hence the contribution of words of length 1 to $I(X^n; Y(X^n))$ is

$$T_1 := d^{n-1}(1-d)\left(\sum_{m=1}^{n} m \log m \binom{n}{m}\left(p^m q^{n-m} + p^{n-m}q^m\right)\right.$$
$$\left. -np\log(np) - nq\log(nq)\right).$$

By using the inequality

$$\log m = \log(np) + \log\left(1 + \frac{m-np}{np}\right) \leq \log(np) + \frac{m-np}{np}$$

we have the inequality

$$\sum_{m=1}^{n} m \log m \binom{n}{m} p^m q^{n-m} \leq \log(np) \sum_{m=1}^{n} m \binom{n}{m} p^m q^{n-m}$$
$$+ \frac{1}{np}\sum_{m=1}^{n} m(m-np)\binom{n}{m} p^m q^{n-m} = \log(np)np + \frac{npq}{np}$$
$$= np\log(np) + q.$$

Putting all parts together this leads to the following upper bound $T_1 \leq d^{n-1}(1-d) \leq (1-d)$. Let T_2 denote the subsum of (1) corresponding to those terms with $|w| \geq 2$. By using the trivial estimate $\Omega_X(w) \leq \binom{n}{|w|}$ and taking absolute values we obtain the upper bound

$$T_2 \leq 2\sum_{\ell=2}^{n} d^{n-\ell}(1-d)^\ell \binom{n}{\ell} \log \binom{n}{\ell}$$
$$\leq 2\sum_{\ell=2}^{n} d^{n-\ell}(1-d)^\ell \frac{n^\ell}{\ell!} \log n^\ell$$
$$= 2d^n \log n \sum_{\ell\geq 2} \left(\frac{n(1-d)}{d}\right)^\ell \frac{1}{(\ell-1)!}$$
$$\leq 2d^n \log n \frac{n(1-d)}{d}\left(e^{n(1-d)/d} - 1\right).$$

If $n(1-d) = o(1)$ this leads to $T_2 \leq C_1 n^2(1-d)^2 \log n$ with an absolute constant $C_1 > 0$. Summing up and using Corollary 1, we obtain the desired upper bound

$$I(d,p) \leq \frac{1}{n} I(X_1^n; Y(X_1^n)) \leq \frac{1-d}{n} + C_1 n(1-d)^2 \log n.$$

Finally by choosing $n = \lfloor (1-d)^{-1/3} \rfloor$ we derive the upper bound

$$I(d,p) \leq C\,(1-d)^{4/3} \log \frac{1}{1-d}$$

for an absolute constant $C > 0$.

C. *Lower Bound for $d \to 0$*

Finally, we comment on the case $d \to 0$ that has been already solved in [9] and [8] where it is shown that $I(d, 0.5) = 1 + d\log d - Ad + O(d^{2-\varepsilon})$ as $d \to 0$ and $C(d) = I(d, 0.5) + O(d^{3/2-\varepsilon})$. The approach presented in [9] is quite different from ours. However, we can use our methods to obtain corresponding bounds. In particular, we easily get the lower bound (8) below for $I(d,p)$ for general p.

Theorem 4. *Suppose that X is generated by a memoryless source with parameter p. Then, as $d \to 0$,*

$$I(d,p) \geq (1-d)H(p) + d\log d \tag{8}$$

$$-d\log(e) + d(q^2 f(p) + p^2 f(p)) + O\left(d^{2-\varepsilon}\right)$$

for every $\varepsilon > 0$, where $f(x)$ denotes the function $f(x) = \sum_{\ell \geq 2} x^\ell \, \ell \log \ell$. Furthermore, as $d \to 0$,

$$I(d,p) \leq H(p) + d\log d + O(d \log\log(1/d)). \tag{9}$$

Proof: The lower bound for $I(d,p)$ follows the ideas similar to those in the proof of Theorem 2. Instead of taking the limit of a_n/n defined in Lemma 7 we derive lower bounds for a_n/n for certain n. We will only consider words w with $|w| = n-1$. Then

$$a_n \geq d(1-d)^{n-1} \sum_{|w|=n-1} \mathbb{E}[\Omega_X(w) \log \Omega_X(w)].$$

Suppose for the moment that w has the form $w = 0^{i_1}1^{j_1}0^{i_2}1^{j_2}\cdots 0^{i_K}1^{j_K}$, where $i_r, j_r \geq 1$; this means that $w_1 = 0$ and $w_{n-1} = 1$ (the other cases can be handled in completely the same way). If $|w| = n-1$, then we have $\Omega_X(w) = \ell$ (for some $\ell > 2$) if and only if there exists r with

$$i_r = \ell - 1 \quad \text{and} \quad X = 0^{i_1}1^{j_1}\cdots 1^{j_{r-1}}0^{i_r+1}1^{j_r}\cdots 0^{i_K}1^{j_K}$$

or there exists r with

$$j_r = \ell - 1 \quad \text{and} \quad X = 0^{i_1}1^{j_1}\cdots 0^{i_r}1^{j_r+1}0^{i_{r+1}}\cdots 0^{i_K}1^{j_K}.$$

Hence,

$$\sum_{|w|=n-1, w_1=0, w_{n-1}=1} \mathbb{E}[\Omega_X(w)\log\Omega_X(w)] +$$

$$\sum_{\ell\geq 2} \ell\log\ell \sum_{w_1=0, w_{n-1}=1} P(w) \sum_{r\geq 1}\left(p\,\mathbf{I}_{[i_r=\ell-1]} + q\,\mathbf{I}_{[j_r=\ell-1]}\right).$$

Now let Y be a random variable defined on words of length $n-1$ as $Y = \sum_{r\geq 1}\left(p\,\mathbf{I}_{[i_r=\ell-1]} + q\,\mathbf{I}_{[j_r=\ell-1]}\right)$. Then we just have to compute the expected value

$$\mathbb{E}[Y|w_0 = 0, w_{n-1} = 1] = \sum_{r\geq 1}\left(p\,\mathbb{P}[i_r = \ell-1] + q\,\mathbb{P}[j_r = \ell-1]\right).$$

Note that the probability distribution on 0-runs (of length $k \geq 1$) is given by $p^k q/(1-q) = p^{k-1}q$ and that the number of runs in a string of length $n-1$ is approximately pqn. Consequently,

$$\mathbb{E}[Y|w_0 = 0, w_{n-1} = 1] \sim npq\left(p\,p^{\ell-2}q + q\,q^{\ell-2}p\right),$$

or

$$\mathbb{E}[Y, w_0 = 0, w_{n-1} = 1] \sim np^2q^2\left(p\,p^{\ell-2}q + q\,q^{\ell-2}p\right),$$

which leads to

$$\sum_{|w|=n-1, w_1=0, w_{n-1}=1} \mathbb{E}[\Omega_X(w)\log\Omega_X(w)]$$

$$\sim npq\sum_{\ell\geq 2}\ell\log\ell\left(p^\ell q^2 + q^\ell p^2\right).$$

Similarly, we find

$$\sum_{|w|=n-1}\mathbb{E}[\Omega_X(w)\log\Omega_X(w)]$$

$$\sim n\sum_{\ell\geq 2}\ell\log\ell\left(p^\ell q^2 + q^\ell p^2\right) = n(q^2 f(p) + p^2 f(q)).$$

Now we choose $n = \lfloor d^{-\varepsilon}\rfloor$ which ensures that $(1-d)^{n-1} = 1 + O(d^{1-\varepsilon})$. This implies $\lambda(d,p) \geq d(q^2 f(p) + p^2 f(q)) + O(d^{2-\varepsilon})$. Since $H(1-d) = -d\log d - (1-d)\log(1-d) = -d\log d + d\log(e) + O(d^2)$ we obtain the lower bound (8).

For the upper bound we proceed as in the proof of Theorem 3. We start with S_1. Let $S_{1,n-1}$ denote the subsum of S_1 corresponding to words of length $n-1$. Then it follows from the above calculations that $S_{1,n-1} = O(nd)$ (actually we can be much more precise). Furthermore it follows as in the proof of Theorem 3 that $S_1 - S_{1,n-1} = O\left(\log n\, d^2 n^2\right)$ if $dn \to 0$. Finally, for S_2 we have (see Lemma 6)

$$S_2 = -n(1-d)H(p) + d(1-d)^{n-1} n\log n + O\left(\log n\, d^2 n^2\right).$$

Consequently, we obtain for $n = n(d) = \lfloor d^{-1}/\log d^{-1}\rfloor$

$$\begin{aligned} I(d,p) &\leq \frac{S_1 - S_2}{n} \\ &= (1-d)H(p) - (1-d)^{n-1} d\log n + O(d) + O\left(\log n\, d^2 n\right) \\ &= H(p) + d\log d + O(d\log\log(1/d)). \end{aligned}$$

This completes the proof of the theorem. ∎

ACKNOWLEDGMENT

M. Drmota was supported in part by the Austrian Science Foundation FWF Grant No. S9604. W. Szpankowski was supported in part by NSF Science and Technology Center on Science of Information Grant CCF-0939370, NSF Grants DMS-0800568, CCF-0830140, AFOSR Grant FA8655-11-1-3076, NSA Grant H98230-11-1-0141.

REFERENCES

[1] J. Bourdon, and B. Vallée, Generalized Pattern Matching Statistics, *Mathematics and Computer Science II*, Trends. Math., 249–26, 2002.

[2] S. Diggavi and M. Grossglauser, Information Transmission over Finite Buffer Channels, *IEEE Information Theory*, 52, 1226-1237, 2006.

[3] R.L. Dobrushin, Shannon's theorem for channles with synchronization errors.]it Prob. Info. Trans., 18-36, 1967.

[4] Tolga M. Duman, On the Capacity of Memoryless Channels with Synchronization Errors, arXiv:1102.2216v1

[5] P. Flajolet, "Singularity analysis and asymptotics of Bernoulli sums," *Theoretical Computer Science*, 215, 371–381, 1999.

[6] P. Flajolet, W. Szpankowski, and B. Vallée, Hidden Word Statistics, *J. ACM*, **53**(1), 147–183, 2006.

[7] P. Jacquet and W. Szpankowski, "Entropy Computations via Analytic Depoissonization," *IEEE Information Theory*, 45, 1072-1081, 1999.

[8] A. Kalai, M. Mitzenmacher, and M. Sudan, Tight Asymptotic Bounds for the Deletion Channel with Small Deletion Probabilities, *ISIT*, Austin, 2010.

[9] Y. Kanoria and A. Montanari, On the deletion channel with small deletion probability, *ISIT*, Austin, 2010.

[10] M. Mitzenmacher, A survey of results for deletion channels and related synchronization channels, *Probab. Surveys*, 1-33, 2009.

[11] M. Mitzenmacher and E. Drinea, A simple lower bound for the capacity of the deletion channel. *IEEE Information Theory*, 4657-4660, 2006.

[12] W. Szpankowski, Average Case Analysis of Algorithms on Sequences, Wiley, New York, 2001.

[13] R. Venkataramanan, S. Tatikonda, and K. Ramchandran, Achievable Rates for Channels with Deletions and Insertions, *ISIT*, St. Petersburg, Russia, 2011.