How to Use Several Noisy Channels with Unknown Error Probabilities*

Olga Mitina

Dept. of General Psychology, Facultet of Psychology, Moscow State University, Mohovaya 8–5, Moscow, Russia 103009.
E-mail: mitina@psych1.cogsci.msu.su

and

Nikolai Vereshchagin†

Dept. of Mathematical Logic,Facultet of Mechanics and Mathematics, Moscow State University, Vorobjevy Gory, Moscow, Russia 119899.
E-mail: ver@mech.math.msu.su

Consider a sender transmitting a given sequence of bits simultaneously through m binary symmetric channels with error probabilities $\varepsilon_1, \ldots, \varepsilon_m$ such that $0 \leq \varepsilon_i < 1/2$ for each $i = 1, \ldots, m$. The receiver can base its guess for the true transmitted bit on the m channel noisy outputs. Knowing the channel error probabilities, the receiver could apply the maximum likelihood estimator (MLE). What if the receiver does not know the channel error probabilities? In the paper, a strategy is derived whose error probability converges to that of the MLE with rate $O(\sqrt{\ln(n)/n})$, where n is the number of bits transmitted so far, whenever $m \geq 3$.

Key Words: noisy channel, maximum likelihood estimator

1. INTRODUCTION

Consider a sender transmitting a given sequence of bits through a binary symmetric channel with error probability $0 \leq \varepsilon < 1/2$. So, each transmitted bit is received incorrectly with probability ε and the receiver

* The preliminary version of the paper was published under the title *How to Use Expert Advice in the Case when Actual Outcomes Remain Unknown* in the Proceedings of 8th ACM Conference on Computational Learning Theory, 1995, pp. 91–97

† This research was in part supported by the grant MQT000 from the International Science Foundation

has clearly no chance of reducing this error probability unless the sender uses some encoding of the bit sequence. Now suppose that the sender transmits each bit of the sequence simultaneously through m binary symmetric channels with error probabilities $\varepsilon_1, \ldots, \varepsilon_m$ such that $0 \leq \varepsilon_i < 1/2$ for each $i = 1, \ldots, m$. The receiver can now base its guess for the true transmitted bit on the m channel noisy outputs. Knowing the channel error probabilities, the receiver could apply the maximum likelihood estimator (MLE) for the true bit and make a guess which is incorrect with probability at most $\min\{\varepsilon_1, \ldots, \varepsilon_m\}$. By definition MLE, given channel outputs $e_1, e_2, \ldots, e_m$, returns the bit v for which the probability of event "for all $i \leq m$ the ith channel outputs e_i provided the true transmitted bit is v" is greater.

What if the receiver does not know the channel error probabilities? In this case it has to base its guess for the true transmitted bit on channel outputs of all the bits transmitted so far. So, receiver's strategy is a function that maps mn-bit strings, where n is the number of bits transmitted so far, to single bits (guessed values of the true nth transmitted bit). The error probability of a strategy S is defined as the maximum over all $v \in \{0,1\}^n$ of the probability that S errs on the last bit of the input sequence v. It is easy to show that the error probability of any strategy is greater than or equal to that of MLE. In the present paper, we construct a strategy whose error probability converges to that of MLE with rate $O(\sqrt{\ln(n)/n})$, whenever $m \geq 3$ (the constant hidden in "O"-notation depends on m and $\varepsilon_1, \ldots, \varepsilon_m$). The strategy is described in Section 3. So, for example, if there are three channels with error probabilities $\varepsilon_1 \leq \varepsilon_2 \leq \varepsilon_3 < 1/2$ then our strategy will err with probability at most $A(\varepsilon_1, \varepsilon_2, \varepsilon_3) + O(\sqrt{\ln n/n})$, where

$$A(\varepsilon_1, \varepsilon_2, \varepsilon_3) = \varepsilon_1\varepsilon_2\varepsilon_3 + \varepsilon_1\varepsilon_2(1-\varepsilon_3) + \varepsilon_1(1-\varepsilon_2)\varepsilon_3 + \min\{\varepsilon_1(1-\varepsilon_2)(1-\varepsilon_3), (1-\varepsilon_1)\varepsilon_2\varepsilon_3\}.$$

In Section 4, we consider an alternate way to define performance of strategies. There, we consider the uniform probability distribution on the set of input sequences and the uniform probability distribution on the set of possible channel error probabilities (that is, on $[0, 1/2)^m$). The performance of a strategy is then defined as the expected probability of correct guessing if both the input sequence and channel error probabilities are chosen at random. For this way of measuring performance, there exists an optimal strategy which can be implemented in polynomial time if the number of channels or the number of input bits is fixed.

In Section 5 we investigate some properties of the maximum likelihood estimator. In particular, for any $\varepsilon_1, \ldots, \varepsilon_m$ we find the maximum t such that adding any channel with error probability $\varepsilon_{m+1} < t$ to m channels with error probabilities $\varepsilon_1, \ldots, \varepsilon_m$ decreases the error probability of MLE.

2. THE MAXIMUM LIKELIHOOD ESTIMATOR

In this section we assume that that we know error probabilities ε_1, ε_2, ..., $\varepsilon_m < 1/2$ of m channels. Based on channel outputs we have to guess the true input bit. So, our strategy is a function from $\{0,1\}^m$ into $\{0,1\}$. To distinguish between strategies in the sense of the next section, which are functions from $\{0,1\}^{mn}$ into $\{0,1\}$, we will call functions from $\{0,1\}^m$ into $\{0,1\}$ *estimators*.

Let $p_1 = 1-\varepsilon_1$, $p_2 = 1-\varepsilon_2$, ..., $p_m = 1-\varepsilon_m$. In the sequel we denote by $\mathbf{p}$ the sequence $p_1, p_2, \ldots, p_m$. Let θ_1^v, θ_2^v, ..., θ_m^v, $v = 0, 1$, be independent random variables such that $\mathrm{P}[\theta_i^v = v] = p_i$. The random variable θ_i^v is the output of ith channel provided the input bit is v.

DEFINITION 2.1. The performance $Q(\mathbf{p}, E)$ of an estimator E with respect to channels with error probabilities $\mathbf{p}$ is the the minimal of two numbers:

$$\mathrm{P}[E(\theta_1^0, \theta_2^0, \ldots, \theta_m^0) = 0], \qquad \mathrm{P}[E(\theta_1^1, \theta_2^1, \ldots, \theta_m^1) = 1].$$

Let us define now the maximum likelihood estimator *MLE*:

$$MLE(e_1, e_2, \ldots, e_m) = \begin{cases} 0 & \text{if } \mathrm{P}[\theta_1^0 = e_1, \ldots, \theta_m^0 = e_m] \\ & \quad > \mathrm{P}[\theta_1^1 = e_1, \ldots, \theta_m^1 = e_m], \\ 1 & \text{if } \mathrm{P}[\theta_1^0 = e_1, \ldots, \theta_m^0 = e_m] \\ & \quad > \mathrm{P}[\theta_1^1 = e_1, \ldots, \theta_m^1 = e_m], \\ e_1 & \text{otherwise} \end{cases}$$

Of course the maximum likelihood estimator depends on the sequence $\mathbf{p} = p_1, \ldots, p_m$. When this sequence is not clear from the context we will write $MLE_{\mathbf{p}}$ instead of *MLE*.

We denote $1 - x$ by $\bar{x}$.

The next lemma, which is proven in Appendix, is well known; it states the optimality of the estimator *MLE*.

LEMMA 2.1. *$Q(\mathbf{p}, MLE_{\mathbf{p}}) \geq Q(\mathbf{p}, E)$ for all E and all $\mathbf{p}$.*

EXAMPLE 2.1. Let $p_1 \geq p_2 \geq p_3 > 1/2$. Then $MLE(000) = MLE(001) = MLE(010) = 0$, $MLE(111) = MLE(110) = MLE(101) = 1$,

$$MLE(011) = \begin{cases} 0, & \text{if } p_1\bar{p}_2\bar{p}_3 \geq \bar{p}_1 p_2 p_3, \\ 1, & \text{else,} \end{cases}$$

and $MLE(100) = \overline{MLE(011)}$. So, MLE returns the output of the first channel if $p_1\bar{p}_2\bar{p}_3 \geq \bar{p}_1 p_2 p_3$ and returns the majority of outputs otherwise.

Its performance is equal to p_1 in the first case and to $p_1p_2p_3 + \bar{p}_1p_2p_3 + p_1\bar{p}_2p_3 + p_1p_2\bar{p}_3$ in the second case.

3. ASYMPTOTICALLY OPTIMAL STRATEGY

In this section we assume that channel error probabilities $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m < 1/2$ *are unknown*. The channels transmit n bits $v_1, v_2, \ldots, v_n$, which are unknown, too. Given channel outputs we have to guess v_n. Possible ways to do this will be called m,n-strategies. Thus a *m,n-strategy* is a function from the set $(\{0,1\}^m)^n$ into $\{0,1\}$. An example of m,n-strategy is the function

$$S_1(e_{11}, e_{21}, \ldots, e_{m1}, e_{12}, e_{22}, \ldots, e_{m2}, \ldots, e_{1n}, e_{2n}, \ldots, e_{mn}) = e_{1n}$$

(trust the first channel). Another example is the function

$$\begin{aligned} &S_2(e_{11}, e_{21}, \ldots, e_{m1}, \ldots, e_{1n}, e_{2n}, \ldots, e_{mn}) \\ &\quad = \mathrm{MAJORITY}(e_{1n}, e_{2n}, \ldots, e_{mn}) \end{aligned}$$

(output the majority; here, MAJORITY denotes the function taking the value 1 iff at least half arguments take the value 1). Let $p_1 = 1 - \varepsilon_1$, $p_2 = 1 - \varepsilon_2$, ..., $p_m = 1 - \varepsilon_m$. We will denote by $MLE_{\mathbf{p},m,n}$ the m,n-strategy

$$S(e_{11}, e_{21}, \ldots, e_{m1}, \ldots, e_{1n}, e_{2n}, \ldots, e_{mn}) = MLE_{\mathbf{p}}(e_{1n}, e_{2n}, \ldots, e_{mn}).$$

A *strategy* is a family $S = \{S_{mn} \mid m,n = 1,2,\ldots\}$ of functions such that S_{mn} is a m,n-strategy for any m,n. The strategy $\{MLE_{\mathbf{p},m,n} \mid m,n = 1,2,\ldots\}$ will be denoted in the same way as the maximum likelihood estimator: $MLE_{\mathbf{p}}$.

Let ξ_{ij}^v, $i \le m$, $j \le n$, $v = 0,1$ denote independent random variables such that $\mathrm{P}[\xi_{ij}^v = v] = p_i$ for all i,j,v. The variable ξ_{ij}^v is the output of ith channel transmitting jth input bit v. Define the *performance $Q(\mathbf{p},S)$ of a m,n-strategy S with respect to* $\mathbf{p} = p_1, p_2, \ldots, p_m$ as follows:

$$Q(\mathbf{p},S) = \min_{v_1,\ldots,v_n} \mathrm{P}[S(\xi_{11}^{v_1}\xi_{21}^{v_1}\cdots\xi_{m1}^{v_1}\cdots\xi_{1n}^{v_n}\xi_{2n}^{v_n}\cdots\xi_{mn}^{v_n}) = v_n].$$

It follows from Lemma 2.1 that $Q(\mathbf{p},S) \le Q(\mathbf{p},MLE_{\mathbf{p}})$ for any m,n-strategy S and for any $\mathbf{p}$.

THEOREM 3.1. *There exists a strategy S computable in polynomial time (in m,n) such that*

$$Q(\mathbf{p},S_{mn}) \ge Q(\mathbf{p},MLE_{\mathbf{p}}) - O(\sqrt{\ln n/n})$$

for any $m \geq 3$ and any $\mathbf{p}$ (the constant hidden in "O"-notation depends on m and on $\mathbf{p}$).

Before to prove the theorem let us remind the Chernoff bound (see [Shiryaev 1995, ch. I, §6, (42)], which will be used in the proof.

THEOREM 3.2. *(Chernoff bound [Chernoff 1952]) Let $\xi_1, \ldots, \xi_n$ be independent random variables in the set $\{0,1\}$ such that $\mathrm{P}[\xi_i = 1] = p$ for all i. Then for any $\gamma \in [0;1]$,*

$$\mathrm{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - p\right| \geq \gamma\right] \leq 2e^{-2\gamma^2 n}.$$

Proof of Theorem 3.1. The m,n-strategy S_{mn} is computed in two stages; on the first stage we use $n-1$ first outputs of channels to compute an approximate $\hat{\mathbf{p}}$ to $\mathbf{p}$, on the second stage we apply the maximum likelihood estimator $MLE_{\hat{\mathbf{p}}}$ to nth output of channels. Let us describe the stages in detail.

First stage. Consider first the case $m = 3$. Let r_1, r_2, r_3 be defined by equalities $p_1 = 1/2 + r_1$, $p_2 = 1/2 + r_2$, $p_3 = 1/2 + r_3$. Then r_1, r_2, r_3 are greater than 0. Let s_{12} stand for the probability that the outputs of the first and second channels (on the same input) coincide. It is easy to see that $s_{12} = 1/2 + 2r_1r_2$. The crucial point is that this probability does not depend on the input bit. In the same way define s_{13} and s_{23}. Note that given s_{12}, s_{13} and s_{23} we can find r_1, r_2 and r_3 by using the simple formulas:

$$r_1 = \sqrt{\tfrac{(2s_{12}-1)(2s_{13}-1)}{4(2s_{23}-1)}}, \quad r_2 = \sqrt{\tfrac{(2s_{12}-1)(2s_{23}-1)}{4(2s_{13}-1)}},$$
$$r_3 = \sqrt{\tfrac{(2s_{13}-1)(2s_{23}-1)}{4(2s_{12}-1)}}.$$

This observation leads to the following algorithm to compute approximations $\hat{p}_1$, $\hat{p}_2$, $\hat{p}_3$ to p_1, p_2, p_3. Let e_{11}, e_{21}, e_{31}, e_{12}, e_{22}, e_{32}, …, e_{1n}, e_{2n}, e_{3n} be outputs of channels. Find first the values $\hat{s}_{12} \rightleftharpoons |\{j < n | e_{1j} = e_{2j}\}|/(n-1)$ and $\hat{s}_{13}$, $\hat{s}_{23}$ defined in the same way (we assume that $n > 1$). If at least one of the numbers $2\hat{s}_{12} - 1, 2\hat{s}_{13} - 1, 2\hat{s}_{23} - 1$ is not positive let $\hat{p}_1 = \hat{p}_2 = \hat{p}_3 = 1$. Otherwise substitute in the three above formulas $\hat{s}_{12}$, $\hat{s}_{13}$, $\hat{s}_{23}$ respectively for s_{12}, s_{13}, s_{23} and denote the resulting values by $\check{r}_1$, $\check{r}_2$, $\check{r}_3$. Find rational numbers $\hat{r}_1$, $\hat{r}_2$, $\hat{r}_3$ of the form k/n being the closest to $\check{r}_1$, $\check{r}_2$, $\check{r}_3$, respectively. Let $\hat{p}_1 = 1/2 + \hat{r}_1$, $\hat{p}_2 = 1/2 + \hat{r}_2$ and $\hat{p}_3 = 1/2 + \hat{r}_3$.

The case $m = 3$ is done. In the case $m > 3$ we find $\hat{p}_1$, $\hat{p}_2$, …, $\hat{p}_m$ as follows: $\hat{p}_1$, $\hat{p}_2$, $\hat{p}_3$ are computed just as earlier and to find $\hat{p}_i$ for $i > 3$ we

group the first, the second and the ith channels and do the same thing. The first stage is completed.

Second stage: Output the value $MLE_{\hat{\mathbf{p}}}(e_{1n}, e_{2n}, \dots, e_{mn})$, where $\hat{\mathbf{p}} = \hat{p}_1, \hat{p}_2, \dots, \hat{p}_m$.

The strategy S_{mn} is defined. Let us prove that for any $\mathbf{p}$,

$$Q(\mathbf{p}, S_{mn}) \geq Q(\mathbf{p}, MLE_{\mathbf{p}}) - O(\sqrt{\ln n/n}).$$

It is easy to see that, for the constructed strategy, the probability of error does not depend on the input sequence. So, we will estimate this probability assuming that the input sequence consists of 1s.

Let us fix arbitrary $p_1, p_2, \dots, p_m > 1/2$. Denote $\min\{p_1 - 1/2, p_2 - 1/2, \dots, p_m - 1/2\}$ by α. Then we have $s_{jk} \geq 1/2 + 2\alpha^2$ for all $j < k \leq m$.

Let $\gamma = \sqrt{\ln n/(n-1)}$. Let us estimate the probability of the event $|\hat{s}_{jk} - s_{jk}| \leq \gamma$ using Chernoff bound. The conditions of Theorem 3.2 are fulfilled. By Chernoff inequality any of the events

$$|\hat{s}_{jk} - s_{jk}| \leq \gamma \tag{1}$$

for $j < k \leq m$ holds with probability at least $1 - 2e^{-2\ln n} = 1 - 2n^{-2}$.

The inequalities (1) imply that $\hat{p}_j$ is close to p_j for all $j \leq m$. More specifically, the following bound holds.

LEMMA 3.1. *If n is large enough and the inequality (1) is true for all $j < k \leq m$ then $|\hat{p}_j - p_j| \leq \alpha^{-5}\gamma + 1/n$ for all $j \leq m$.*

Proof. Let n be so large that $\gamma < 1.5\alpha^2$ and let (1) be true for all $j < k \leq m$. Let us prove that $|\hat{p}_1 - p_1| < \alpha^{-5}\gamma + 1/n$. We have $\hat{s}_{jk} \geq s_{jk} - 1.5\alpha^2 \geq 0.5 + 2\alpha^2 - 1.5\alpha^2 = 0.5 + 0.5\alpha^2$. Denote

$$\begin{aligned} a &= 2s_{12} - 1, \quad \hat{a} = 2\hat{s}_{12} - 1, \\ b &= 2s_{13} - 1, \quad \hat{b} = 2\hat{s}_{13} - 1, \\ c &= 2s_{23} - 1, \quad \hat{c} = 2\hat{s}_{23} - 1. \end{aligned}$$

We have

$$\begin{aligned} &4\alpha^2 \leq a \leq 1, \ \ \alpha^2 \leq \hat{a} \leq 1, \ \ |\hat{a} - a| \leq 2\gamma, \\ &4\alpha^2 \leq b \leq 1, \ \ \alpha^2 \leq \hat{b} \leq 1, \ \ |\hat{b} - b| \leq 2\gamma, \\ &4\alpha^2 \leq c \leq 1, \ \ \alpha^2 \leq \hat{c} \leq 1, \ \ |\hat{c} - c| \leq 2\gamma. \end{aligned}$$

Therefore

$$\begin{aligned}|p_1 - \hat{p}_1| &= |r_1 - \hat{r}_1| \le |r_1 - \check{r}_1| + 1/n \\ &= \left| \sqrt{\frac{ab}{4c}} - \sqrt{\frac{\hat{a}\hat{b}}{4\hat{c}}} \right| + \frac{1}{n} = \frac{|ab\hat{c} - \hat{a}\hat{b}c|}{2\left(\sqrt{abc\hat{c}^2} + \sqrt{\hat{a}\hat{b}\hat{c}c^2}\right)} + \frac{1}{n} \\ &\le \frac{3 \cdot 2\gamma}{2(\sqrt{64\alpha^{10}} + \sqrt{16\alpha^{10}})} + \frac{1}{n} \le \frac{\gamma}{\alpha^5} + \frac{1}{n}.\end{aligned}$$

∎

By Lemma 3.1,

$$\begin{aligned}&\mathrm{P}\left[\exists j \ |\hat{p}_j - p_j| > \gamma/\alpha^5 + 1/n\right] \\ &\le \mathrm{P}\left[\exists \ j < k \le m \ |\hat{s}_{jk} - s_{jk}| > \gamma\right] \le \binom{m}{2} 2n^{-2} \le m^2 n^{-2}.\end{aligned}$$

It is easy to verify that for any two events A, B it holds

$$\mathrm{P}[A] \ge \mathrm{P}[A|B] - \mathrm{P}[\bar{B}].$$

Using this inequality we get

$$\begin{aligned}&Q(\mathbf{p}, S_{mn}) \\ &\ge \mathrm{P}\left[MLE_{\hat{\mathbf{p}}}(\xi^1_{1n}, \ldots, \xi^1_{mn}) = 1 \mid \forall j \ |\hat{p}_j - p_j| \le \gamma/\alpha^5 + 1/n\right] \\ &\quad - m^2 n^{-2}. \qquad (2)\end{aligned}$$

For any fixed estimator E the events $E(\xi^1_{1n}, \ldots, \xi^1_{mn}) = 1$ and $\forall j \ |\hat{p}_j - p_j| \le \gamma/\alpha^5 + 1/n$ are independent, as the former event depends only on $\xi^1_{1n}, \ldots, \xi^1_{mn}$ and the latter one only on $\xi^1_{11}, \ldots, \xi^1_{m1}, \ldots, \xi^1_{1(n-1)}, \ldots, \xi^1_{m(n-1)}$. If $\hat{p}_j$ is close to p_j for all $j \le m$, then the estimator $MLE_{\hat{\mathbf{p}}}$ is close to the estimator $MLE_{\mathbf{p}}$. The value $Q(\mathbf{p}, MLE_{\hat{\mathbf{p}}})$ is continuous when $\hat{\mathbf{p}}$ varies. Therefore, the value

$$\mathrm{P}\left[MLE_{\hat{\mathbf{p}}}(\xi^1_{1n}, \ldots, \xi^1_{mn}) = 1 \mid \forall j \ |\hat{p}_j - p_j| \le \gamma/\alpha^5 + 1/n\right]$$

is close to $Q(\mathbf{p}, E_{\mathbf{p}})$. An appropriate upper bound follows from the next lemma proven in the Appendix.

LEMMA 3.2. *For any $\hat{\mathbf{p}}$ such that $|\hat{p}_i - p_i| \le \delta$ for all $i \le m$ we have*

$$|\mathrm{P}[MLE_{\hat{\mathbf{p}}}(\xi^1_{1n}, \ldots, \xi^1_{mn}) = 1] - Q(\mathbf{p}, MLE_{\mathbf{p}})| \le m 2^m \delta.$$

Let $\delta = \gamma/\alpha^5 + 1/n$. For any $\xi^1_{11}, \ldots, \xi^1_{m1}, \ldots, \xi^1_{1(n-1)}, \ldots, \xi^1_{m(n-1)}$ such that $\forall j \ |\hat{p}_j - p_j|$ we have by Lemma 3.2 $\mathrm{P}[MLE_{\hat{\mathbf{p}}}(\xi^1_{1n}, \ldots, \xi^1_{mn}) = 1] \geq Q(\mathbf{p}, MLE_{\mathbf{p}}) - m2^m\delta$. Therefore we can continue inequality (2)

$$\begin{aligned} Q(\mathbf{p}, S_{mn}) &\geq Q(\mathbf{p}, MLE_{\mathbf{p}}) - m2^m(\gamma/\alpha^5 + 1/n) - m^2n^{-2} \\ &= Q(\mathbf{p}, MLE_{\mathbf{p}}) - O(\sqrt{\ln n/n}). \end{aligned}$$

Remark 3. 1. The upper bound for the gap between the performance of the strategy and that of the ML estimator is exponential in the number of channels. So it is not clear whether the fact that it is polynomial-time in the number of channels is useful, since number of outcomes needs to be exponentially large in the number of channels for a given desired relative performance.

4. THE OPTIMAL STRATEGY FOR AN ALTERNATE WAY TO DEFINE PERFORMANCE OF STRATEGIES

In this section we will measure the performance of strategies as follows: for a m, n-strategy S let

$$\check{Q}(S) = \int \Big\{ 1/2^n \sum_{v_1, \ldots, v_n} \mathrm{P}[S(\xi^{v_1}_{11} \cdots \xi^{v_n}_{mn}) = v_n] \Big\} dp_1 \cdots dp_m,$$

where the integral is taken over $(1/2, 1]^m$.

An m, n-strategy S is called *optimal* if $\check{Q}(S) \geq \check{Q}(S')$ for all m, n-strategies S'. A strategy $S = \{S_{mn} \mid m, n = 0, 1, 2, \ldots\}$ is called *optimal* if S_{mn} is optimal for all m, n.

THEOREM 4.1. *There is an optimal strategy computable in polynomial time if n is fixed or m is fixed.*

Proof. Let $\check{Q}_{mn} = \max_S \check{Q}(S)$, where the maximum is taken over all m, n-strategies S. By definition we have

$$\check{Q}_{mn} = \max_S \int 2^{-n} \sum_{\mathbf{v} = v_1, \ldots, v_n} \mathrm{P}\left[S(\xi^{v_1}_{11}, \ldots, \xi^{v_n}_{mn}) = v_n\right] dp_1 \cdots dp_m.$$

Obviously,

$$\begin{aligned} &\mathrm{P}[S(\xi^{v_1}_{11}, \ldots, \xi^{v_n}_{mn}) = v_n] \\ &\quad = \sum_{e_{11}, \ldots, e_{mn}} \mathrm{P}[S(e_{11}, \ldots, e_{mn}) = v_n \wedge \forall i \leq m \forall j \leq n \ \xi^{v_j}_{ij} = e_{ij}]. \end{aligned}$$

Therefore,

$$\begin{aligned}
&\int 2^{-n} \sum_{\mathbf{v}} \mathrm{P}[S(\xi_{11}^{v_1}, \ldots, \xi_{mn}^{v_n}) = v_n] dp_1 \ldots dp_m \\
&= \int 2^{-n} \sum_{\mathbf{v}} \sum_{e_{11},\ldots,e_{mn}} \mathrm{P}[S(e_{11}, \ldots, e_{mn}) = v_n \wedge \forall ij\ \xi_{ij}^{v_j} = e_{ij}] dp_1 \ldots dp_m \\
&= \sum_{e_{11},\ldots,e_{mn}} \int 2^{-n} \sum_{\mathbf{v}} \mathrm{P}[S(e_{11}, \ldots, e_{mn}) = v_n \wedge \forall ij\ \xi_{ij}^{v_j} = e_{ij}] dp_1 \cdots dp_m.
\end{aligned}$$

Given $e \in \{0,1\}$ denote by $F(e, e_{11}, \ldots, e_{mn})$ the value

$$\int 2^{-n} \sum_{\mathbf{v}} \mathrm{P}[e = v_n \wedge \forall ij\ \xi_{ij}^{v_j} = e_{ij}] dp_1 \cdots dp_m. \tag{3}$$

Then

$$\begin{aligned}
\check{Q}_{mn} &= \max_S \sum_{e_{11},\ldots,e_{mn}} F(S(e_{11}, \ldots, e_{mn}), e_{11}, \ldots, e_{mn}) \\
&= \sum_{e_{11},\ldots,e_{mn}} \max_{e \in \{0,1\}} F(e, e_{11}, \ldots, e_{mn}).
\end{aligned}$$

Therefore the strategy that given a tuple $\langle e_{11}, \ldots, e_{mn} \rangle$ returns 0 if

$$F(0, e_{11}, \ldots, e_{mn}) > F(1, e_{11}, \ldots, e_{mn})$$

and 1 otherwise is optimal. Note that this is again the maximum likelihood estimator. Indeed, $F(e, e_{11}, \ldots, e_{mn})$ is equal to the probability that the last bit in a random input sequence $\mathbf{v}$ is equal to e and channels chosen at random output $e_{11}, \ldots, e_{mn}$ on $\mathbf{v}$.

Let us prove that this strategy is polynomial-time if either m or n is fixed.

Assume first that n is fixed. It suffices to prove that given e and $\langle e_{11}, \ldots, e_{mn} \rangle$ we can compute $F(e, e_{11}, \ldots, e_{mn})$ in polynomial time. We

have

$$\begin{aligned}
&F(e, e_{11}, \ldots, e_{mn}) \\
&= \int \Big(2^{-n} \sum_{\mathbf{v}} \mathrm{P}[e = v_n \wedge \forall ij \xi_{ij}^{v_j} = e_{ij}]\Big) dp_1 \cdots dp_m \\
&= \int \Big(2^{-n} \sum_{\mathbf{v}: v_n = e} \mathrm{P}[\forall ij\ \xi_{ij}^{v_j} = e_{ij}]\Big) dp_1 \cdots dp_m \\
&= \int \Big(2^{-n} \sum_{\mathbf{v}: v_n = e} \prod_{ij} \mathrm{P}[\xi_{ij}^{v_j} = e_{ij}]\Big) dp_1 \cdots dp_m \\
&= \sum_{\mathbf{v}: v_n = e} 2^{-n} \int \Big(\prod_{ij} \mathrm{P}[\xi_{ij}^{v_j} = e_{ij}]\Big) dp_1 \cdots dp_m.
\end{aligned}$$

As the number of different $\mathbf{v}$'s is 2^{n-1} (hence does not depend on m), it suffices to show that given $\mathbf{v}$ we can compute in polynomial time the value

$$\int (\prod_{ij} \mathrm{P}[\xi_{ij}^{v_j} = e_{ij}]) dp_1 \ldots dp_m.$$

This expression can be rewritten as

$$\prod_{i=1}^{m} (\int \prod_{j=1}^{n} \mathrm{P}[\xi_{ij}^{v_j} = e_{ij}] dp_i).$$

The value

$$\mathrm{P}[\xi_{ij}^{v_j} = e_{ij}] = \begin{cases} p_i & \text{if } v_j = e_{ij}, \\ 1 - p_i & \text{otherwise} \end{cases}$$

is a linear polynomial in p_i with coefficients not exceeding 1 in absolute value. By multiplying $\mathrm{P}[\xi_{ij}^{v_j} = e_{ij}]$ for $j = 1, 2, \ldots, n$ we find in time not depending on m the coefficients of the polynomial $g_i(p_i) = \prod_{j=1}^{n} \mathrm{P}[\xi_{ij}^{v_j} = e_{ij}]$. Those coefficients do not exceed 2^n in absolute value. Then we find $\int g_i(p_i) dp_i$ and multiply the resulting values for $i = 1, 2, \ldots, m$.

Assume now that m is fixed. Again it suffices to prove that given e and $\langle e_{11}, \ldots, e_{mn} \rangle$ we can compute $F(e, e_{11}, \ldots, e_{mn})$ in polynomial time. In this case we write $F(e, e_{11}, \ldots, e_{mn})$ as

$$\begin{aligned}
F(e, e_{11}, \ldots, e_{mn}) &= \int \Big(2^{-n} \sum_{\mathbf{v}: v_n = e} \mathrm{P}[\forall_{ij} \xi_{ij}^{v_j} = e_{ij}]\Big) dp_1 \cdots dp_m \\
&= \int \Big(2^{-n} \sum_{\mathbf{v}: v_n = e} \prod_{j} \prod_{i} \mathrm{P}[\xi_{ij}^{v_j} = e_{ij}]\Big) dp_1 \cdots dp_m.
\end{aligned}$$

Let us note that

$$\sum_{\mathbf{v}:v_n=e} \prod_j \prod_i \mathrm{P}[\xi_{ij}^{v_j} = e_{ij}]$$
$$= \prod_{j=1}^{n-1} \Big\{ \Big(\prod_{i=1}^{m} \mathrm{P}[\xi_{ij}^0 = e_{ij}] + \prod_{i=1}^{m} \mathrm{P}[\xi_{ij}^1 = e_{ij}] \Big) \prod_{i=1}^{m} \mathrm{P}[\xi_{in}^e = e_{in}] \Big\}.$$

Therefore,

$$F(e, e_{11}, \ldots, e_{mn}) = \int 2^{-n} \prod_{j=1}^{n} f_j(p_1, p_2, \ldots, p_m) dp_1 \cdots dp_m,$$

where

$$f_j(p_1, p_2, \ldots, p_m) = \prod_{i=1}^{m} \mathrm{P}[\xi_{ij}^0 = e_{ij}] + \prod_{i=1}^{m} \mathrm{P}[\xi_{ij}^1 = e_{ij}]$$

for $j = 1, \ldots, n-1$ and $f_n(p_1, p_2, \ldots, p_m) = \prod_{i=1}^{m} \mathrm{P}[\xi_{in}^e = e_{in}]$. All the $f_1, \ldots, f_n$ are multilinear polynomials (i.e. of degree 1 in every variable) in $p_1, \ldots, p_m$ and have at most 2^m integer coefficients not exceeding 2^{m+1} in absolute value. In time depending only on m we can find the coefficients of any of $f_1, \ldots, f_n$. Therefore, in time linear in n we can find the coefficients of all $f_1, \ldots, f_n$.

Now we have to multiply n multilinear polynomials. Note that we cannot do that directly because direct multiplying of n polynomials having 2^m terms yields $(2^m)^n$ terms. To avoid this difficulty let us note that the resulting polynomial $f = f_1 f_2 \cdots f_n$ has degree n in every variable and therefore has $(n+1)^m$ coefficients. Therefore we can compute the coefficients of polynomials $f_1 f_2$, $f_1 f_2 f_3$, ..., $f_1 f_2 \cdots f_n$ in succession. On every step we multiply two polynomials of degree at most n. This multiplying requires only $(n+1)^{2m}$ arithmetical operations.

Then we can find the integral of that polynomial by integrating separately all its terms. ∎

5. COMPARING DIFFERENT CHANNEL SETS

The value $A(p_1, p_2, \ldots, p_m) =_{\mathrm{def}} Q(p_1, p_2, \ldots, p_m, MLE)$ will be called *the performance of the channel set* $p_1, p_2, \ldots, p_m$.

THEOREM 5.1. *The function* $A(p_1, p_2, \ldots, p_m)$ *is monotone, i.e.* $p_1 \geq p_1', \ldots, p_m \geq p_m' \Rightarrow A(p_1, \ldots, p_m) \geq A(p_1', \ldots, p_m')$.

We will prove this theorem later.

The next theorem answers the following question: suppose we want to increase the performance of a channel set by adding a new channel. How low should be its error probability?

Let $x^{(1)}$ denote $1-x$ and let $x^{(0)}$ denote x. Let $\psi(y,z) = \frac{\max\{y,z\}}{y+z}$ and let

$$t = \min_{e_1,\dots,e_{m-1}} \psi(p_1^{(e_1)} p_2^{(e_2)} \cdots p_{m-1}^{(e_{m-1})}, p_1^{(\bar{e}_1)} p_2^{(\bar{e}_2)} \cdots p_{m-1}^{(\bar{e}_{m-1})}).$$

THEOREM 5.2. *$A(p_1,\dots,p_{m-1},p_m) > A(p_1,\dots,p_{m-1})$ for any $p_m > t$ and $A(p_1,\dots,p_{m-1},p_m) = A(p_1,\dots,p_{m-1})$ for any $1/2 \le p_m \le t$.*

For example, assume that $p_i = p$ for $i = 1,2,\dots,m-1$. Then it is easy to see that $t = 1/2$ if $m-1$ is even and $t = p$ otherwise. Thus for even $m-1$ adding any new channel will increase the performance of the set, whereas for odd $m-1$ only adding a channel with error probability lower than $1-p$ can do so.

Proof of Theorem 5.1 and Theorem 5.2. To prove the former theorem it suffices to prove that the function $A(p_1,p_2,\dots,p_m)$ is monotone in p_m.

The equality (A.2) from the Appendix implies that

$$\begin{aligned} A(p_1,\dots,p_m) &= 0.5 \sum_{e_1,\dots,e_m} \max\{p_1^{(e_1)} \cdots p_m^{(e_m)}, p_1^{(\bar{e}_1)} \cdots p_m^{(\bar{e}_m)}\} \\ &= 0.5 \sum_{e_1,\dots,e_{m-1}} \Big(\max\{p_1^{(e_1)} \cdots p_{m-1}^{(e_{m-1})} p_m, p_1^{(\bar{e}_1)} \cdots p_{m-1}^{(\bar{e}_{m-1})}(1-p_m)\} \\ &\quad + \max\{p_1^{(e_1)} \cdots p_{m-1}^{(e_{m-1})}(1-p_m), p_1^{(\bar{e}_1)} \cdots p_{m-1}^{(\bar{e}_{m-1})} p_m\}\Big) \end{aligned}$$

For any positive α, β the function

$$f_{\alpha\beta}(x) = \max\{\alpha x, \beta(1-x)\} + \max\{\alpha(1-x), \beta x\}$$

takes the value $\max\{\alpha,\beta\}$ if $1/2 \le x \le \frac{\max\{\alpha,\beta\}}{\alpha+\beta}$ and $(\alpha+\beta)x$ if $x \ge \frac{\max\{\alpha,\beta\}}{\alpha+\beta}$. Thus, this function is monotone on the segment $[1/2;1]$. Therefore, the function $A(p_1,p_2,\dots,p_m)$ is monotone in the variable p_m as a sum of functions of the type $f_{\alpha\beta}(x)$.

Moreover, the above equations show that

$$\begin{aligned} &A(p_1,p_2,\dots,p_m) \\ &= 0.5 \sum_{e_1,\dots,e_{m-1}} \max\{p_1^{(e_1)} \cdots p_{m-1}^{(e_{m-1})}, p_1^{(\bar{e}_1)} \cdots p_{m-1}^{(\bar{e}_{m-1})}\} = A(p_1,\dots,p_{m-1}) \end{aligned}$$

for any $p_m \le t$ and $A(p_1, p_2, \dots, p_m) > A(p_1, p_2, \dots, p_{m-1})$ for all $p_m > t$. This proves Theorem 5.2.

6. CONCLUSION

We left open the following question: how tight is the bound $O(\sqrt{\ln n/n})$ in Theorem 3.1? That is, how fast can tend to zero the value $Q(\mathbf{p}, S_{mn}) - Q(\mathbf{p}, MLE_{\mathbf{p}})$ when n goes to infinity, where S be an arbitrary strategy (not necessary polynomial-time).

Another question is: what is the performance of the following strategy $\tilde{S}$: given $e_{11}, \dots, e_{mn}$ find those real number $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m \in [0; 1/2)$ and bits $v_1, v_2, \dots, v_n$ for which the probability of the event "the channels with error probabilities $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ output $e_{11}, \dots, e_{mn}$ on input sequence is v_1, v_2, ..., v_n" is maximal, then output v_n. It is easy to see that this is again the maximal likelihood estimator. We do not know if the strategy $\tilde{S}$ is asymptotically optimal neither we know whether it is polynomial-time.

APPENDIX

Proof of Lemma 2.1. It is easy to see that

$$MLE_{\mathbf{p}}(\bar{e}_1, \dots, \bar{e}_m) = \overline{MLE_{\mathbf{p}}(e_1, \dots, e_m)}. \tag{A.1}$$

Let us prove that

$$\begin{aligned} Q(\mathbf{p}, MLE_{\mathbf{p}}) &= \mathrm{P}\big[MLE_{\mathbf{p}}(\theta_1^0, \dots, \theta_m^0) = 0\big] = \mathrm{P}\big[MLE_{\mathbf{p}}(\theta_1^1, \dots, \theta_m^1) = 1\big] \\ &= (1/2) \sum_{\mathbf{e}=e_1,\dots,e_m} \mathrm{P}[\theta_1^{MLE_{\mathbf{p}}(\mathbf{e})} = e_1, \dots, \theta_m^{MLE_{\mathbf{p}}(\mathbf{e})} = e_m] \end{aligned} \tag{A.2}$$

Indeed, let $\oplus$ denote addition modulo 2, let $x^{(1)}$ denote $1 - x$ and let $x^{(0)}$ denote x. Then

$$\begin{aligned} &\mathrm{P}\big[MLE_{\mathbf{p}}(\theta_1^0, \dots, \theta_m^0) = 0\big] \\ &= \sum_{e_1,\dots,e_m\,:\,MLE(e_1,\dots,e_m)=0} p_1^{(e_1 \oplus 0)} \cdots p_m^{(e_m \oplus 0)} \\ &= \sum_{e_1,\dots,e_m\,:\,MLE(e_1,\dots,e_m)=1} p_1^{(\bar{e}_1 \oplus 0)} \cdots p_m^{(\bar{e}_m \oplus 0)} \\ &= \sum_{e_1,\dots,e_m\,:\,MLE(e_1,\dots,e_m)=1} p_1^{(e_1 \oplus 1)} \cdots p_m^{(e_m \oplus 1)} \\ &= \mathrm{P}\big[MLE_{\mathbf{p}}(\theta_1^1, \dots, \theta_m^1) = 1\big], \end{aligned} \tag{A.3}$$

where the second equality holds because of Equation (A.1). The equalities (A.3) imply that

$$Q(\mathbf{p}, MLE_{\mathbf{p}}) = (1/2) \sum_{e_1,\ldots,e_m} p_1^{(e_1 \oplus MLE_{\mathbf{p}}(e_1,\ldots,e_m))} \cdots p_m^{(e_m \oplus MLE_{\mathbf{p}}(e_1,\ldots,e_m))}.$$

Let E be an estimator. Then

$$\begin{aligned}
Q(\mathbf{p}, E) &\leq 1/2\big(\mathrm{P}\big[E(\theta_1^0,\ldots,\theta_m^0)=0\big] + \mathrm{P}\big[E(\theta_1^1,\ldots,\theta_m^1)=1\big]\big) \\
&= 1/2 \sum_{e_1,\ldots,e_m,v\,:\,E(e_1,\ldots,e_m)=v} p_1^{(e_1\oplus v)} \cdots p_m^{(e_m \oplus v)} \\
&= (1/2) \sum_{e_1,\ldots,e_m} p_1^{(e_1 \oplus E(e_1,\ldots,e_m))} \cdots p_m^{(e_m \oplus E(e_1,\ldots,e_m))} \\
&\leq (1/2) \sum_{e_1,\ldots,e_m} p_1^{(e_1 \oplus MLE(e_1,\ldots,e_m))} \cdots p_m^{(e_m \oplus MLE(e_1,\ldots,e_m))} \\
&= Q(\mathbf{p}, MLE_{\mathbf{p}}).
\end{aligned}$$

Proof of Lemma 3.2. Let $|\hat{p}_i - p_i| < \delta$ for all $i \leq m$. Equation (A.3) implies that

$$\mathrm{P}[MLE_{\hat{\mathbf{p}}}(\xi_{1n}^1,\ldots,\xi_{mn}^1)=1] = Q(\mathbf{p}, MLE_{\hat{p}}).$$

Just as in the proof of Lemma 2.1, we can show that

$$\begin{aligned}
Q(\mathbf{p}, MLE_{\hat{\mathbf{p}}}) &= 1/2 \sum_{e_1,\ldots,e_m} p_1^{(e_1 \oplus MLE_{\hat{\mathbf{p}}}(e_1,\ldots,e_m))} \cdots p_m^{(e_m \oplus MLE_{\hat{\mathbf{p}}}(e_1,\ldots,e_m))} \\
Q(\mathbf{p}, MLE_{\mathbf{p}}) &= 1/2 \sum_{e_1,\ldots,e_m} p_1^{(e_1 \oplus MLE_{\mathbf{p}}(e_1,\ldots,e_m))} \cdots p_m^{(e_m \oplus MLE_{\mathbf{p}}(e_1,\ldots,e_m))}.
\end{aligned}$$

So it suffices to prove that for any $e_1,\ldots,e_m$,

$$\begin{aligned}
\Big| p_1^{(e_1 \oplus MLE_{\hat{\mathbf{p}}}(e_1,\ldots,e_m))} \cdots p_m^{(e_m \oplus MLE_{\hat{\mathbf{p}}}(e_1,\ldots,e_m))} \\
- p_1^{(e_1 \oplus MLE_{\mathbf{p}}(e_1,\ldots,e_m))} \cdots p_m^{(e_m \oplus MLE_{\mathbf{p}}(e_1,\ldots,e_m))} \Big| \leq 2m\delta.
\end{aligned}$$

Let us fix arbitrary $e_1,\ldots,e_m$. If

$$MLE_{\hat{\mathbf{p}}}(e_1,\ldots,e_m) = MLE_{\mathbf{p}}(e_1,\ldots,e_m)$$

we have nothing to do. Otherwise without loss of generality assume that

$$MLE_{\hat{\mathbf{p}}}(e_1,\dots,e_m)=1,\quad MLE_{\mathbf{p}}(e_1,\dots,e_m)=0.$$

This means that

$$\hat{p}_1^{(e_1\oplus 0)}\cdots\hat{p}_m^{(e_m\oplus 0)}\le\hat{p}_1^{(e_1\oplus 1)}\cdots\hat{p}_m^{(e_m\oplus 1)}$$

and

$$p_1^{(e_1\oplus 0)}\cdots p_m^{(e_m\oplus 0)}\ge p_1^{(e_1\oplus 1)}\cdots p_m^{(e_m\oplus 1)}.$$

It is easy to see that

$$|\hat{p}_1^{(e_1\oplus v)}\cdots\hat{p}_m^{(e_m\oplus v)}-p_1^{(e_1\oplus v)}\cdots p_m^{(e_m\oplus v)}|\le m\delta.$$

for any $v\in\{0,1\}$.

Therefore

$$\begin{aligned}&p_1^{(e_1\oplus 0)}\cdots p_m^{(e_m\oplus 0)}\\ &\le\ \hat{p}_1^{(e_1\oplus 0)}\cdots\hat{p}_m^{(e_m\oplus 0)}+m\delta\le\hat{p}_1^{(e_1\oplus 1)}\cdots\hat{p}_m^{(e_m\oplus 1)}+m\delta\\ &\le\ p_1^{(e_1\oplus 1)}\cdots p_m^{(e_m\oplus 1)}+2m\delta.\end{aligned}$$

In a similar way we can prove that

$$p_1^{(e_1\oplus 1)}\cdots p_m^{(e_m\oplus 1)}\le p_1^{(e_1\oplus 0)}\cdots p_m^{(e_m\oplus 0)}+2m\delta.$$

ACKNOWLEDGMENT

The authors are sincerely grateful to Pavel Naumov for helpful discussions, to Vladimir Vovk for useful comments, to Michael Gambaryan for the implementing the described algorithms and to unknown referees for helpful suggestions.

REFERENCES

Chernoff 1952. Chernoff, H. (1952), A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Annals of Mathematical Statistics*, **23** 493–509.

Shiryaev 1995. Shiryaev, A. N. (1995), "Probability. Second edition." Springer Verlag, New-York, Berlin, Heidelberg.