New Upper Bounds in the Hypothesis Testing Problem with Information Constraints

M. V. Burnashev

Kharkevich Institute for Information Transmission Problems,
Russian Academy of Sciences, Moscow, Russia

e-mail: burn@iitp.ru

Received April 10, 2020; revised May 15, 2020; accepted May 19, 2020

Abstract—We consider a hypothesis testing problem where a part of data cannot be observed. Our helper observes the missed data and can send us a limited amount of information about them. What kind of this limited information will allow us to make the best statistical inference? In particular, what is the minimum information sufficient to obtain the same results as if we directly observed all the data? We derive estimates for this minimum information and some other similar results.

Key words: hypothesis testing, information constraints, error probabilities.

DOI: 10.1134/S0032946020020027

1. INTRODUCTION AND MAIN RESULTS

1. Statement of the problem. Similarly to [1,2], we consider a binary symmetric channel BSC(\(p\)) on length \(n\) with input and output alphabets \(E = \{0, 1\}\) and unknown crossover probability \(p\). To distinguish between input and output sets of blocks \(E^n = \{0, 1\}^n\), we denote them by \(E^n_{in}\) and \(E^n_{out}\), respectively. There are two hypotheses concerning the value of \(p\) (one of them being true): \(H_0: p = p_0\) and \(H_1: p = p_1\), where \(0 < p_0, p_1 \leq 1/2\).

Denote by \(P(y|x)\) and \(Q(y|x)\) the probabilities to receive the block \(y = (y_1, \ldots, y_n)\) at the channel output provided that the input block was \(x = (x_1, \ldots, x_n)\) for the hypotheses \(H_0\) and \(H_1\), respectively. Then

\[
P(y|x) = (1 - p_0)^{n - d(x,y)}p_0^{d(x,y)}, \quad Q(y|x) = (1 - p_1)^{n - d(x,y)}p_1^{d(x,y)},
\]

where \(d(x,y)\) is the Hamming distance between blocks \(x\) and \(y\) (i.e., the number of noncoinciding components of those vectors on length \(n\)).

We consider the following problem of minimax testing of the hypotheses \(H_0\) and \(H_1\). We (i.e., a “statistician”) observe only the channel output block \(y \in E^n_{out}\), while our “helper” observes only the channel input \(x \in E^n_{in}\). It is assumed that we do not have any prior information on the input block \(x\). Clearly, based only on the output block \(y\), we cannot make any reasonable conclusions on the unknown value of \(p\).

Assume further that for a prescribed value \(R > 0\), the helper is allowed to partition in advance the input space \(E^n_{in} = \{0, 1\}^n\) into \(N \leq 2^{Rn}\) arbitrary parts \(\{X_1, \ldots, X_N\}\) and to inform us (in some additional way) only about to which part \(X_i\) the input block \(x\) belongs. Clearly, the case \(N < 2^n\), i.e., \(R < 1\), is only interesting (otherwise, the helper can simply communicate to us the block \(x\)).

For example, the helper may transmit to the statistician exact values of the first \(Rn\) components \(x_1, \ldots, x_{Rn}\) (but after that, report nothing concerning the next values \(x_i\)). However, such a simple
partitioning of the input space $E^n_{\text{in}}$ (into cylinder sets $\{X_i\}$), generally speaking, is not optimal. From the point of view of the statistician, the input data $(x_1, \ldots, x_n)$ are a very strong nuisance vector.

There are many practical situations where such a problem arises. For example, in some applications, the input block $x \in E^n_{\text{in}}$ is “nuisance noise,” which “contaminated” the output block $y \in E^n_{\text{out}}$, and therefore we would like to “reduce” (as far as possible) this “contamination” to improve statistical inferences. Of course, in this case the quality of the channel from the helper to the statistician is very important. In order to simplify the problem, here we only consider the ideal case of a noiseless channel with a limited capacity.

We can also say that the optimal limited information on the block $x \in E^n_{\text{in}}$ means the optimal “contraction” of the full information on the block $x$. Of course, such optimal “contraction” depends on available prior information on the crossover probability $p$ and on a quality criteria used.

**Remark 1.** Clearly, the problem does not change if the statistician observes the channel input, and the helper, the channel output.

Based on the observation $y \in E^n_{\text{out}}$ and the index $i$ of the part $X_i$, the statistician makes a decision in favor of one of the hypotheses $H_0$ or $H_1$. In order to avoid overcomplication, we consider only nonrandomized decision methods (the problem essence and the results remaining the same).

We are interested in partitions $\{X_1, \ldots, X_N\}$ and decision methods that are asymptotically (as $n \to \infty$) optimal. Similar but much more general problem settings were considered, e.g., in [3–8].

**Remark 2.** Anticipating things, let us note that, as far as we know, all results in this area (see, e.g., [1–8]) have the form: “it is possible to obtain the following hypothesis-testing performance: . . .”. Our goal is to prove opposite results, i.e., show that “it is impossible to get a better performance than . . .”

Below we denote $\log x = \log_2 x$. Introduce balls and spheres in $E^n$:

\[
\begin{align*}
B_x(p) &= \{u : d(x, u) \leq pn\}, \\
S_x(p) &= \{u : d(x, u) = pn\},
\end{align*}
\]

(1)

2. **Error probability exponents and the dual problem.** Let a partition $\{X_1, \ldots, X_N\}$ of the input space $E^n_{\text{in}} = \{0, 1\}^n$ be chosen. Then a general decision-making rule can be described as follows. For each partition element $X_i$, we choose a set $A(X_i) \subset E^n_{\text{out}}$, and based on the observation $y$ and the known element $X_i$, make a decision ($A^c = E^n_{\text{out}} \setminus A$):

$y \in A(X_i) \implies H_0$, \quad $y \in A^c(X_i) \implies H_1$.

Define error probabilities of the 1st kind, $\alpha_n$, and 2nd kind, $\beta_n$:

\[
\begin{align*}
\alpha_n &= \Pr(H_1 \mid H_0) = \max_{i=1, \ldots, N} \max_{x \in X_i} P(A^c(X_i) \mid x), \\
\beta_n &= \Pr(H_0 \mid H_1) = \max_{i=1, \ldots, N} \max_{x \in X_i} Q(A(X_i) \mid x).
\end{align*}
\]

Let $\gamma \geq 0$ be a given constant. We require the 1st kind error probability $\alpha_n$ to satisfy the condition

\[
\alpha_n = \Pr(H_1 \mid H_0) \leq 2^{-\gamma n}.
\]

We are interested in the minimum possible (over all partitions $\{X_i\}$ of the input space $E^n_{\text{in}}$ and all decisions) 2nd kind error probability $\min \beta_n$. We investigate the asymptotic case as $n \to \infty$.
and $N = 2^{Rn}$, where $0 < R < 1$ is a given constant.\footnote{In order to simplify formulas, we do not use the integral part sign hereinafter.} Then for the best partition $\{X_i\}$ and best decision methods, denote
\[
e(\gamma, R) = \lim_{n \to \infty} \frac{1}{n} \log_2 \frac{1}{\min \beta_n} > 0,
\]
where the minimum is over all partitions $\{X_i\}$ and decision methods satisfying condition (2).

Our main goal is to find upper bounds for the function $e(\gamma, R)$ (for lower bounds, see [1]). In this paper we restrict ourselves to the case $\gamma \to 0$, evaluating the function $e(0, R) = e(R)$ and the related function $r_{cair}(p_0, p_1)$ (this case is sometimes called the Neyman–Pearson problem). We will consider the case of $\gamma > 0$ in a separate paper.

It will be convenient for us to consider also an equivalent dual problem (without the helper). Let an $r$, $0 < r < 1$, be given, and assume that we are allowed to choose in advance any set $X \subset E^n$ consisting of $X = 2^n$ input blocks. It is also known that the input block $x$ belongs to the chosen set $X$. We observe the channel output $y$ and, having the set $X$ known, consider the problem of testing of the hypotheses $H_0$ and $H_1$. We choose a set $\mathcal{A}$ and, depending on observation $y$, make the decision:
\[
y \in \mathcal{A} \implies H_0, \quad y \in \mathcal{A}^c \implies H_1.
\]
Define the 1st kind, $\alpha_n$, and 2nd kind, $\beta_n$, error probabilities as
\[
\alpha_n = \max_{x \in X} P(\mathcal{A}^c | x), \quad \beta_n = \max_{x \in X} Q(\mathcal{A} | x).
\]

Let condition (2) for the 1st kind error probability $\alpha_n$ be fulfilled; we want to choose a set $X \subset E^n$ of cardinality $X = 2^n$ and a decision method so that to achieve the minimum possible 2nd kind error probability $\min \beta_n$. Similarly to (3), for this dual problem define the function
\[
e_d(\gamma, r) = \lim_{n \to \infty} \frac{1}{n} \log_2 \frac{1}{\min \beta_n} > 0,
\]
where the minimum is over all sets $X \subset E^n$ of cardinality $X = 2^n$ and all decision methods.

The following result establishes a simple relation between the functions $e(\gamma, R)$ and $e_d(\gamma, r)$.

**Proposition 1** [1, Proposition 1]. We have
\[
e(\gamma, 1 - R) = e_d(\gamma, R), \quad 0 \leq R \leq 1, \quad \gamma \geq 0.
\]

By Proposition 1 and equation (5), it suffices to investigate the function $e_d(\gamma, r)$. In the paper we confine ourselves with the case $\gamma \to 0$, investigating the function $e_d(0, r)$.

**Remark 3.** Essentially, in this paper we consider the case where the distributions $P(x, y)$ and $Q(x, y)$ are of the form $P(x, y) = p(x)P(y | x)$ and $Q(x, y) = p(x)Q(y | x)$, where the distribution $p(x)$ is the same for $P(x, y)$ and $Q(x, y)$. In a more general problem setting, this need not be the case.

**3. Known input block.** Assume that we know the input block $x \in E^n$ (then we may set $x = 0$) and observe an output block $y \in E^n$. If it is only required that $\alpha_n \to 0$, $n \to \infty$ (i.e., $\gamma = 0$), and we are only interested in the exponent (in $n$) of the 2nd kind error probability $\beta_n$, then, as $n \to \infty$, by the central limit theorem and Neyman–Pearson lemma, the optimal decision set in favor of $H_0$ (i.e., $p_0$) is the spherical slice $B_0(p_0 + \delta) \setminus B_0(p_0 - \delta)$ in $E^n$ (see (1)), where $\delta > 0$ is small. Then for the exponent (in $n$) of the 2nd kind error probability $\beta_n$ we have
\[
\frac{1}{n} \log \beta_n = \frac{1}{n} \log \left[ \left( \frac{p_0}{p_1} \right)^n \left( 1 - p_0 \right)^{1-p_1} n^{p_0} \right] + o(1), \quad n \to \infty,
\]
and therefore we obtain, as $n \to \infty$,
\[
\frac{1}{n} \log \frac{1}{\beta_n} = -(1 - p_0) \log(1 - p_1) - p_0 \log p_1 - h(p_0) + o(1) = D(p_0 \parallel p_1) + o(1),
\]
where $h(p) = -p \log p - (1 - p) \log (1 - p)$ and
\[
D(a \parallel b) = a \log \frac{a}{b} + (1 - a) \log \frac{1 - a}{1 - b}.
\]

Remark 4. The function $D(a \parallel b)$ from (7) is the divergence for two binomial random variables with parameters $a$ and $b$, respectively. In Russian literature it is more often referred to as the Kullback–Leibler distance. The quantity $D(a \parallel b)$ gives the best exponent for the 2nd kind error probability under a fixed 1st kind error probability (i.e., when its exponent is zero) when testing a simple hypothesis $H_0: p = a$ versus a simple alternative $H_1: p = b$.

When $\gamma = r = 0$, for $e_d(0, 0)$ (see (4)) from (6) we get
\[
e_d(0, 0) = D(p_1 \parallel p_0).
\]

4. Unknown input block and the critical rate. If we know the input block $x \in E_{in}^n$, and $\alpha_n \to 0$, then the best exponent $e_d(0, 0)$ for the 2nd kind error probability $\beta_n$ is given by (8).

If we only know that the input block $x$ belongs to a set $X \subseteq E_{in}^n$ of cardinality $X \sim 2^{rn}$, then, for the best such set $X$, the exponent $e_d(0, r)$ of the 2nd kind error probability $\beta_n$ is given by (4). It is clear that
\[
e_d(\gamma, r) \leq e_d(\gamma, 0), \quad \gamma \geq 0, \quad 0 \leq r \leq 1.
\]

The function $e_d(\gamma, r)$ is nonincreasing in $r$. Therefore, the following natural question arises: Does there exist $r(\gamma) > 0$ for which the equality in (9) holds, and, if so, what is the maximum rate $r_{\text{crit}}(\gamma)$? Confining ourselves with the case $\gamma = 0$, define the critical rate $r_{\text{crit}}(p_0, p_1) = r_{\text{crit}}(p_0, p_1, 0)$ as (see (8))
\[
r_{\text{crit}} = r_{\text{crit}}(p_0, p_1) = \sup\{r : e_d(0, r) = e_d(0, 0) = D(p_0 \parallel p_1)\}.
\]

In other words, what is the maximum cardinality $2^{rn}$ of the “best” set $X$ for which we can achieve the same asymptotic efficiency as for a known input block $x$ (though we do not know the input block $x$)?

Similarly, introduce the critical rate $R_{\text{crit}}$ for the original problem (see (3)):
\[
R_{\text{crit}}(p_0, p_1) = \inf\{R : e(0, R) = e(0, 1) = D(p_0 \parallel p_1)\}.
\]

By Proposition 1 and (11), we have
\[
R_{\text{crit}}(p_0, p_1) = 1 - r_{\text{crit}}(p_0, p_1).
\]

The main result of the paper is as follows.

Theorem 1. If $p_1 < p_0 \leq 1/2$, then there exists $p_1^*(p_0)$ such that for any $p_1 \leq p_1^*(p_0)$ we have
\[
r_{\text{crit}}(p_0, p_1) = 1 - R_{\text{crit}}(p_0, p_1) = 1 - h(p_0), \quad 0 < p_1 \leq p_1^*(p_0) < p_0 \leq 1/2.
\]

Remark 5. Although $r_{\text{crit}}(p_0, p_1)$ in (13) coincides with the capacity of the BSC$(p_0)$ channel, its origin (10) is related to the function $e_d(0, r)$, which is similar to the channel reliability function $E(r, p)$ in information theory [9, 10]. The exact form of the function $E(r, p)$ is only partially known at present [11]. Therefore, the proof of Theorem 1 (as well as [11–13]) uses rather recent results on the spectrum of binary codes. A complete description of the function $e_d(\gamma, r)$ seems to be a rather difficult problem.
In Section 2, a lower bound for $r_{\text{crit}}$ is presented (Proposition 2). In Section 3, the general formula for the 2nd kind error probability $\beta_n$ is derived (Lemma 1). In Section 4, using the method of “two hypotheses,” Theorem 1 is proved. Generally speaking, the upper bound (13) for $r_{\text{crit}}$ is weaker than the corresponding lower bound from Section 2. In Section 5, using additional combinatorial arguments, another upper bound for $r_{\text{crit}}$ (Proposition 3) is derived. In Section 6, exactness of the lower bound for $r_{\text{crit}}$ from Proposition 2 is shown provided that some additional condition is fulfilled. In the Appendix, some necessary analytic results are presented.

Throughout what follows, $f \sim g$ means $n^{-1}\ln f = n^{-1}\ln g + o(1)$, $n \to \infty$, and $f \lesssim g$ means $n^{-1}\ln f \leq n^{-1}\ln g + o(1)$, $n \to \infty$.

## 2. LOWER BOUND FOR $r_{\text{crit}}$

The following result is a consequence of [1, Proposition 2].

**Proposition 2.** For $r_{\text{crit}}(p_0, p_1)$ we have the lower bounds

$$r_{\text{crit}}(p_0, p_1) \geq 1 - h(p_0) \quad \text{if} \quad 0 < p_1 < p_0 \leq 1/2,$$

and

$$r_{\text{crit}}(p_0, p_1) \geq 1 - h(p_0) - D(p_0 \| p_1) \quad \text{if} \quad 0 < p_0 < p_1 \leq 1/2. \quad (14)$$

**Proof.** For a given $r$, $0 < r < 1$, randomly and equiprobably choose a set $\mathcal{X}$ of $2^m$ input blocks $x$. It was shown in [1, Proposition 2] that if $p_0 < p_1 \leq 1/2$, then for any $\tau$, $0 < \tau \leq p_1$, there exist $x$ and a decision method for which we have the following inequalities:

$$\frac{1}{n} \log \frac{1}{\alpha_n} \geq D(\tau \| p_0), \quad \frac{1}{n} \log \frac{1}{\beta_n} \geq \min\{D(\tau \| p_1), 1 - h(\tau) - r\}. \quad (16)$$

If it suffices to have $\alpha_n \to 0$, $n \to \infty$, then by setting $\tau = p_0$ in (16), from (10) we obtain (15).

Similarly, if $p_1 < p_0 \leq 1/2$, then by replacing $p_0$ with $p_1$ and $\alpha_n$ with $\beta_n$ in (16), for any $\tau$ we obtain

$$\frac{1}{n} \log \frac{1}{\alpha_n} \geq \min\{D(\tau \| p_0), 1 - h(\tau) - r\}, \quad \frac{1}{n} \log \frac{1}{\beta_n} \geq D(\tau \| p_1). \quad (17)$$

If $\alpha_n \to 0$, $n \to \infty$, then by setting $\tau = p_0$ in (17), from (10) we obtain (14). \(\triangle\)

## 3. GENERAL FORMULA FOR THE 2ND KIND ERROR PROBABILITY $\beta_n$

Let $\mathcal{C}_n(r) = \{x_1, \ldots, x_M\}$ be a set (code) of $M = 2^n$ different input code blocks. For the code $\mathcal{C}_n(r)$ and the 1st kind error probability $\alpha_n$, denote by $D_0 = D_0(\mathcal{C}_n, \alpha_n) \subseteq E_{\text{out}}^n$ the optimal decision set in favor of $H_0$ minimizing the 2nd kind error probability $\beta_n$. Although the set $D_0$ has a rather complicated form, we can establish some of its properties that are sufficient for proving Theorem 1.

Take a small $\delta > 0$, and for each $x_k$, $k = 1, \ldots, M$, introduce the spherical slice in $E_{\text{out}}^n$

$$SL_{x_k}(p_0, \delta) = B_{x_k}(p_0 + \delta) \setminus B_{x_k}(p_0 - \delta) = \{u : |d(x_k, u) - p_0n| \leq \delta n\}, \quad (18)$$

where $B_x(p)$ is defined in (1). For each $x_k$, introduce also the set

$$D_{x_k}(\delta) = D_0 \cap SL_{x_k}(p_0, \delta). \quad (19)$$

Since we need $\alpha_n \to 0$, $n \to \infty$, the optimal set $D_0$ contains a “substantial” part of each set $SL_{x_k}(p_0, \delta)$, $k = 1, \ldots, M$. In order to evaluate this, note that for any $x_k$ and $u, z \in SL_{x_k}(p_0, \delta)$ we have

$$P(u | p_0, x_k) = \left(\frac{q_0}{p_0}\right)^{d(z, x_k) - d(u, x_k)} \leq \left(\frac{q_0}{p_0}\right)^{2\delta n}, \quad q_0 = 1 - p_0. \quad (20)$$
By the Chebyshev exponential inequality (Chernoff bound), for any \( x_k \) and small \( \delta > 0 \) we get
\[
\log P\{u \notin SL_{x_k}(p_0, \delta) \mid x_k, p_0\} \leq -\frac{n\delta^2}{2p_0q_0}. \tag{21}
\]
Then by (18), (19), and (21) for any \( x_k \) we have
\[
P\{D_{x_k}(\delta) \mid p_0, x_k\} \geq 1 - P\{u \notin D_0 \mid p_0, x_k\} - P\{u \notin SL_{x_k}(p_0, \delta) \mid p_0, x_k\} \geq 1 - \alpha_n - e^{-n^2\delta^2/(2p_0q_0)}, \tag{22}
\]
and by (20) we also have
\[
\delta_1 \mid SL_{x_k}(p_0, \delta) \mid \leq \mid D_{x_k}(\delta) \mid \leq \mid SL_{x_k}(p_0, \delta)\mid, \quad \delta_1 = (1 - \beta_n - e^{-n^2\delta^2/(2p_0q_0)})\left(\frac{p_0}{q_0}\right)^{2\delta_n}. \tag{23}
\]
Since \( D_{x_k}(\delta) \subseteq D_0 \) for any \( x_k \), then by (19), (22), and (23) for the probability \( P(e \mid p_1, x_i) \) we have
\[
P(e \mid p_1, x_i) = P\{D_0 \mid p_1, x_i\} \sim P\left\{\bigcup_{k=1}^{M} D_{x_k}(\delta) \mid p_1, x_k\right\} \sim \delta_1 P\left\{\bigcup_{k=1}^{M} SL_{x_k}(p_0, \delta) \mid p_1, x_i\right\}. \tag{24}
\]
For \( t > 0 \) and each \( x_i \), introduce the set
\[
D_{x_i}(t, p) = \{u: \text{there exists } x_k \neq x_i \text{ such that } d(x_i, u) = tn, d(x_k, u) = pn\}. \tag{25}
\]

**Lemma 1.** For the 2nd kind error probability \( \beta_n \) of a code \( C_n = \{x_1, \ldots, x_M\} \) and the optimal set \( D_0 \) in favor of \( H_0 \), as \( n \to \infty \) we have
\[
\frac{\log \beta_n}{n} \sim \max_{t > 0} \left\{ \frac{1}{n} \log \left[ \frac{1}{M} \sum_{i=1}^{M} |D_{x_i}(t, p_0)| \right] + t \log p_1 + (1 - t) \log(1 - p_1) \right\}. \tag{26}
\]

The critical rate \( r_{\text{crit}}(p_0, p_1) \) is defined by \( (M = 2^n) \)
\[
r_{\text{crit}}(p_0, p_1) = \sup \{r : F(p_0, p_1, r) \leq 0\} = \inf \{r : F(p_0, p_1, r) > 0\}, \tag{27}
\]
where
\[
F(p_0, p_1, r) = \lim_{n \to \infty} \min_{|C_n| \leq M} \max_{t} F(p_0, p_1, r, C_n, t),
\]
\[
F(p_0, p_1, r, C_n, t) = \frac{1}{n} \log \left[ \sum_{i=1}^{M} |D_{x_i}(t, p_0)| \right] + (p_0 - t) \log \frac{1 - p_1}{p_1} - r - h(p_0). \tag{28}
\]

**Proof.** Using (24) with \( \delta = o(1) \) and \( \delta_1 = e^{o(n)} \) as \( n \to \infty \), we have
\[
\beta_n = \max_i P(e \mid p_1, x_i) \sim \frac{1}{M} \sum_{i=1}^{M} P(e \mid p_1, x_i) \sim \frac{\delta_1}{M} \sum_{i=1}^{M} P\left\{\bigcup_{k=1}^{M} SL_{x_k}(p_0, \delta) \mid p_1, x_i\right\}. \tag{29}
\]
From (25) and (26), for each \( x_i \) we have
\[
P\left\{\bigcup_{k=1}^{M} SL_{x_k}(p_0, \delta) \mid p_1, x_i\right\} \sim P\left\{\bigcup_{t>0} D_{x_i}(t, p_0) \mid p_1, x_i\right\} \sim \max_{t > 0} \left\{p_1^n(1 - p_1)^{1-t}n|D_{x_i}(t, p_0)|\right\}. \tag{30}
\]
Therefore, (29) and (30) imply (26).
Since
\[
\mathbb{P}\left\{ SL_{x_i}(p_0, \delta) \mid p_1, x_i \right\} \sim \mathbb{P}\left\{ d(x_i, u) \geq p_0n \mid p_1, x_i \right\} \sim 2^{-D(p_0 \| p_1)n},
\]
the right-hand side of (26) increases with \( r \) (i.e., with \( M = 2^n \)) starting from \(-D(p_1 \| p_0)\). Therefore, it follows from (6) and (26) that the critical rate \( r_{\text{crit}} \) is the largest rate \( r \) such that
\[
\min_{\{x_i\}} \max_{t>0} \left\{ \frac{1}{n} \log \left[ \sum_{i=1}^{M} |D_{x_i}(t, p_0)| \right] + t \log p_1 + (1-t)(1-p_1) \right\} - r \leq -D(p_0 \| p_1).
\]
Note that
\[
D(p_0 \| p_1) + t \log p_1 + (1-t)(1-p_1) = -h(p_0) + (p_0-t) \log \frac{1-p_1}{p_1}.
\]
From (31) and (32), equations (27) and (28) follow. \( \triangle \)

Note in particular that from (53) with \( t = p_0 \) we have
\[
F(p_0, p_1, r, C_n, p_0) = o(1), \quad n \to \infty.
\]

The main difficulty in the analysis of relations (27) and (28) consists in evaluation of the cardinalities \(|D_{x_i}(t, p_0)|\) in (28), which depend on the geometry of the code \( C_n \). A similar problem appeared in [11–13], where the reliability function \( E(R, p) \) of the BSC(\( p \)) channel was investigated. Direct estimation of those cardinalities leads to very cumbersome formulas.

4. UPPER BOUND FOR \( r_{\text{crit}} \): TWO HYPOTHESES

Now we derive a simple (but not too tight) upper bound for \( r_{\text{crit}}(p_0, p_1) \) by using the method of “two hypotheses,” quite popular in mathematical statistics (mainly, in estimation theory). Using (26), choose any two codewords from the code \( C_n(r) = \{x_1, \ldots, x_M\}, M = 2^n \), say \( x_1 \) and \( x_2 \), with \( d(x_1, x_2) = \omega n \). We may assume that for the rate \( r > 0 \), the parameter \( \omega \) satisfies the constraints
\[
0 < \omega \leq \omega_{\text{min}}(r),
\]
where \( \omega_{\text{min}}(r) \) will be specified later. Replace the code \( C_n(r) \) by a code \( C' \) consisting of the two chosen codewords \( C' = \{x_1, x_2\} \). Then \( \beta_n(C) \geq \beta_n(C') \). Similarly to (29) and (30), we have
\[
\beta_n(C') \sim 2^{-D(p_0 \| p_1)n} + \mathbb{P}\left\{ SL_{x_2}(p_0, \delta) \mid p_1, x_1 \right\}.
\]
We are interested in when \( x_1 \) and \( x_2 \) satisfy the inequality
\[
\frac{1}{n} \log \mathbb{P}\left\{ SL_{x_2}(p_0, \delta) \mid p_1, x_1 \right\} > -D(p_0 \| p_1).
\]
Let us evaluate the probability on the left-hand side of (33). For \( d(x_i, x_k) = \omega n \), denote
\[
S_{x_i, x_k}(t, p, \omega) = \{ u : d(x_i, u) = tn, d(x_k, u) = pn, d(x_i, x_k) = \omega n \}.
\]
Then (see the Appendix)
\[
\frac{1}{n} \log |S_{x_i, x_k}(t, p, \omega)| = g(t, p, \omega) + o(1), \quad n \to \infty,
\]
\[
\frac{1}{n} \log \mathbb{P}\left\{ S_{x_i, x_k}(t, p, \omega) \mid p_1, x_i \right\} = g(t, p, \omega) - t \log \frac{1-p_1}{p_1} + \log(1-p_1) + o(1),
\]

PROBLEMS OF INFORMATION TRANSMISSION Vol. 56 No. 2 2020
where \( g(t,p,\omega) \) is defined in (78). Therefore, as \( n \to \infty \) (see (76) and (77)),

\[
\frac{1}{n} \log P \left\{ SL_{x_2}(p_0, \delta) \mid p_1, x_1 \right\} = \frac{1}{n} \max_t \log P \left\{ S_{x_1,x_2}(t,p_0,\omega) \mid p_1, x_1 \right\} + o(1)
\]

\[
= f(p_0,p_1,\omega) + o(1),
\]

where

\[
f(p_0,p_1,\omega) = \max_t f(p_0,p_1,\omega,t),
\]

\[
f(p_0,p_1,\omega,t) = g(t,p_0,\omega) - t \log \frac{1-p_1}{p_1} + \log(1-p_1).
\]

We have

\[
f_t'(p_0,p_1,\omega,t) = \log \frac{\omega-t}{t} - \log \frac{p_0+t-\omega}{1-p_0-t} - \frac{1-p_1}{p_1}, \quad f_t''(p_0,p_1,\omega,t) < 0. \tag{38}
\]

By (32) and (35)–(37), inequality (33) takes the form

\[
\max_t F(p_0,p_1,\omega,t) > 0, \tag{39}
\]

where

\[
F(p_0,p_1,\omega,t) = f(p_0,p_1,\omega,t) + D(p_0 \parallel p_1)
\]

\[
= g(t,p_0,\omega) + (p_0 - t) \log \frac{1-p_1}{p_1} - h(p_0). \tag{40}
\]

If some \( p_0, p_1, \) and \( \omega \) satisfy inequality (39), then an appropriate upper bound (14) or (15) holds.

Denote by \( t_1^0 = t_1^0(p_0,p_1,\omega) \) the maximizing value \( t \) in (37) (it remains to be maximizing in (39) too). Then

\[
f(p_0,p_1,\omega) = f(p_0,p_1,\omega,t_1^0(p_0,p_1,\omega)). \tag{41}
\]

From the equation \( f_t'(p_0,p_1,\omega,t) = 0 \) for \( t_1^0 \), from (38) we obtain

\[
t_1^0 = t_1^0(p_0,p_1,\omega) = \sqrt{1+(v_0-1)[(\omega-p_0)^2 v_0 - (1-\omega-p_0)^2 + 1]} - 1
\]

\[
v_0(p_1) = \left( \frac{1-p_1}{p_1} \right)^2 \geq 1. \tag{42}
\]

Then from (40) and (42) we have

\[
F(p_0,p_1,\omega,t_1^0) = g(t_1^0,p_0,\omega) + (p_0 - t_1^0) \log \frac{1-p_1}{p_1} - h(p_0). \tag{43}
\]

One can check that for the function \( F(p_0,p_1,\omega,t_1^0) \), from (43) we have \( F(p_0,p_1,0,t_1^0) = 0 \) and \( F_\omega'' < 0, \omega > 0 \). Therefore, it suffices to check inequality (39) with \( t = t_1^0 \) for the minimum value of \( \omega \) for the code \( C_n(r) \) only (i.e., for its code distance \( d(C) \)).

Let \( \omega_{\min}(r)n \) be the maximum possible code distance of \( C_n(r) \). For \( \omega_{\min}(r) \), the following bound is known [14, equation (1.5)]:

\[
r \leq h \left[ \frac{1}{2} - \sqrt{\omega_{\min}(1 - \omega_{\min})} \right], \quad \omega_{\min} = \omega_{\min}(r). \tag{44}
\]

Consider two possible cases: (1) \( p_1 < p_0 \leq 1/2 \) and (2) \( p_0 < p_1 \leq 1/2 \).
(1) Case $p_1 < p_0 \leq 1/2$. Setting $r = 1 - h(p_0)$, denote by $\omega_0 = \omega_0(p_0)$ the root of the equation (see (44))

$$1 - h(p_0) = h\left[\frac{1}{2} - \sqrt{\omega(1 - \omega)}\right].$$

Then inequality (39) takes the form $(\omega_0 = \omega_0(p_0))$

$$F(p_0, p_1, \omega_0, t_1^0) = g(t_1^0, p_0, \omega_0) + (p_0 - t_1^0)\log \frac{1-p_1}{p_1} - h(p_0) > 0. \quad (45)$$

One can check (using Maple) that inequality (45) is satisfied if $p_1 \leq p_1^0(p_0)$, where

| $p_0$  | 0.1  | 0.12 | 0.15 | 0.2  | 0.3  | 0.4  | 0.45 | 0.49 |
|--------|------|------|------|------|------|------|------|------|
| $p_1^0(p_0)$ | 0.0003 | 0.003 | 0.016 | 0.056 | 0.17 | 0.317 | 0.4  | 0.48 |

For the case of $p_0 \leq 0.20707$ (i.e., $\omega < 0.273$), in [14, equation (1.4)] there is a slightly more precise bound than (44) (though much more bulky).

(2) Case $p_0 < p_1 \leq 1/2$. One can check that inequality (39) is not satisfied for any $p_0 < p_1$!

5. UPPER BOUND FOR $r_{\text{crit}}$: COMBINATORICS

Now we derive one more upper bound for $r_{\text{crit}}$, based on the same equation (26) but using additional combinatorial arguments.

1. Combinatorial lemma. In the code $C_n = \{x_i\}$, we call $(x_i, x_j)$ an $\omega$-pair if $d(x_i, x_j) = \omega n$. We say that a point $y \in E^n$ is $(\omega, p, t)$-covered if there exists an $\omega$-pair $(x_i, x_j)$ such that $d(x_i, y) = pn$ and $d(x_j, y) = tn$. Denote by $K(y, \omega, p, t)$ the number of $(\omega, p, t)$-coverings of a point $y$ (taking into account the multiplicity of coverings), i.e.,

$$K(y, \omega, p, t) = |\{(x_i, x_j) : d(x_i, x_j) = \omega n, d(x_i, y) = pn, d(x_j, y) = tn\}|, \quad \omega > 0. \quad (46)$$

Introduce the sets (see (25))

$$D_{x_i}(t, p, \omega) = \bigcup_{x_k} S_{x_i, x_k}(t, p, \omega)$$

$$= \{u : \text{there exists } x_k \text{ such that } d(x_i, x_k) = \omega n, d(x_i, u) = tn, d(x_k, u) = pn\}. \quad (47)$$

Then

$$D_{x_i}(t, p) = \bigcup_{\omega > 0} D_{x_i}(t, p, \omega).$$

For $t > 0$, introduce the quantity

$$m_t(y) = |\{x_i : x_i \in S_y(t)\}|. \quad (48)$$

Then for any $y, p, t > 0$ we have

$$K(y, t, p) = m_t(y)m_p(y). \quad (49)$$

Lemma 2. For a code $\{x_i\}$ and $\omega, p, t > 0$ we have the inequality (see (46) and (47))

$$\sum_{i=1}^M |D_{x_i}(t, p, \omega)| \leq \sum_{y \in E^n} K(y, \omega, t, p). \quad (50)$$
Also, if (see (48))
\[ \max_y m_p(y) = 2^{o(n)}, \quad n \to \infty, \tag{51} \]
then for any \( \omega, t > 0, \)
\[ \sum_{i=1}^{M} |D_{x_i}(t, p, \omega)| = 2^{o(n)} \sum_{y \in E^n} K(y, \omega, t, p), \quad n \to \infty. \tag{52} \]

**Proof.** Let \( y \in E^n \) and assume that there are \( m \) ordered pairs \((x_i, x_j)\) with \( d(x_i, x_j) = \omega n, \)
\( d(x_i, y) = t n, \) and \( d(x_j, y) = p n. \) These \( m \) pairs \((x_i, x_j)\) have \( m_1 \leq m \) different first arguments \( \{x_i\}. \) Then \( y \) occurs \( m \) times on the right-hand side of (50) and \( m_1 \) times on the left-hand side, which proves (50). If the condition (51) is satisfied, then \( m_1 = me^{o(n)}, \) whence (52) follows. Note also that by (49) we have
\[ \sum_{i=1}^{M} |D_{x_i}(t, p)| = \sum_{y : m_p(y) \geq 1} K(y, t, p) = \sum_{y : m_p(y) \geq 1} m_t(y) \sim M2^{h(t)n} - \sum_{y : m_p(y) = 0} m_t(y). \tag{53} \]

From the first equality in (53), equations (50) and (52) follow as well. △

Equation (53) looks simple and attractive, but its right-hand side is of the form large minus large, which is inconvenient. Note that in (53) we cannot neglect the last sum, because then we get only \( r_{\text{crit}} \leq 1, \) which is of no interest.

2. **One more upper bound for \( r_{\text{crit}}.** We upper bound the last sum in (53) as follows. We have
\[ \sum_{y : m_p(y) = 0} m_t(y) \leq 2^{h(t)n} |\{y : m_p(y) = 0\}|. \tag{54} \]
The maximum of the cardinality \( |\{y : m_p(y) = 0\}| \) is attained when the code \( C \) is the ball \( B_0(\tau) \) of radius \( \tau n, \) where \( r = h(\tau). \) Therefore,
\[ \max_C |\{y : m_p(y) = 0\}| = 2^n \quad \text{attained when } \quad C = B_0(\tau + p_0), \quad \tau + p_0 \geq 1/2, \]
\[ \tau + p_0 \leq 1/2. \tag{55} \]

If \( \tau + p_0 \geq 1/2, \) i.e., if \( r \geq h(1/2 - p_0), \) then from (53)–(55) we get
\[ \sum_{i=1}^{M} |D_{x_i}(t, p_0)| \geq 2^{h(t)n} [M - 2^{h(\tau + p_0)n}] = 2^{h(t)n} [2^{h(\tau)n} - 2^{h(1-\tau - p_0)n}] \sim M2^{h(t)n} \]
if \( \tau > 1 - \tau - p_0, \) i.e., if \( \tau > (1 - p_0)/2, \) or equivalently, if \( r > h[(1 - p_0)/2]. \)

Therefore, if \( r \geq \max\{h(1/2 - p_0), h[(1 - p_0)/2]\} = h[(1 - p_0)/2], \) then for any \( p_0 \neq p_1, \) equation (28) takes the form
\[ F(p_0, p_1, r) = \max_{t > 0} \left\{ h(t) + (p_0 - t) \log \frac{1 - p_1}{p_1} \right\} - h(p_0) \]
\[ = h(p_1) + (p_0 - p_1) \log \frac{1 - p_1}{p_1} - h(p_0) > 0, \quad p_0 \neq p_1, \]

since the maximum over \( t \) is attained at \( t = p_1. \) Therefore, this gives the following upper bound for \( r_{\text{crit}} \) (weaker than (13)):
\[ r_{\text{crit}}(p_0, p_1) \leq h[(1 - p_0)/2], \quad p_0 \neq p_1. \tag{56} \]
Remark 6. Note that $1 - h(p_0) < h(1/2 - p_0) < h(1 - p_0)/2$, $0 < p_0 < 1/2$.

Let us improve the bound (56). In addition to (54), we also have

$$\sum_{y: m_p(y) = 0} m_t(y) \leq M |\{y : m_p(y) = 0\}|.$$ 

Therefore, if $\tau + p_0 \geq 1/2$ and $t \geq 1 - \tau - p_0$, then

$$\sum_{i=1}^{M} |D_{x_i}(t, p_0)| \geq M \left[ 2^{h(t)n} - 2^{h(1-\tau-p_0)n} \right] \sim M 2^{h(t)n}.$$ 

By (39) and (40), we should necessarily have

$$\max_{t \geq 1 - \tau - p_0} f(t, p_0, p_1) > 0,$$

$$f(t, p_0, p_1) = h(t) + (p_0 - t) \log \frac{1 - p_1}{p_1} - h(p_0). \tag{57}$$

The maximum of $f(t, p_0, p_1)$ over $t \geq 1 - \tau - p_0$ is attained at $t = \max\{p_1, 1 - \tau - p_0\}$, since

$$\max_t f(t, p_0, p_1) = f(p_1, p_0, p_1) > 0, \quad p_0 \neq p_1, \quad f(p_0, p_0, p_1) = 0,$$

$$f'_t(t, p_0, p_1) = \log \frac{1 - t}{t} - \log \frac{1 - p_1}{p_1}, \quad f''_t(t, p_0, p_1) < 0, \tag{58}$$

$$\text{sign} f'_t(t, p_0, p_1) = \text{sign}(p_1 - t).$$

Therefore, if $p_1 \geq 1 - \tau - p_0$, then from (57) and (58) for $p_0 \neq p_1$ we obtain

$$\max_{t \geq 1 - \tau - p_0} f(t, p_0, p_1) = h(p_1) + (p_0 - p_1) \log \frac{1 - p_1}{p_1} - h(p_0) > 0. \tag{59}$$

Hence, if $\tau \geq \max\{1/2 - p_0, 1 - p_0 - p_1\} = 1 - p_0 - p_1$, then for $p_0 \neq p_1$, inequality (59) holds, which implies the estimate

$$\tau_{\text{crit}} \leq 1 - p_0 - p_1, \quad r_{\text{crit}} = h(\tau_{\text{crit}}). \tag{60}$$

If $p_1 < 1 - \tau - p_0$, then the maximum in (57) is attained at $t = 1 - \tau - p_0$, and then

$$\max_{t \geq 1 - \tau - p_0} f(t, p_0, p_1) = f(1 - \tau - p_0, p_0, p_1).$$

Note that

$$f(p_0, p_0, p_1) = 0, \quad f'_t(p_0, p_0, p_1) \neq 0, \quad p_0 \neq p_1,$$

$$f''_t(p_0, p_0, p_1) < 0, \quad \text{sign} f'_t(p_0, p_0, p_1) = \text{sign}(p_1 - t).$$

Let also $p_0 > 1 - \tau - p_0$ (i.e., $\tau > 1 - 2p_0$). Then $\max_{t \geq 1 - \tau - p_0} f(t, p_0, p_1) > 0$ (it suffices to choose $t$ close to $p_0$). Therefore,

$$\tau_{\text{crit}} \leq 1 - 2p_0, \quad r_{\text{crit}} = h(\tau_{\text{crit}}). \tag{61}$$

As a result, from (60) and (61) we obtain the following.

**Proposition 3.** For any $p_0, p_1 \in [0, 1/2]$, for $r_{\text{crit}}$ we have the upper bound

$$\tau_{\text{crit}}(p_0, p_1) \leq \min \{1 - p_0 - p_1, 1 - 2p_0\}, \quad r_{\text{crit}} = h(\tau_{\text{crit}}). \tag{62}$$

**Corollary.** If $p_0 = 1/2$, then it follows from (62) that $\tau_{\text{crit}}(1/2, p_1) = r_{\text{crit}}(1/2, p_1) = 0$.

This particular result was proved earlier in [1, Proposition 3] using another method. The best exponent $e_d(\gamma, r)$ from (4) for $\gamma \geq 0$ and $0 \leq r \leq 1$ was also obtained there.
6. “POTENTIAL” ADDITIVE UPPER BOUND FOR $r_{\text{crit}}$

Theorem 1 was proved by replacing the exponential number $M$ of codewords $\{x_i\}$ in (26) with two closest codewords $(x_i, x_j)$. Such a method gives optimal results only if it is possible to choose a pair $(x_i, x_j)$ with $d(x_i, x_j) = \omega n$ and small $\omega > 0$. In our problem setting, we cannot do this.

In order to improve Theorem 1, we need to consider an exponential number $M$ of codewords $\{x_i\}$ in (26), which is much more difficult (see [11–13]). We improve Theorem 1 provided that it is possible to use an additive approximation in (26).

Assume that, as $n \to \infty$, for all $\{x_i\}$ in (26) we have an additive approximation

$$P \left( \bigcup_{k \neq i} SL_{x_k}(p_0, \delta) \mid p_1, x_i \right) = 2^{o(n)} \sum_{k \neq i} P \left( SL_{x_k}(p_0, \delta) \mid p_1, x_i \right).$$

(63)

Then (see (36)), with $d(x_i, x_k) = \omega_{ik} n$,

$$P \left( \bigcup_{k \neq i} SL_{x_k}(p_0, \delta) \mid p_1, x_i \right) = 2^{o(n)} \sum_{k \neq i} 2f(p_0, p_1, \omega_{ik}) n$$

and

$$\sum_{i=1}^{M} P \left( \bigcup_{k \neq i} SL_{x_k}(p_0, \delta) \mid p_1, x_i \right) = 2^{o(n)} \sum_{i=1}^{M} \sum_{k \neq i} 2f(p_0, p_1, \omega_{ik}) n.$$

(64)

In order to further develop (64), we introduce some additional notions. The code spectrum (distance distribution) $B(C) = (B_0, B_1, \ldots, B_n)$ of a code $C$ of length $n$ is the $(n+1)$-tuple with components

$$B_i = |C|^{-1} |\{(x, y) : x, y \in C, d(x, y) = i\}|, \quad i = 0, 1, \ldots, n.$$

(65)

In other words, $B_i$ is the average number of codewords $y$ at distance $i$ from the codeword $x$. The total number of ordered code pairs $(x, y) \in C$ with $d(x, y) = i$ is $|C|B_i$. Denote also $B_{\omega n} = 2^{b(\omega, r)n}$.

Then we can continue (64) as follows:

$$\sum_{i=1}^{M} P \left( \bigcup_{k \neq i} SL_{x_k}(p_0, \delta) \mid p_1, x_i \right) = 2^{o(n)} M \sum_{\omega > 0} 2^{b(\omega, r) + f(p_0, p_1, \omega)} n.$$

Therefore (see (36) and (37)),

$$\frac{1}{n} \log \left[ \sum_{i=1}^{M} P \left( \bigcup_{k \neq i} SL_{x_k}(p_0, \delta) \mid p_1, x_i \right) \right] = r + \max_{\omega, t} \{b(\omega, r) + f(p_0, p_1, \omega, t)\} + o(1),$$

(66)

where $f(p_0, p_1, \omega, t)$ is defined in (37). Then for the function $F(p_0, p_1, r)$, from (28) and (66) we have

$$F(p_0, p_1, r) = \max_{\omega, t} \left\{ b(\omega, r) + g(p_0, t, \omega) + (p_0 - t) \log \frac{1 - p_1}{p_1} - h(p_0) \right\}. $$

(67)

As an estimate for $b(\omega, r)$ in (67), we use a function $b_{\text{low}}(\omega, r)$ with the following property: there exists $\omega_{\text{max}} = \omega_{\text{max}}(r) > 0$ such that

$$\max_{0 < \omega \leq \omega_{\text{max}}} [b(\omega, r) - b_{\text{low}}(\omega, r)] \geq 0, \quad r > 0.$$

(68)

Then, for the inequality $F(p_0, p_1, r) > 0$ (see (27)) to be valid, the following condition is sufficient (see (37) and (67)):

$$\min_{0 < \omega \leq \omega_{\text{max}}} \max_{t > 0} \left\{ b_{\text{low}}(\omega, r) + g(p_0, t, \omega) + (p_0 - t) \log \frac{1 - p_1}{p_1} - h(p_0) \right\} > 0.$$

(69)
As \( b_{\text{low}}(\omega, r) \) in (69), we use the best of such known functions \( \mu(r, \alpha, \omega) \), \( h_2(\tau) = h_2(\alpha) - 1 + r \), with an arbitrary \( \alpha \in [\delta_{\text{GV}}(r), 1/2] \) (see (81), (82), and Theorem 2 in the Appendix). The function \( \mu(r, \alpha, \omega) \) satisfies condition (68). Moreover, it monotonically increases in \( r \), and \( \omega_{\text{max}} = G(\alpha, \tau) \), where \( G(\alpha, \tau) \) is defined in (79). Then for inequality (69) to be satisfied, it suffices that

\[
\min_{0 < \omega \leq \omega_{\text{max}}} \max_{t > 0} K(p_0, p_1, r, \omega, t) > 0, \tag{70}
\]

where

\[
K(p_0, p_1, r, \omega, t) = \mu(r, p_0, \omega) + g(p_0, t, \omega) + (p_0 - t) \log \frac{1 - p_1}{p_1} - h(p_0).
\]

Note that \( K(p_0, p_1, r, 0, p_0) = 0 \). In order to avoid bulky computations, we set \( t = p_0 \). The function \( K(p_0, p_1, r, \omega, p_0) \) is concave in \( \omega \), i.e., \( K''(p_0, p_1, r, \omega, p_0)_{\omega\omega} < 0 \) (the simplest way to check this is to use Maple). Therefore, the minimum over \( \omega \) is attained at \( \omega = \omega_{\text{max}} = G(\alpha, \tau) \), and it suffices to check condition (70) for \( \omega = G(\alpha, \tau) \). The following useful relation [11, Lemma 4] is known:

\[
\mu(r, \alpha, G(\alpha, \tau)) = h_2(G(\alpha, \tau)) + r - 1, \quad h_2(\alpha) - h_2(\tau) = 1 - r.
\]

Below we only consider the following simpler case.

**Case** \( p_1 < p_0 \leq 1/2 \). Set \( r = r_0 = 1 - h(p_0) \) and \( \alpha = p_0 \) (note that in this case \( \delta_{\text{GV}}(r_0) = p_0 \) and \( \tau = 0 \)). Then \( G(\alpha, \tau) = 2p_0(1 - p_0) \), and it suffices to check (70) for \( \omega = 2p_0(1 - p_0) \). From (71) and (72) with \( \alpha = p_0, \tau = 0, r = r_0 = 1 - h(p_0), t = p_0, \) and \( \omega_{\text{max}} = G(\alpha, \tau) = 2p_0(1 - p_0) \), we have

\[
K(p_0, p_1, 1 - h(p_0), \omega_{\text{max}}, p_0) = h_2(\omega_{\text{max}}) + g(p_0, p_0, \omega_{\text{max}}) - 2h(p_0),
\]

where

\[
g(p, p, 2p(1 - p)) = 2p(1 - p) + (1 - 2p(1 - p))h\left[\frac{p^2}{1 - 2p(1 - p)}\right].
\]

One can check that for \( \omega_0 = 2p_0(1 - p_0) \) we have the equality

\[
K(p_0, p_1, 1 - h(p_0), \omega_0, p_0) = h_2(\omega_0) + \omega_0 + (1 - \omega_0)h\left[\frac{p_0^2}{1 - \omega_0}\right] - 2h(p_0) = 0. \tag{73}
\]

We also have

\[
[K(p_0, p_1, 1 - h(p_0), \omega_0, t)]' = \frac{1}{2} \log \frac{(1 - t)^2 - (1 - \omega_0 - p_0)^2}{t^2 - (\omega_0 - p_0)^2} - \log \frac{1 - p_1}{p_1},
\]

\[
[K(p_0, p_1, 1 - h(p_0), \omega_0, t)]'' \bigg|_{t=p_0} < 0.
\]

Therefore, for \( t = p_0 \) we have

\[
[K(p_0, p_1, 1 - h(p_0), \omega_0, t)]'_{t=p_0} = \log \frac{1 - p_0}{p_0} - \log \frac{1 - p_1}{p_1} < 0, \quad p_1 < p_0.
\]

It follows from (73)–(75) that

\[
K(p_0, p_1, 1 - h(p_0), \omega_0, t) > 0, \quad t < p_0.
\]

Therefore, inequality (70) holds for any \( r > r_0 = 1 - h(p_0) \) and \( p_1 < p_0 \leq 1/2 \).

Thus, we get the following conditional result.

**Proposition 4.** If the additive approximation (63) holds, then \( r_{\text{crit}}(p_0, p_1) = 1 - h(p_0), 0 < p_1 < p_0 \leq 1/2 \).
Remark 7. One can show that Theorem 1 and equation (13) hold for any \( p_1 < p_0 \leq 1/2 \). To this end, one can proceed similarly to [11], using Lemma 2 and considering separately the case of equality in (50) (in essence, this is equivalent to the case considered in Section 6) and the case of inequality in it. The proof in the second case turns out to be too bulky (and oriented to the binary channel BSC(\( p \)) only). Therefore, we omit this proof. Apparently, a simpler proof should exist.

**APPENDIX**

1. Function \( g(t, p, \omega) \) and equation (35). Consider codewords \( x = 0 \) and \( x_1 \) with \( d(x, x_1) = w(x_1) = \omega n \), and also the set \( S_{x, x_1}(t, p, \omega) \) from (34). We may assume that \( x_1 = (1, \ldots, 1, 0, \ldots, 0) \), and \( x_1 \) has first \( \omega n \) “ones” and then \( (1 - \omega)n \) “zeros.” Let also \( u \in S_{x, x_1}(t, p, \omega) \) have \( u_1 n \) “ones” in the first \( \omega n \) positions and then \( u_2 n \) “ones” in the next \( (1 - \omega)n \) positions. Since \( u_1 + u_2 = t \) and \( \omega - u_1 + u_2 = p \), we have

\[
    u_1 = \frac{t - p + \omega}{2}, \quad u_2 = \frac{t + p - \omega}{2},
\]

and as \( n \to \infty \) we obtain

\[
    \frac{1}{n} \log |S_{x, x_1}(t, p, \omega)| = \frac{1}{n} \log \left( \frac{\omega n}{u_1 n} \right) \left( \frac{(1 - \omega)n}{u_2 n} \right) = \omega h\left( \frac{u_1}{\omega} \right) + (1 - \omega) h\left( \frac{u_2}{1 - \omega} \right) + o(1) = g(t, p, \omega) + o(1),
\]

where

\[
    g(t, p, \omega) = \omega h\left( \frac{t + \omega - p}{2\omega} \right) + (1 - \omega) h\left( \frac{t + p - \omega}{2(1 - \omega)} \right).
\]

We also have

\[
    2g'_p(p, t, \omega) = -2 \log \frac{1 - \omega}{\omega} + \log \frac{(1 - \omega)^2 - (1 - t - p)^2}{\omega^2 - (t - p)^2},
\]

\[
    2g'_t(p, t, \omega) = \log \frac{(1 - t)^2 - (1 - \omega - p)^2}{t^2 - (\omega - p)^2}, \quad g''_u(p, t, \omega) < 0, \quad g''_{u\omega}(p, t, \omega) \leq 0.
\]

For the root \( \omega_0 \) of the equation \( g'(t, p, \omega) = 0 \), we have

\[
    \omega_0 = \frac{p - t}{1 - 2t}, \quad g(t, p, \omega_0) = h(t).
\]

2. Function \( \mu(R, \alpha, \omega) \). Introduce the function [14] \((0 \leq \tau \leq \alpha \leq 1/2)\)

\[
    G(\alpha, \tau) = 2frac{\alpha(1 - \alpha) - \tau(1 - \tau)}{1 + 2\sqrt{\tau(1 - \tau)}} \geq 0.
\]

For \( \alpha \) and \( \tau \) such that \( 0 \leq \tau \leq \alpha \leq 1/2 \) and \( h_2(\alpha) - h_2(\tau) = 1 - R \), introduce the function [16]

\[
    \mu(R, \alpha, \omega) = h_2(\alpha) - 2 \int_0^{\omega/2} \log \frac{P + \sqrt{P^2 - 4Qy^2}}{Q} dy - (1 - \omega) h_2\left( \frac{\alpha - \omega/2}{1 - \omega} \right),
\]

\[
    P = \alpha(1 - \alpha) - \tau(1 - \tau) - y(1 - 2y), \quad Q = (\alpha - y)(1 - \alpha - y).
\]

Define the function \( \delta_{GV}(R) \leq 1/2 \) (Gilbert–Varshamov bound) as

\[
    1 - R = h_2(\delta_{GV}(R)), \quad 0 \leq R \leq 1.
\]
The importance of the function \( \mu(R, \alpha, \omega) \) and its relation to the code spectrum \( \{B_i\} \) (see (65)) is described by the following version of Theorem 3 from [15].

**Theorem 2** [15, Theorem 3]. For any \((R, n)\) code and any \( \alpha \in [\delta_{GV}(R), 1/2] \) there exist \( r_1(R, \alpha) > 0 \) and \( \omega, \, 0 < r_1(R, \alpha) \leq \omega \leq G(\alpha, \tau) \), where \( h_2(\tau) = h_2(\alpha) - 1 + R \) and \( G(\alpha, \tau) \) is defined in (79), such that

\[
n^{-1} \log B_{\omega n} \geq \mu(R, \alpha, \omega) + o(1), \quad n \to \infty. \tag{82}
\]

For \( \mu(R, \alpha, \omega) \) from (80), the nonintegral representation (83)–(85) also holds.

**Remark 8.** Theorem 2 refines Theorem 5 from [16] (see also [12, Theorem 2]). With \( r_1 = 0 \), Theorem 2 turns into Theorem 5 from [16]. In [15, Theorem 3] one can find estimates for \( r_1(R, \alpha) > 0 \).

**Proposition 5** [11, Proposition 3]. For the function \( \mu(R, \alpha, \omega) \), we have the representation

\[
\mu(R, \alpha, \omega) = (1 - \omega) h_2\left(\frac{\alpha - \omega/2}{1 - \omega}\right) - h_2(\alpha) + 2h_2(\omega) + \omega \log \frac{2\omega}{e} - T(A, B, \omega), \tag{83}
\]

where

\[
T(A, B, \omega) = \omega \log(v - 1) - (1 - \omega) \log \frac{v^2 - A^2}{v^2 - B^2} + B \log \frac{v + B}{v - B} - A \log \frac{v + A}{v - A} - \frac{(v - 1)(B^2 - A^2)}{(v^2 - B^2) \ln 2}, \quad v = \sqrt{B^2 \omega^2 - 2a_1 \omega + a_1^2 + a_1}, \quad a_1 = \frac{B^2 - A^2}{2}, \tag{84}
\]

and

\[
h_2(\alpha) - h_2(\tau) = 1 - R, \quad A = 1 - 2\alpha, \quad B = 1 - 2\tau, \quad 0 \leq \tau \leq \alpha \leq 1/2. \tag{85}
\]

For any \( \alpha_0(R) \leq \alpha < 1/2 \) and \( \omega > 0 \), we have

\[
\frac{d\mu(R, \alpha, \omega)}{d\alpha} > 0, \quad \alpha_0(R) = h_2^{-1}(1 - R).
\]

For any \( \alpha > 0 \) and \( R > 0 \), we also have \( \mu(R, \alpha, 0) = 0 \) and \( \mu'(R, \alpha, \omega)|_{\omega=0} > 0 \). Moreover, for any \( 0 \leq \tau \leq \alpha \leq 1/2 \) and \( 0 < \omega < G(\alpha, \tau) \),

\[
\mu''(R, \alpha, \omega) > 0.
\]

For any \( \omega > 0 \), we have \( \mu(0, 1/2, \omega) = 0 \).

**ACKNOWLEDGEMENTS**

The author is grateful to Shun Watanabe and a reviewer for useful discussions and constructive critical remarks, which improved the paper.

**FUNDING**

The research was supported in part by the Russian Foundation for Basic Research, project no. 19-01-00364.
REFERENCES

1. Burnashev, M.V., Amari, S., and Han, T.S., On Some Problems of Hypothesis Testing with Information Constraints, *Teor. Veroyatnost. i Primenen.*, 2000, vol. 45, no. 4, pp. 625–638 [Theory Probab. Appl. (Engl. Transl.), 2002, vol. 45, no. 4, pp. 558–568].

2. Burnashev, M.V., Han, T.S., and Amari, S., On Some Estimation Problems with Information Constraints, *Teor. Veroyatnost. i Primenen.*, 2001, vol. 46, no. 2, pp. 233–246 [Theory Probab. Appl. (Engl. Transl.), 2003, vol. 46, no. 2, pp. 214–225].

3. Ahlswede, R. and Csiszár, I., Hypothesis Testing with Communication Constraints, *IEEE Trans. Inform. Theory*, 1986, vol. 32, no. 4, pp. 533–542.

4. Han, T.S. and Kobayashi, K., Exponential-type Error Probabilities for Multiterminal Hypothesis Testing, *IEEE Trans. Inform. Theory*, 1989, vol. 35, no. 1, pp. 2–14.

5. Ahlswede, R. and Burnashev, M.V., On Minimax Estimation in the Presence of Side Information about Remote Data, *Ann. Statist.*, 1990, vol. 18, no. 1, pp. 141–171.

6. Han, T.S. and Amari, S., Statistical Inference under Multiterminal Data Compression, *IEEE Trans. Inform. Theory*, 1998, vol. 44, no. 6, pp. 2300–2324.

7. Shimokawa, H., Han, T.S., and Amari, S., Error Bounds of Hypothesis Testing with Data Compression, in *Proc. 1994 IEEE Int. Symp. on Information Theory (ISIT'94), Trondheim, Norway, June 27–July 1, 1994*, p. 114.

8. Watanabe, S., Neyman–Pearson Test for Zero-Rate Multiterminal Hypothesis Testing, in *Proc. 2017 IEEE Int. Sympos. on Information Theory (ISIT'2017), Aachen, Germany, June 25–30, 2017*, pp. 116–120.

9. Elias, P., Coding for Noisy Channels, *IRE Conv. Rec.*, 1955, vol. 4, pp. 37–46. Reprinted in *Key Papers in the Development of Information Theory*, Slepian, D., Ed., New York: IEEE Press, 1974, pp. 102–111.

10. Gallager, R.G., *Information Theory and Reliable Communication*, New York: Wiley, 1968.

11. Burnashev, M.V., On the BSC Reliability Function: Expanding the Region Where It Is Known Exactly, *Probl. Peredachi Inf.*, 2015, vol. 51, no. 4, pp. 3–22 [Probl. Inf. Transm. (Engl. Transl.), 2015, vol. 51, no. 4, pp. 307–325].

12. Burnashev, M.V., Code Spectrum and the Reliability Function: Binary Symmetric Channel, *Probl. Peredachi Inf.*, 2006, vol. 42, no. 4, pp. 3–22 [Probl. Inf. Transm. (Engl. Transl.), 2006, vol. 42, no. 4, pp. 263–281].

13. Burnashev, M.V., Sharpening of an Upper Bound for the Reliability Function of a Binary Symmetric Channel, *Probl. Peredachi Inf.*, 2005, vol. 41, no. 4, pp. 3–22 [Probl. Inf. Transm. (Engl. Transl.), 2005, vol. 41, no. 4, pp. 301–318].

14. McEliece, R.J., Rodemich, E.R., Rumsey, H., Jr., and Welch, L.R., New Upper Bounds on the Rate of a Code via the Delsarte–MacWilliams Inequalities, *IEEE Trans. Inform. Theory*, 1977, vol. 23, no. 2, pp. 157–166.

15. Burnashev, M.V., On Lower Bounds on the Spectrum of a Binary Code, *Probl. Peredachi Inf.*, 2019, vol. 55, no. 4, pp. 76–85 [Probl. Inf. Transm. (Engl. Transl.), 2019, vol. 55, no. 4, pp. 366–375].

16. Litsyn, S., New Bounds on Error Exponents, *IEEE Trans. Inform. Theory*, 1999, vol. 45, no. 2, pp. 385–398.