On some estimation problems with information constraints *

Burnashev M. V.†, Han T. S.‡, Amari S.§

Abstract

The paper is the second part of author's article [1]. Parameter estimation problem is considered where some part of data can not be directly observed. Our helper observes those data and can send us some limited amount of information about them. What kind of that information allows us to get minimal mean–square error in parameter estimate ? In particular, what is the minimal information sufficient to get the same mean–square error as if we could observe directly all data ? Some upper bounds for that minimal amount of information and some related results are obtained.

Key words and phrases: parameter estimate, mean–square error, Fisher information, rate of transmission, critical rate.

The paper is the second part of author's article [1]. We also use notations from [1].

For the BSC model (Example 2 in [1]) we consider parameter p estimation problem.

Below we denote $\log x = \log_2 x$, $\exp_2 x = 2^x$. For any set A via $|A|$ cardinality of set A is denoted. For any function $f(x)$, $x \in A$, via $|f|$ cardinality of set $f(A)$ is denoted.

---

*The research described in this publication was made possible in part by Grant N 98-01-04108 from the Russian Fund for Fundamental Reseacrh.

†Institute for Problems of Information Transmission, Russian Academy of Sciences, Bolshoi Karetni 19, 101447, Moscow, Russia.

‡Graduate School of Information Systems, University of Electro–Communications, Chfugaoka 1 - 5 - 1, Chofu, Tokyo 182, Japan.

§RIKEN Brain Science Institute, Wako-shi, Hirosawa 2-1, Saitama 351-0198, Japan.

I. Parameter estimation

1. Statement of the problem and some auxilieri results

It will be more convinient for us to consider first a more general than for BSC parameter estimation problem. For the sake of variety in contrast to [1] we will change places of statistician and helper (i.e. statistician observes channel input and helper observes channel output). Of course, it does not change problem essence, but makes more clear its relation with other problems.

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets, representing input and output channel alphabets, respectively. There is given transition probability $p(y|x,\theta)\,; x \in \mathcal{X}, y \in \mathcal{Y}$, i.e. probability to get y at the output provided that x was at the input. It is supposed that it depends on some unknown parameter $\theta \in [a,b] \subset \mathbf{R}^1$ to be estimated. There is given also some prior distribution $\pi(x)$ on input alphabet $\mathcal{X}^n$ and input block x^n is vector of n independent random variables with that distribution $\pi(x)$. We designate via $\pi(x^n)$ its distribution

Let transmission rate $R > 0$ is given and we are allowed to partition output space $\mathcal{Y}^n$ on $N \leq 2^{Rn}$ parts $\{\mathcal{Y}_1, \ldots, \mathcal{Y}_N\}$. Now statistician observes only channel input x^n and helper observes only channel output y^n. After that helper informs statistician to which part $\mathcal{Y}_i$ belongs observation y^n. In other words, we choose on the set $\mathcal{Y}^n$ arbitrary function $f(y^n)$ that takes no more than N different values (i.e. $|\log f| \leq Rn$). Based on his own observation x^n and additional information $f(y^n)$ statistician constructs some estimate $\hat{\theta}_n = \hat{\theta}_n(x^n, f(y^n))$ for unknown parameter θ. We are interested in asymptotics $(n \to \infty)$ of mean–square error of estimate $\hat{\theta}_n$. More exactly, we are interested in behavior of function

$$e(\theta, \hat{\theta}_n, R, f) = \limsup_{n\to\infty} \left\{ n\, E_\theta \left[\hat{\theta}_n(x^n, f(y^n)) - \theta\right]^2 \right\} = \tag{1}$$

$$= \limsup_{n\to\infty} \left\{ n \sum_{x^n \in \mathcal{X}^n} \pi(x^n) E_\theta \left[(\hat{\theta}_n(x^n, f(y^n)) - \theta)^2 | x^n\right] \right\}, \quad R > 0\,,$$

for good functions $f(y^n)$ (partitions $\{\mathcal{Y}_i\}$) and good estimates $\hat{\theta}_n$.

For that purpose it is natural first to understand behavior of that function when transmission rate R is large enough such that helper can transfer to statistician exact value of block y^n (e.g. $R \geq \log|\mathcal{Y}|$). Then we come to standard statistical parameter estimation problem on observations x^n, y^n. We consider the regular case of that problem when transition probability $p(y|x,\theta)$ is continiously differentiable (on θ) function for any $x \in \mathcal{X}, y \in \mathcal{Y}$, and Fisher information for it is defined

$$I(\theta) = \sum_{x \in \mathcal{X}} \pi(x) \sum_{y \in \mathcal{Y}} \frac{p'^2(y|x,\theta)}{p(y|x,\theta)} > 0\,. \tag{2}$$

Since $\mathcal{X}, \mathcal{Y}$ are finite sets function $I(\theta)$ is uniformly continious for $\theta \in [a,b]$.

In standard estimation theory there is always some assumption that ensures existence of some uniformly consistent estimate θ [4, 5]. Then the problem reduces to investigation of properties of the best estimates $\hat{\theta}_n$ in a small vicinity of true value θ. In our case it is sufficient to assume that the family of probability measures $\{p(y|x,\theta),\, \theta \in [a,b]\}$ is "uniformly separable" on θ, i.e.

$$\inf_{|\theta-\theta_1|\ge\varepsilon} \sum_{x \in \mathcal{X}} \pi(x) \sum_{y \in \mathcal{Y}} |p(y|x,\theta) - p(y|x,\theta_1)| > 0 \quad \text{for any} \quad \varepsilon > 0\,. \tag{3}$$

Let $\hat{\theta}_n(x^n, y^n)$ be maximum likelyhood estimate for parameter θ. From standard estimation theory it follows that uniformly on $\theta \in [a,b]$ the following relation holds

$$\lim_{n\to\infty} \left\{ n\, E_\theta \left[\hat{\theta}_n(x^n, f(y^n)) - \theta \right]^2 \right\} \le \frac{1}{I(\theta)}\,, \tag{4}$$

(inequality in (4) arises at the endpoints of the segment $[a,b]$).

On the other hand it is known that for any $\delta > 0$ and $c \in [a,b]$ the following inequality holds ($c_\delta = (c-\delta, c+\delta) \cap [a,b]$)

$$\lim_{n\to\infty} \inf_{\tilde{\theta}_n} \sup_{\theta \in c_\delta} \left\{ n\, E_\theta \left[\hat{\theta}_n(x^n, f(y^n)) - \theta \right]^2 - \frac{1}{I(\theta)} \right\} \ge 0\,. \tag{5}$$

Therefore from (4)–(5) it follows

$$\lim_{n\to\infty} \inf_{\tilde{\theta}_n} \sup_{\theta \in [a,b]} \left\{ n\, E_\theta \left[\hat{\theta}_n(x^n, f(y^n)) - \theta \right]^2 - \frac{1}{I(\theta)} \right\} = 0\,. \tag{6}$$

We are interested in question what is the minimal rate $R_{\rm crit} = R_{\rm crit}([a,b])$ such that natural analog of relation (6) holds ? In other words, for what rate R there exists function $f(y^n)$ with $|\log f| \le Rn$ such that the following relation holds

$$\lim_{n\to\infty} \inf_{\tilde{\theta}_n} \sup_{\theta\in[a,b]} \left\{ n\, E_\theta \left[\hat{\theta}_n(x^n, f(y^n)) - \theta\right]^2 - \frac{1}{I(\theta)} \right\} = 0\,. \qquad (7)$$

We will show first that it is sufficient to establish relation (7) in some local $\delta(n)$–vicinity of each point $\theta \in [a,b]$.

Proposition 1.. *For validity of relation (7) it is sufficient that for any $c \in [a,b]$ and $\delta(n) = (\ln n)/\sqrt{n}$ the following relation holds*

$$\lim_{n\to\infty} \inf_{\tilde{\theta}_n} \sup_{\theta\in c_{\delta(n)}} \left\{ n\, E_\theta \left[\hat{\theta}_n(x^n, f(y^n)) - \theta\right]^2 - \frac{1}{I(\theta)} \right\} = 0\,. \qquad (8)$$

Proof. Suppose that statistician knows exact value of block y^m for some $m < n$. Then due to assumptions (2)–(3) there exist positive constants K_1, K_2 (independent on n, θ) such that for maximum likelyhood estimate $\hat{\theta}_m = \hat{\theta}_m(x^m, y^m)$ the inequality holds ([4], Theorem 5.3)

$$\sup_{\theta\in[a,b]} P_\theta \left\{ |\hat{\theta}_m - \theta| > z \right\} \le K_1 e^{-K_2 m z^2}\,, \quad z > 0\,. \qquad (9)$$

Therefore for sequence $\delta(m) = (\ln m)/\sqrt{m}$, starting with sufficiently large m, we have

$$\sup_{\theta\in[a,b]} E_\theta \left[\left(\hat{\theta}_m - \theta\right)^2 ; I_{|\hat{\theta}_m - \theta| > \delta(m)/4} \right] \le \frac{1}{m^2}\,. \qquad (10)$$

Let now helper transfers to statistician exact values of the first $m = \sqrt{n}$ values $(y_1, \ldots, y_m)$. Based on them statistician can construct some interval of length $\delta(n)/2$ where with large probability parameter θ is located (due to (10) we may neglect remaining small probability).

Now on the segment $[a,b]$ we construct the following system of segments of length $\delta(n)$. As the first segment we choose $[a, a+\delta(n)]$. Then we shift it sequentially to the right on $\delta(n)/2$ until its right end exceeds point b. In a result, total number of such segments of length $\delta(n)$ in such system does not exceed $2(b-a)/\delta(n)+1$. On the other hand, any segment of length $\delta(n)$ on $[a,b]$ completely belongs to some segment of that system.

Now for each segment of that system helper can use its own function f and transfer its value to statistician. Since total number of segments of that system is less than $\sqrt{n}$ (i.e. nonexponential) it does not change the transmission rate R. Statistician based on the first $\sqrt{n}$ observations can correctly determine the system segment (of length $\delta(n)$) where parameter to be estimated is located. After that he can use corresponding to that segment function f. Therefore validity of Proposition 1 follows. $\square$.

Remark. Although Proposition 1 is a rather simple and understandable result, it was not known in time to authors of paper [2] and that created certain troubles in [2].

Corollary 1. *The formula is valid*

$$R_{\text{crit}}([a,b]) = \sup_{\theta\in[a,b]} R_{\text{crit}}(\theta_{\delta(n)})\,. \tag{11}$$

Due to Proposition 1 and Corollary 1 the problem reduces to investigation of local (in a small vicinity of true value of θ) behavior of good estimates $\tilde{\theta}_n$. It is clear that that behavior will be expressed via correspondingly modified Fisher information. Let us introduce some necessary for that purpose notions.

For arbitrary partition $\{\mathcal{Y}_i; i=1,\ldots,N=2^{Rn}\}$ (defining function $f_n(y^n)$ with $|\log f_n|\le Rn$) define conditional and unconditional Fisher informations

$$I(\theta;f_n|x^n) = I(\theta;\{\mathcal{Y}_i\}|x^n) = \frac{1}{n}\sum_{i=1}^{N}\frac{p'^2(\mathcal{Y}_i|x^n,\theta)}{p(\mathcal{Y}_i|x^n,\theta)} \tag{12}$$

and

$$I(\theta;f_n) = I(\theta;\{\mathcal{Y}_i\}) = \sum_{x^n\in\mathcal{X}^n}\pi(x^n)\,I(\theta;\{\mathcal{Y}_i\}|x^n) = \tag{13}$$

$$= \frac{1}{n}\sum_{x^n\in\mathcal{X}^n}\pi(x^n)\sum_{i=1}^{N}\frac{p'^2(\mathcal{Y}_i|x^n,\theta)}{p(\mathcal{Y}_i|x^n,\theta)}\,.$$

Introduce also maximal achievable Fisher information for given rate R

$$I(\theta;R) = \lim_{n\to\infty}\sup_{|\log f_n|\le Rn} I(\theta;f_n)\,,\qquad R>0\,. \tag{14}$$

Since in relation (8) we deal only with consistent estimates in "regular experiment" [4, 5] then for any $R>0$ the following formula holds [2]

$$\lim_{n\to\infty}\inf_{\tilde{\theta}_n}\sup_{\theta\in c_{\delta(n)}}\left\{n\,E_\theta\left[\hat{\theta}_n(x^n,f(y^n))-\theta\right]^2-\frac{1}{I(\theta;R)}\right\} = 0\,. \tag{15}$$

It is clear also that for $I(\theta;R)$ and $I(\theta)$ the inequality holds

$$I(\theta;R) \le I(\theta)\,, \quad R > 0\,. \tag{16}$$

In order to prove it we notice that for any $\mathcal{Y}_i \subseteq \mathcal{Y}^n$ and $x^n \in \mathcal{X}^n$ due to Cauchy–Buniakovsky–Schwarz inequality we have

$$p'^2(\mathcal{Y}_i|x^n,\theta) = \left[\sum_{y^n\in\mathcal{Y}_i}\sqrt{p(y^n|x^n,\theta)}\frac{p'(y^n|x^n,\theta)}{\sqrt{p(y^n|x^n,\theta)}}\right]^2 \le$$

$$\le p(\mathcal{Y}_i|x^n,\theta)\sum_{y^n\in\mathcal{Y}_i}\frac{p'^2(y^n|x^n,\theta)}{p(y^n|x^n,\theta)}$$

Therefore for any partition $\{\mathcal{Y}_i\}$ and any $x^n \in \mathcal{X}^n$ we get

$$I(\theta;\{\mathcal{Y}_i\}|x^n) = \frac{1}{n}\sum_i\frac{p'^2(\mathcal{Y}_i|x^n,\theta)}{p(\mathcal{Y}_i|x^n,\theta)} \le \frac{1}{n}\sum_i\sum_{y^n\in\mathcal{Y}_i}\frac{p'^2(y^n|x^n,\theta)}{p(y^n|x^n,\theta)}\,,$$

from where together with (12)–(14) inequality (16) follows.

Therefore we come to the pure analytical problem: for which R there is equality in the inequality (16) ? Clearly, there exists certain critical rate $R_{\rm crit}(\theta) \ge 0$ such that equality in the inequality (16) holds for $R > R_{\rm crit}(\theta)$ and does not hold for $R < R_{\rm crit}(\theta)$. Define formally $R_{\rm crit}(\theta) \ge 0$ as

$$R_{\rm crit}(\theta) = \inf\left\{R : I(\theta;R) = I(\theta)\right\}, \quad \theta \in [a,b]\,. \tag{17}$$

2. Upper bound for $R_{crit}(\theta)$ (Fisher information for randomly chosen points)

In order to get upper bound for $R_{crit}(\theta)$ we estimate from below the value $I(\theta, R)$ and will get therefore the sufficient condition for equality in relation (16).

In turn in oder to estimate from below the value $I(\theta, R)$ we use method of "random choice" of sets $\{\mathcal{Y}_i\}$ which has been used already in the first part of paper [1].

Further we consider binary symmetrical channel (BSC) from Example 2 [1] with unknown transition probability $0 < \theta \le 1/2$ and $\mathcal{X} = \mathcal{Y} = E = \{0; 1\}$. Probability to get at the output block $y^n = (y_1, \ldots, y_n)$ provided that at the input was block $x^n = (x_1, \ldots, x_n)$ is given by formula

$$P(y^n|x^n) = (1-\theta)^{n-d(x^n,y^n)}\theta^{d(x^n,y^n)},$$

where $d(x^n, y^n)$ is Hamming distance between blocks x^n and y^n. We assume that input block x^n has equiprobable prior distribution on E^n.

Let us choose N natural integers $Y_1, \ldots, Y_N$ with $Y_1 + \ldots + Y_N = 2^n$. Each Y_i is cardinality of corresponding partition part $\mathcal{Y}_i$. We consider now all possible partitions of set $\mathcal{Y}^n$ on N sets $(\mathcal{Y}_1, \ldots, \mathcal{Y}_N)$ with $|\mathcal{Y}_i| = Y_i$; $i = 1, \ldots, N$. In other words, we choose first randomly Y_1 elements from set $\mathcal{Y}^n$ (it will be set $\mathcal{Y}_1$). After that from the remaining part of set $\mathcal{Y}^n$ we choose next Y_2 elements (it will be set $\mathcal{Y}_2$) and so on.

Due to the symmetry Fisher information $I(\theta, \{\mathcal{Y}_i\}|x^n)$, averaged over all partitions $\{\mathcal{Y}_i\}$, does not depend on input block x^n. Therefore we may assume that x^n zero–block $\mathbf{0} \in E^n$. In a result for the averaged value $EI(\theta, \{\mathcal{Y}_i\}|x^n)$ we have from (13)

$$E\,I(\theta, \{\mathcal{Y}_i\}|x^n) = \frac{1}{n}\sum_{i=1}^{N} E\,\frac{p'^2(\mathcal{Y}_i|\mathbf{0},\theta)}{p(\mathcal{Y}_i|\mathbf{0},\theta)} = \sum_{i=1}^{N} i(\theta, Y_i)\,; \quad \sum_{i=1}^{N} Y_i = 2^n\,, \tag{18}$$

where

$$i(\theta, A) = \frac{1}{n}\,E\,\frac{p'^2(\mathcal{A}|\mathbf{0},\theta)}{p(\mathcal{A}|\mathbf{0},\theta)}\,, \tag{19}$$

and last average is taken over all possible equiprobable sets $\mathcal{A}$ of cardinality A.

We investigate now the value $i(\theta; A)$ from (19) and show that it is almost linear function on A if A is not too large. For that purpose we consider first the case $0 < \theta < 1/2$. Choose some $\theta < r < 1/2$ and introduce the following classes $\{\mathcal{B}_i\}$ $\{\mathcal{C}_i\}$ of sets $\mathcal{A} = \{\mathbf{y}_1, \ldots, \mathbf{y}_A\}$, where $\mathbf{y}_i \in E^n$; $i = 1, \ldots, A$.

$$\mathcal{B}_i = \{\mathcal{A} : w(\mathbf{y}_i) \le rn\}\ ; \quad i = 1, \ldots, A\,;$$

$$\mathcal{C}_i = \left\{\mathcal{A} : \min_{j \ne i} w(\mathbf{y}_j) > rn\right\}\ ; \quad i = 1, \ldots, A\,,$$

where $w(\mathbf{y})$ – weight (i.e. the number of ones) of vector $\mathbf{y}$.

Since sets $\{\mathcal{B}_i \cap \mathcal{C}_i;\ i = 1, \ldots, A\}$ does not intersect between each other then due to the symmetry we have ($E[\xi; \mathcal{A}] = \int_{\mathcal{A}} x dF_\xi(x)$)

$$E\,\frac{p'^2(\mathcal{A}|\mathbf{0},\theta)}{p(\mathcal{A}|\mathbf{0},\theta)} \ge AE\left[\frac{p'^2(\mathcal{A}|\mathbf{0},\theta)}{p(\mathcal{A}|\mathbf{0},\theta)}\,;\ \mathcal{B}_1 \cap \mathcal{C}_1\right]. \tag{20}$$

Then we have ($M = 2^n$, V_{rn} – cardinality of ball of radius rn in E^n)

$$P(\mathcal{B}_1) = V_{rn} 2^{-n}\,;$$

$$P(\mathcal{C}_1|\mathcal{B}_1) = \binom{M - V_{rn}}{A-1} \Big/ \binom{M-1}{A-1} = \prod_{i=1}^{V_{rn}-1}\left[1 - \frac{(A-1)}{(M - V_{rn} - i)}\right] \ge \tag{21}$$

$$\ge \left[1 - \frac{(A-1)}{(M - 2V_{rn} + 1)}\right]^{V_{rn}-1} \ge 1 - \frac{AV_{rn}}{(2^n - 2V_{rn})}\,.$$

Notice that

$$\frac{p'(y^n|x^n,\theta)}{p(y^n|x^n,\theta)} = \frac{(d(x^n,y^n) - \theta n)}{\theta(1-\theta)}\,; \quad \frac{p'(\mathbf{y}|\mathbf{0},\theta)}{p(\mathbf{y}|\mathbf{0},\theta)} = \frac{(w(\mathbf{y}) - \theta n)}{\theta(1-\theta)}\,;$$

$$i(\theta, 1) = \frac{1}{n}\,E\,\frac{p'^2(\mathbf{y}|\mathbf{0},\theta)}{p(\mathbf{y}|\mathbf{0},\theta)} = \frac{1}{n}\,E\,\frac{(w(\mathbf{y}) - \theta n)^2}{\theta^2(1-\theta)^2} p(\mathbf{y}|\mathbf{0},\theta) = \frac{2^{-n}}{\theta(1-\theta)}\,.$$

Therefore for $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_A\} \in \mathcal{C}_1$ the estimate holds

$$0 \le \sum_{i=2}^{A} p'(\mathbf{y}_i|\mathbf{0},\theta) \le \frac{n}{\theta} \sum_{i=2}^{A} p(\mathbf{y}_i|\mathbf{0},\theta)\,. \tag{22}$$

In addition using Chebichev inequality we have

$$P\left\{\sum_{i=2}^{A} p(\mathbf{y}_i|\mathbf{0},\theta) > z\right\} \le (A-1)2^{-n}z^{-1}\,; \quad z>0\,. \tag{23}$$

Introduce the set of random events $\mathcal{D}_z \subseteq \mathcal{C}_1\,,\ z>0$:

$$\mathcal{D}_z = \left\{\{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_A\} \in \mathcal{C}_1 : \sum_{i=2}^{A} p(\mathbf{y}_i|\mathbf{0},\theta) \le z\right\}.$$

Now for $\mathcal{A} = \{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_A\} \in \mathcal{D}_z$ using (22) we have

$$\frac{p'^2(\mathcal{A}|\mathbf{0},\theta)}{p(\mathcal{A}|\mathbf{0},\theta)} \ge \frac{p'^2(\mathbf{y}_1|\mathbf{0},\theta) - 2nz\theta^{-1}|p'(\mathbf{y}_1|\mathbf{0},\theta)|}{p(\mathbf{y}_1|\mathbf{0},\theta)+z} \ge$$

$$\ge \frac{p'^2(\mathbf{y}_1|\mathbf{0},\theta)}{p(\mathbf{y}_1|\mathbf{0},\theta)} - \frac{zp'^2(\mathbf{y}_1|\mathbf{0},\theta)}{p^2(\mathbf{y}_1|\mathbf{0},\theta)} - \frac{2nz|p'(\mathbf{y}_1|\mathbf{0},\theta)|}{\theta p(\mathbf{y}_1|\mathbf{0},\theta)}\,.$$

Therefore for right side of (20) we get

$$E\left[\frac{p'^2(\mathcal{A}|\mathbf{0},\theta)}{p(\mathcal{A}|\mathbf{0},\theta)}\,;\ \mathcal{B}_1\cap\mathcal{C}_1\right] \ge E\left[\frac{p'^2(\mathcal{A}|\mathbf{0},\theta)}{p(\mathcal{A}|\mathbf{0},\theta)}\,;\ \mathcal{B}_1\cap\mathcal{D}_z\right] \ge$$

$$\ge E\left[\frac{p'^2(\mathbf{y}_1|\mathbf{0},\theta)}{p(\mathbf{y}_1|\mathbf{0},\theta)} - \frac{zp'^2(\mathbf{y}_1|\mathbf{0},\theta)}{p^2(\mathbf{y}_1|\mathbf{0},\theta)} - \frac{2nz|p'(\mathbf{y}_1|\mathbf{0},\theta)|}{\theta p(\mathbf{y}_1|\mathbf{0},\theta)};\mathcal{B}_1\cap\mathcal{D}_z\right] = \tag{24}$$

$$= E\left[\frac{p'^2(\mathbf{y}_1|\mathbf{0},\theta)}{p(\mathbf{y}_1|\mathbf{0},\theta)} - \frac{zp'^2(\mathbf{y}_1|\mathbf{0},\theta)}{p^2(\mathbf{y}_1|\mathbf{0},\theta)} - \frac{2nz|p'(\mathbf{y}_1|\mathbf{0},\theta)|}{\theta p(\mathbf{y}_1|\mathbf{0},\theta)};\mathcal{B}_1\right] P(\mathcal{D}_z|\mathcal{B}_1)\,.$$

Now we estimate three terms in right side of (24) :

$$E\left[\frac{p'^2(\mathbf{y}_1|\mathbf{0},\theta)}{p(\mathbf{y}_1|\mathbf{0},\theta)};\mathcal{B}_1\right] \ge \frac{n\,2^{-n}}{\theta(1-\theta)} - \frac{n^2}{\theta^2}E\left[p(\mathbf{y}_1|\mathbf{0},\theta);\mathcal{B}_1^c\right] =$$

$$= \frac{n\,2^{-n}}{\theta(1-\theta)} - \frac{n^2\,2^{-n}}{\theta^2}P_\theta(\mathcal{B}_1^c)\,;$$

$$E\left[\frac{p'^2(\mathbf{y}_1|\mathbf{0},\theta)}{p^2(\mathbf{y}_1|\mathbf{0},\theta)};\mathcal{B}_1\right] \le \frac{n^2V_{rn}2^{-n}}{\theta^2}\,;\ E\left[\frac{|p'(\mathbf{y}_1|\mathbf{0},\theta)|}{p(\mathbf{y}_1|\mathbf{0},\theta)};\mathcal{B}_1\right] \le \frac{nV_{rn}2^{-n}}{\theta}\,.$$

Now using (21) and (23) for $r \le 1/2$ and $A \ge 2$ we have

$$P(\mathcal{D}_z|\mathcal{B}_1) \ge P(\mathcal{C}_1|\mathcal{B}_1) - P\left\{\sum_{i=2}^{A} p(\mathbf{y}_i|\mathbf{0},\theta) > z\right\} \ge 1 - A2^{-n}\left(2V_{rn} + z^{-1}\right).$$

Therefore from (20) and (24) we get

$$E\,\frac{p'^2(\mathcal{A}|\mathbf{0},\theta)}{p(\mathcal{A}|\mathbf{0},\theta)} \ge \frac{A\,n\,2^{-n}}{\theta(1-\theta)}\left[1 - \frac{nP_\theta(\mathcal{B}_1^c) + 3nzV_{rn}}{\theta} - \frac{A(2V_{rn} + z^{-1})}{2^n}\right]. \quad (25)$$

It remains to estimate values $P_\theta(\mathcal{B}_1^c)$ and V_{rn} in right side of (25) and choose parameters r, z. First, using standard estimates for binomial coefficients [6, 7] and Taylor formula for function $h(r)$ we have

$$P_\theta(\mathcal{B}_1^c) = \sum_{i>rn}\binom{n}{i}\theta^i(1-\theta)^{n-i} \le \frac{\theta^{rn+1}(1-\theta)^{n(1-r)}2^{nh(r)}}{2(r-\theta)\sqrt{nr}} \le$$

$$\le \frac{\theta}{2(r-\theta)\sqrt{nr}}\exp\left\{-2n(r-\theta)^2\right\}; \quad \theta < r \le 1/2;$$

$$V_{rn} \le \frac{\sqrt{1-r}}{2(1-2r)\sqrt{rn}}\,2^{nh(r)}, \quad r < 1/2.$$

Puting now

$$r = \theta + \sqrt{(\ln n)/n}; \quad z = 1/(n^2\,V_{rn}),$$

we get

$$P_\theta(\mathcal{B}_1^c) \le n^{-2}\sqrt{\theta/(\ln n)}; \quad \theta + \sqrt{(\ln n)/n} < 1/2; \quad (26)$$

$$V_{rn} \le \frac{2^{nh(\theta)}}{2n^2\left(1 - 2\theta - 2\sqrt{(\ln n)/n}\right)\sqrt{n\theta}}\left(\frac{1-\theta}{\theta}\right)^{\sqrt{n\ln n}}.$$

In a result from (25–26) follows the following result having also is its own interest.

Proposition 2. *For value $i(\theta, A)$, $0 < \theta < 1/2$, estimates hold*

$$1 - \frac{1}{n\sqrt{\theta\ln n}} - \frac{3}{n\theta} - \frac{A\,2^{nh(\theta)-n}}{\left(1 - 2\theta - 2\sqrt{(\ln n)/n}\right)\sqrt{\theta n}}\left(\frac{1-\theta}{\theta}\right)^{\sqrt{n\ln n}} \le$$

$$\leq \frac{\theta(1-\theta)2^n}{A}\, i(\theta, A) \leq 1\,; \quad \theta + \sqrt{(\ln n)/n} < 1/2\,. \tag{27}$$

We come back now to relation (18). Let cardinality Y_i of partition part $\mathcal{Y}_i$ satisfyes condition

$$Y_i \leq 2^{n-nh(\theta)}\, (\theta/(1-\theta))^{\sqrt{n \ln n}}\,; \quad i = 1, \ldots, N\,. \tag{28}$$

For such partitions $\{\mathcal{Y}_i\}$ due to (27) uniformly on $\theta \in [a,b]$; $0 < a < b < 1/2$, we have

$$EI(\theta, \{\mathcal{Y}_i\}) = \frac{n}{\theta(1-\theta)}\,(1 + o(1))\,, \quad n \to \infty\,. \tag{29}$$

It is clear that for existence of such collection $\{Y_i,\, i = 1, \ldots, N\}$ satisfying condition (28) it is sufficient that total number N of parts satisfyes condition

$$N \geq 2^{nh(\theta)}\, ((1-\theta)/\theta)^{\sqrt{n \ln n}}\,. \tag{30}$$

Notice also that for the value $I(\theta; R)$ condition is fullfilled

$$I(\theta; R) = I(1-\theta; R)\,; \quad 0 < \theta < 1\,; \quad R > 0\,.$$

In a result, from (11) and (27)–(30) we get

Corollary 2. *The relation holds*

$$I(\theta; R) = I(\theta) \qquad R \geq h(\theta)\,; \quad 0 < \theta < 1\,; \tag{31}$$

and, therefore,

$$R_{\text{crit}}(\theta) \leq h(\theta); \quad R_{\text{crit}}([a,b]) \leq \max_{x \in [a,b]} h(x)\,. \tag{32}$$

2. Point $\theta = 1/2$

In the case of $\theta = 1/2$ calculations concerning Fisher information become much simpler that allows to find exactly the function $I(\theta = 1/2; R) \quad R > 0$.

First from definition (13) we have

$$I(\theta; \{\mathcal{Y}_i\}) = n^{-1} 2^{-n} \sum_{x^n \in E^n} \sum_{i=1}^{N} \frac{p'^2(\mathcal{Y}_i | x^n, \theta)}{p(\mathcal{Y}_i | x^n, \theta)} = \sum_{i=1}^{N} i(\theta, \mathcal{Y}_i)\,, \tag{33}$$

where it is designated

$$i(\theta, \mathcal{A}) = \frac{2^{-n}}{n} \sum_{x^n \in E^n} \frac{p'^2(\mathcal{A} | x^n, \theta)}{p(\mathcal{A} | x^n, \theta)} =$$

$$= \frac{2^{-n}}{n\theta^2(1-\theta)^2} \sum_{x^n \in E^n} \frac{\left[\sum_{y^n \in \mathcal{A}} (d(x^n, y^n) - n\theta) p(y^n | x^n, \theta)\right]^2}{\sum_{y^n \in \mathcal{A}} p(y^n | x^n, \theta)}\,. \tag{34}$$

For $\theta = 1/2$ value $i(\theta, \mathcal{A})$ from (34) takes the form ($A = |\mathcal{A}|$)

$$i(1/2, \mathcal{A}) = \frac{2^{4-2n}}{nA} \sum_{y_1^n \in \mathcal{A}} \sum_{y_2^n \in \mathcal{A}} \sum_{x^n \in E^n} \left(d(y_1^n, x^n) - \frac{n}{2}\right) \left(d(y_2^n, x^n) - \frac{n}{2}\right) =$$

$$= A\, 2^{2-n} \left\{ 1 - \frac{2}{nA^2} \sum_{y_1^n \in \mathcal{A}} \sum_{y_2^n \in \mathcal{A}} d(y_1^n, y_2^n) \right\}\,, \tag{35}$$

where the folowing calculations have been used ($d = d(y_1^n, y_2^n)$)

$$\sum_{x^n \in E^n} \left(d(y_1^n, x^n) - \frac{n}{2}\right) \left(d(y_2^n, x^n) - \frac{n}{2}\right) =$$

$$= \sum_{w=0}^{d} \sum_{v=0}^{n-d} \binom{d}{w} \binom{n-d}{v} \left(w + v - \frac{n}{2}\right) \left(d - w + v - \frac{n}{2}\right) =$$

$$= 2^{d-2} \sum_{v=0}^{n-d} \binom{n-d}{v} \left[4v^2 + 4v(d-n) + d(d-1) - n(2d-n)\right] =$$

$$= 2^{n-2} \left(n - 2d(y_1^n, y_2^n)\right)\,.$$

In order to find the maximum of right side of (35) for given cardinality A we need to find the minimum (over all sets $\mathcal{A}$) of

$$\frac{2}{nA^2}\sum_{y_1^n\in\mathcal{A}}\sum_{y_2^n\in\mathcal{A}} d(y_1^n,y_2^n)\,.$$

Moreover, we are interested in asymptotic ($n\to\infty$) value of that minimum when $A\sim 2^{rn}\,, 0<r<1/2$. It is known [8] that asymptotically that minimum is attained when set $\mathcal{A}$ is ball $B_{\alpha n}\subset E^n$ of some radius $\alpha n\,, 0<\alpha<1/2$. Since $A\sim 2^{rn}\sim |B_{\alpha n}|\sim 2^{nh(\alpha)}$ then

$$h(\alpha)=r=\frac{\log A}{n}\,. \tag{36}$$

In order to calculate sum in the right side of (35) we imagine that all blocks $y^n\in\mathcal{A}$ are written as $A\times n$ matrix (i.e. each matrix row is some block y^n). Let m_i; $i=1,\ldots,n$, be the number of ones in i–th column of that matrix. Then we have

$$\frac{2}{nA^2}\sum_{y_1^n\in\mathcal{A}}\sum_{y_2^n\in\mathcal{A}} d(y_1^n,y_2^n)=\frac{4}{nA^2}\sum_{i=1}^{n} m_i(A-m_i)\,.$$

For total number of ones we have

$$\sum_{i=1}^{n} m_i=\sum_{j=0}^{\alpha n} j\binom{n}{j}=\alpha n|B_{\alpha n}|\left(1+o(1)\right)\,.$$

Therefore

$$m_1=\ldots=m_n=\alpha|B_{\alpha n}|\left(1+o(1)\right)\,,\ n\to\infty\,.$$

and for ball $\mathcal{A}$ of radius αn we have

$$\frac{2}{nA^2}\sum_{y_1^n\in\mathcal{A}}\sum_{y_2^n\in\mathcal{A}} d(y_1^n,y_2^n)=4\alpha(1-\alpha)\left(1+o(1)\right)\,,\ n\to\infty\,. \tag{37}$$

In a result, from (35) we get

$$\max_{\mathcal{A}:|\mathcal{A}|=A} i(1/2,\mathcal{A})=A\,2^{2-n}\left(1-2\alpha\right)^2\left(1+o(1)\right)\,,\ n\to\infty\,, \tag{38}$$

where α is defined in (36).

In order using formula (38) to get necessary result for whole space E^n we will need the following refinement of Lemma 1 on "coverings" from [1].

Lemma 1. *Let $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_X\} \subset E^n$ be arbitrary set of cardinality $X \leq 2^{n-2}$ and $K =](n+1)2^n/X[$. Then there exist K "shifts" $\{\mathbf{y}_1, \ldots, \mathbf{y}_K\} \subset E^n$ such that:*
1) sets $\mathcal{X}_i = \mathcal{X} + \mathbf{y}_i; i = 1, \ldots, K$, cover space E^n and moreover each point $\mathbf{x} \in E^n$ is covered not more than $4n$ times;
2) if $X \geq 6n^2$ then there exist a system of disjoint sets $\mathcal{Y}_i \subseteq \mathcal{X}_i \,; i = 1, \ldots, K$ covering space E^n with $|\mathcal{Y}_i| \geq X/(8n) \,; i = 1, \ldots, K$.

Proof. Let K, m be arbitrary natural integers. In order to get the first lemma's statement we choose K shifts randomly and independently (with returns). Then we have

$$P_0 = \Pr \left\{ \begin{array}{c} \text{there exists noncovered point } \ \mathbf{x} \in E^n, \text{ or} \\ \text{there exists point } \ \mathbf{x} \in E^n \text{ covered more than } m \text{ times} \end{array} \right\} \leq$$

$$\leq 2^n \Pr \left\{ \text{point } \mathbf{0} \text{ is not covered} \right\} +$$

$$+ 2^n \Pr \left\{ \text{point } \mathbf{0} \text{ is covered more than } m \text{ times} \right\} = P_1 + P_2 \,.$$

For any $K > (n+1)2^n \ln 2/X$ we get

$$P_1 = 2^n \left(1 - X 2^{-n}\right)^K \leq \exp\left\{-XK2^{-n} + n \ln 2\right\} < 1/2 \,.$$

Using known estimates for binomial coefficients [7, 8], we have ($M = 2^n$)

$$\ln P_2 = \ln \left\{ M \sum_{i=m+1}^{K} \binom{K}{i} \left(\frac{X}{M}\right)^i \left(1 - \frac{X}{M}\right)^{K-i} \right\} \leq$$

$$\leq \ln M + K h_e \left(\frac{m+1}{K}\right) + K \ln \left(1 - \frac{X}{M}\right) - (m+1) \ln \frac{M-X}{X} \leq$$

$$\leq \ln M - \frac{KX}{M} + (m+1) \ln \frac{eKX}{(m+1)(M-X)} \,,$$

where inequality $h_e(x) \leq x \ln(e/x)$ was used. Now for K defined in lemma and $m = 4n$ we get $P_2 \leq 1/2$. In a result, for such K, m we have $P_0 < 1$ and therefore there exist a collection of shifts satisfying the first Lemma 1 statement.

In order to prove the second lemma's statement consider any point $x \in E^n$. It is covered by not more than m sets $\mathcal{X}_i$. We attribute that point to one of those sets by random equiprobable way. Repeat the same procedure for all points $x \in E^n$. In a result, we get a system of disjoint sets $\mathcal{Y}_i \subseteq \mathcal{X}_i\,; i = 1, \ldots, K$ covering space E^n. For that system for $2mN \leq X$ we have

$$Q_0 = \Pr\left\{\text{there exists set } \mathcal{Y}_i \text{ with } |\mathcal{Y}_i| \leq N\ \right\} \leq$$

$$\leq KP\left\{|\mathcal{Y}_1| \leq N\right\} \leq K \sum_{j=0}^{N} \binom{X}{j} m^{-j} \left(1 - \frac{1}{m}\right)^{X-j} \leq$$

$$\leq K \left(1 - \frac{1}{m}\right)^{X} (m-1)^{-N} \exp\left\{X h_e\left(\frac{N}{X}\right)\right\} <$$

$$< K \exp\left\{N \ln \frac{eX}{(m-1)N} - \frac{X}{m}\right\} \leq K e^{-X/(2m)} < 1\,,$$

if $X \geq 6n^2$. Therefore, there exists a collection of shifts satisfying the second Lemma 1 statement. $\Box$.

We come now back to relation (33) and let $\mathcal{A} = B_{\alpha n}(\mathbf{0})$ with $|\mathcal{A}| = A \simeq 2^{rn}$. Due to Lemma 1 there exist $N = (n+1)2^n/A \simeq n2^{n(1-r)}$ shifts $\{\mathbf{y}_1, \ldots, \mathbf{y}_N\} \subset E^n$ such that balls $B_{\alpha n}(\mathbf{y}_i) = \mathcal{A} + \mathbf{y}_i$; $i = 1, \ldots, N$, cover all space E^n. Due to the second Lemma 1 statement there exist disjoint sets $\mathcal{Y}_i \subset B_{\alpha n}(\mathbf{y}_i)$; $i = 1, \ldots, N$, that still cover all space E^n and moreover each set $\mathcal{Y}_i$ has cardinality $|\mathcal{Y}_i| \geq |B_{\alpha n}|/(8n)$.

It is left to show that relation (37) remains valid for such set $\{\mathcal{Y}_i\}$ as well. For that purpose we notice that in ball $B_{\alpha n}$ the number of ordered pairs (y_1^n, y_2^n) with $|d(y_1^n, y_2^n) - 2\alpha(1-\alpha)n| > \sqrt{n \ln n}$ does not exceed $|B_{\alpha n}|^2 n^{-3}$. Since $\mathcal{Y}_i \subset B_{\alpha n}$ and total number of ordered pairs (y_1^n, y_2^n) $\mathcal{Y}_i$ has order $|B_{\alpha n}|^2 n^{-2}$ then for "main part" of such pairs relation holds $|d(y_1^n, y_2^n) - 2\alpha(1-\alpha)n| \leq \sqrt{n \ln n}$ from where we get

$$\frac{2}{nY_i^2} \sum_{y_1^n \in \mathcal{Y}_i} \sum_{y_2^n \in \mathcal{Y}_i} d(y_1^n, y_2^n) = 4\alpha(1-\alpha)\,(1 + o(1))\ ;\ i = 1, \ldots, N,\ ; \qquad (39)$$

and

$$i(1/2, \mathcal{Y}_i) = Y_i\, 2^{2-n}\, (1 - 2\alpha)^2\, (1 + o(1))\ ;\ i = 1, \ldots, N;\ n \to \infty\,. \qquad (40)$$

Since $\sum_i Y_i = 2^n$, then from (33) and (40) follows

$$I(1/2; \{\mathcal{Y}_i\}) = 4\,(1-2\alpha)^2 \,. \tag{41}$$

In a result, from (33)–(41) we get

Proposition 3. *Function $I(1/2;R)$ is defined by relation*

$$I(1/2;R) = 4\,(1-2\alpha)^2 \;; \quad 0 < R < 1 \,, \tag{42}$$

where $h(\alpha) = 1 - R$. In particular, $R_{\mathrm{crit}}(1/2) = 0$.

For comparison purpose we will find also the value $i(1/2, A)$ which is value $i(1/2, \mathcal{A})$ averaged over all randomly chosen sets $\mathcal{A}$ of cardinality A. We may assume that set $\mathcal{A}$ has a form $\mathcal{A} = \{y_1^n = \mathbf{0}, y_2^n, \ldots, y_A^n\}$. Then we have

$$A^{-2} E \sum_{y_1^n \in \mathcal{A}} \sum_{y_2^n \in \mathcal{A}} d(y_1^n, y_2^n) = A^{-1} E \sum_{y \in \mathcal{A}} w(y) = \frac{n}{2}\left(1 - \frac{1}{A}\right)\left(1 + \frac{1}{2^n - 1}\right) \,.$$

Therefore from (35) we get

$$i(1/2, A) = E\, i(1/2, \mathcal{A}) = 2^{2-n}\left(1 - \frac{A-1}{2^n - 1}\right) \,.$$

In a result from (33) for $0 \le R < 1$ we have

$$EI(1/2; \{\mathcal{A}_i\}) = 2^{2-n}(N-1) = 2^{2-(1-R)n}\left(1 - 2^{-Rn}\right) \to 0 \,,\; n \to \infty \,. \tag{43}$$

It means that for the case $\theta = 1/2$ random choice of sets $\{\mathcal{A}_i\}$ is very bad (Fisher information obtained goes even to 0 when $n \to \infty$). It possible to make a conjecture that random choice of sets $\{\mathcal{A}_i\}$ gives asymptotically optimal results only when $R > R_{\mathrm{crit}}(\theta)$. In the case $R < R_{\mathrm{crit}}(\theta)$ the optimal set $\mathcal{A}_i$, apparently, is a set of "well separated" balls of certain raduis. It shows a certain difference of situation in this problem with respect to traditional situation in information theory (where "averaged code" is always rather good).

REFERENCES

1. *Burnashev M. V., Amari S., Han T. S.* On some testing of hypotheses problems with information constraints.

2. *Ahlswede R., Burnashev M. V.* On minimax estimation in the presence of side information about remote data. – The Annals of Statistics, 1990, v. 18, No. 1, p. 141–171.

3. *Han T. S., Amari S.* Parameter estimation with multiterminal data compression. – IEEE Trans. on Inform. Theory, 1965, v. IT–41, No. 6, p. 1802–1833.

4. *Ibragimov I. A., Has'minskii R. Z.* Statistical Estimation. Asymptotic Theory. Springer–Verlag, New York–Berlin–Heidelberg, 1981.

5. *Le Cam L.* Asymptotic methods in statistical decision theory. New York, Springer–Verlag, 1986.

6. *Fano R. M.* Transmission of Information. A statistical theory of communication. MIT&Wiley, New York–London, 1961.

7. *Gallager R. G.* Information Theory and Reliable Communication. Wiley, New York–London–Sydney–Toronto, 1968.

8. *Ahlswede R., Althöfer I.* The asymptotic behavior of diameters in the average. – Journal of Combinatorial Theory, Series B, 1994, v. 61, No. 2, p. 167–177.