# 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 $\theta$. 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