Sequential recovery of analytic periodic edges in the binary image models

G. Golubev[1] B. Levit[2]

[1]CMI, Université de Provence
39, rue F. Joliot-Curie
13453 Marseille, Cedex 13
France
e-mail: golubev@gyptis.univ-mrs.fr

[2]Department of Mathematics and Statistics
Queen's University
Kingston ON, K7L 3N6
Canada
e-mail: blevit@mast.queensu.ca

Mathematics Subject Classification (2000): 62G05, 62G20.
Key words: binary image models, analytic functions, sequential designs, minimax risk, sup-norm.

In the binary image model, we consider the problem of the edge restoration based on noisy data. Assuming that the edge function is analytic and the design points are chosen sequentially, we derive lower and upper bounds for the minimax risk of recovery in the sup-norm.

1 Introduction

The model. This paper deals with the problem of edge restoration based on noisy data in a simple binary image model. We assume that an image I is described by a function $I(x),\ x = (x_1, x_2)$ on the unit square $S = [0,1] \times [0,1]$, with two possible values 0 and 1, which are defined as

$$I(x) = \mathbf{1}\{x_2 \le f(x_1)\}.$$

Here $f(x),\ x \in [0,1]$ is the unknown edge function. Since the image is completely characterized by its edge function f, we will use the notation $I_f(x)$.

Suppose that at any n chosen design points $X_k = (X_{k1}, X_{k2}) \in S$ we can observe an image $I_f(x)$ corrupted by the binary noise. In other words, our observations are random variables

$$Y(X_k) = I_f(X_k) \oplus \xi_k, \quad k = 1, \dots, n, \tag{1}$$

where ξ_k are i.i.d. Bernoulli random variables, independent of the design points X_k, with $\mathbf{P}\{\xi_k = 0\} = p$. Throughout the paper we will assume that $0.5 < p \le 1$.

Our goal is to estimate the edge function f, with a reasonably high efficiency, based on the observations $Y(X_k),\ k = 1, \dots, n$. To achieve this, we will explore two important ingredients. First, the design points $X_k,\ k = 1, \dots, n$ will be chosen sequentially. This means that each new observation point X_k can be chosen using all the information which can be derived from the previous design points $X_1, \dots, X_{k-1}$ and the observations $Y(X_1), \dots, Y(X_{k-1})$. Formally,

$$X_k = \mathbf{X}_k(Y(X_1), \dots, Y(X_{k-1}), X_1, \dots, X_{k-1}), \tag{2}$$

where $\mathbf{X}_k(\cdot)$ can be arbitrary measurable function $\mathbb{R}^{2k-2} \to [0,1]^2$. Such approach has been proposed earlier in Korostelev (1999). Note that the equations (1), (2) allow to interpret our observations as memoryless binary symmetric channel with feedback (cf. Cover and Thomas (1991), Ch. 8.12). With such interpretation, an *"encoder"* uses the unknown function $f(x)$ and the observed data $Y(X_i),\ i = 1, ..., k-1$ to specify next transmitted signal $I_f(X_k)$ distorted by the binary noise in the transmission channel.

Another novelty feature is concerned with the edge functions f. We will assume that f is a periodic analytic function. More precisely, let

$$\varphi_k(x) = \exp(2\pi i k x), \quad k = 0, \pm 1, ...$$

be the standard orthonormal trigonometric basis in $\mathbf{L}_2[0,1]$ equipped with the inner product

$$\langle f, g \rangle = \int_0^1 f(u) g^*(u)\, du,$$

where z^* stands for the complex conjugate $z \in \mathbf{C}$. Consider the following functional class

$$\mathcal{F}(\gamma, L) = \left\{ f(x) = \sum_{k=-\infty}^{\infty} f_k \varphi_k(x) : f_k = f_{-k}^*,\ |f_k| \le L \exp(-2\gamma |k|) \right\}. \tag{3}$$

We will assume that the unknown edge function f is an element of this compact set in $\mathbf{L}_2(0,1)$, with known positive parameters L and γ.

We define the minimax risk, with respect to the functional class $\mathcal{F}(\gamma, L)$, in the usual way

$$r_n(\mathcal{F}(\gamma, L)) = \inf_{\hat{f}, \mathbf{X}_1, \dots, \mathbf{X}_n} \sup_{f \in \mathcal{F}(\gamma, L)} \mathbf{E}_f \|\hat{f} - f\|. \tag{4}$$

Here and further $\mathbf{E}_f$ is the expectation with respect to the measure induced by the observations (1), the norm $\|\cdot\|$ is defined by

$$\|g\| = \sup_{x \in [0,1]} |g(x)|$$

and *inf* is taken over all possible sequential designs as well as arbitrary estimates $\hat{f}$.

2 An upper bound

The noise-free case: $p = 1$. An essential feature of the above model is the use of sequential designs in choosing the observation points X_k. In the problem of edge recovery such sequential designs can dramatically improve the accuracy of algorithms. To explain why sequential algorithms have advantage over non-sequential procedures in such models, let us start with a special case in which the observations are not corrupted by noise. To simplify the situation even further, let us assume first that $f(x) = f = \text{const}$. Then we arrive at the following model

$$Y(X_k) = \mathbf{1}\{X_{k2} \le f\}, \quad k = 1, \dots, n \tag{5}$$

where $f \in [0, 1]$ is unknown. In this case it is not difficult to construct an algorithm which converges to f exponentially fast: one can use, for instance, the famous Brent algorithm (Brent (1973)) converging at the rate 2^{-n}. It fact, it is well known that there is no algorithm which converges faster. Now compare this rate of convergence with the rate n^{-1} obtainable when one uses a fixed equidistant design.

Suppose next that we want to recover an analytic boundary $f \in \mathcal{F}(\gamma, L)$ based on the noise-free observations

$$Y(X_k) = I_f(X_k), \quad k = 1, \dots, n. \tag{6}$$

How should one change then the scenario of edge recovery and what would be the effect of our underlying assumption $f \in \mathcal{F}(\gamma, L)$? To answer these questions, let us recall some results concerning interpolation of analytic functions. We start with what can be defined as a "noise-free interpolation". Suppose $x_1, \dots, x_m$ is a uniform grid on $[0, 1]$,

$$x_i = \frac{i-1}{m}, \quad i = 1, \dots, m,$$

and the function $f \in \mathcal{F}(\gamma, L)$ is to be on recovered on the hole interval $[0, 1]$ given the data

$$y_i = f(x_i), \quad i = 1, \dots, m.$$

Proposition 1 *There exists an interpolation procedure* $\bar{f}(x) = \bar{f}(x, y_1, \dots, y_m)$ *such that*

$$\sup_{f \in \mathcal{F}(\gamma, L)} \|\bar{f} - f\| \le c(\gamma) L \exp(-\gamma m),$$

where $c(\gamma)$ *depends only on* γ*.*

This result is well-known; we provide its proof in the Appendix only for the reader's convenience. One can interpolate data using for instance the following formula going back to Cauchy and Gauss:

$$\bar{f}(x) = \sum_{i=1}^{m} y_i \, \frac{\sin(\pi(x - x_i)m)}{m \sin(\pi(x - x_i))} \,. \tag{7}$$

Using Proposition 1, one can solve the following "noisy interpolation" problem. Suppose one wants to recover $f \in \mathcal{F}(\gamma, L)$ on the interval $[0, 1]$ based on the data

$$y_i = f(x_i) + \zeta_i, \quad i = 1, \dots, m. \tag{8}$$

Proposition 2 *There exist an interpolation procedure $\bar{f}(x) = \bar{f}(x, y_1, \ldots, y_m)$ such that*

$$\sup_{f\in\mathcal{F}(\gamma,L)} \|\bar{f} - f\| \le c \max_{i=1,\ldots,m} |\zeta_i| \log m + c(\gamma) L \exp(-\gamma m). \tag{9}$$

A straightforward proof of this result will be produced in the Appendix. Here again the interpolation formula (7) does the job.

Let us return now to the problem of noise-free recovery of an analytic edge function. Consider the following simple strategy. Fix an uniform grid $x_1, \ldots, x_{m(n)}$ consisting of $m(n)$ points on the interval $[0, 1]$. In each of the corresponding cross-sections $(x_i, \cdot) \subset S$ carry out $n/m(n)$ sequential observations to recover $f(x_i)$. (For the sake of exposition, we are dealing with $n/m(n)$ as with an integer; minor adjustments are needed to justify this!) The choice of the observation points in each of the cross-section is governed by the sequential Brent algorithm.

As we have seen, the error of estimating $f(x_i)$ is then $\exp[-\log 2 \cdot n/m(n)]$, for each $i = 1, ..., m(n)$. Therefore by (9) the error of the edge recovery, in the uniform norm, is bounded by

$$c \log(m(n)) \exp[-\log 2 \cdot n/m(n)] + c(\gamma) L \exp[-\gamma m(n)]. \tag{10}$$

Next we could choose $m(n)$ to minimize this bound. In fact, throughout the rest of the paper we will be mainly interested in the logarithmic behaviour of the attainable risk. Then it is not difficult to see that exponents in (10) should be made equal, that is

$$\log 2 \cdot n/m(n) = \gamma m(n),$$

or equivalently when $m(n) = \sqrt{\log 2 \cdot n/\gamma}$. By substituting this in (10), we obtain the following rate of convergence: $\log(n) \cdot \exp(-\sqrt{\log 2 \cdot \gamma n})$.

However, as we will see below, this simple method is not optimal, and we will show the above rate of convergence can be improved. More precisely, it will be shown, see Section 3.2, Remark 1, that the best (logarithmic) rate of convergence is $\exp(-\sqrt{2 \log 2 \cdot \gamma n})$. This demonstrates that the problem of recovery of analytic edges is non-trivial, even in the case of noise-free images.

Constant boundary with noisy data. To introduce the algorithm which will be used in the general case, when the observations are corrupted by a binary noise, let us return for a moment to the simplest case $f(x) = f = \text{const}$. The unknown value f is *a priory* restricted to the interval $[0, 1]$. The method of estimation in this situation was proposed by Burnashev and Zigangirov (1974). To describe the quality of an estimate $\hat{f}$, they used the error probability

$$P_e(\Delta, \hat{f}) = \max_{f\in[0,1]} \mathbf{P}_f \left\{ |\hat{f} - f| > \Delta \right\},$$

where the parameter Δ was typically depending on n in the following form: $\Delta = \exp(-Rn)$, for a given $R > 0$. Denote

$$E(R, \hat{f}) = -\frac{\log P_e(\Delta, \hat{f})}{n}.$$

Recall that the *capacity* of a symmetric binary channel, with the error probability $0 < x < 1$, is defined by

$$C(x) = \log 2 + H(x), \tag{11}$$

where

$$H(x) = x\log x + (1-x)\log(1-x).$$

Let

$$R_{cr} = C\left(\frac{\sqrt{p}}{\sqrt{p}+\sqrt{1-p}}\right).$$

For $R_{cr} \le R \le C(p)$, denote by $\rho(R) = \rho$ be the unique root of the equation $R = C(\rho)$ on $(0.5, p)$.

Theorem 1 *(Burnashev and Zigangirov (1974)). For any n and R there exists an estimator f^* such that*

$$E(R, f^*) \ge E(R) - \mathrm{e}^{-Rn}, \tag{12}$$

where $E(R)$, called reliability function, is defined as

$$E(R) = \begin{cases} \log \dfrac{2}{(\sqrt{p}+\sqrt{1-p})^2} - R, & 0 < R \le R_{cr}, \\ \\ \rho(R)\log\dfrac{\rho(R)}{1-p} + [1-\rho(R)]\log\dfrac{1-\rho(R)}{p}, & R_{cr} \le R < C(p). \end{cases}$$

Note that if the unknown f was a priory restricted to an interval $[0, e^{-dn}],\ 0 < d < R$ then $E(R)$ in (12) would be replaced, according to the rescaling argument, by $E(R-d)$. For reader's convenience the Burnashev-Zigangirov (BZ) procedure is outlined in the Appendix, whereas the typical shape of reliability function $E(R)$ is shown in Figure 1.

The general case. Our interest in the edges recovery problem in binary images models was triggered by a recent paper Korostelev (1999) which advocated the use of the BZ algorithm for the restoration of $\mathbf{C}^k$-smooth boundaries. In this paper, the edge function was recovered first on a regular grid $\{x_k\},\ k = 1, ..., m(n)$, by the implementing the BZ algorithm, in each of the cross-sections $(x_k, \cdot) \subset S$, using $n/m(m)$ observations. This followed by an interpolation procedure to recover the edge function on the whole interval $[0, 1]$. According to (12) and (9), using this strategy in the case of periodic analytic edge functions will result in the risk bounded by

$$1 \cdot m(n)\exp[-E(R)n/m(n)] + \exp[-Rn/m(n)]\log[m(n)] + c(\gamma)\exp[-\gamma m(n)].$$

Minimizing the logarithmic error, with respect to R and $m(n)$, we obtain the following equations

$$\frac{E(R)n}{m(n)} = \frac{Rn}{m(n)} = \gamma m(n)$$

leading to the following upper bound for the minimax risk:

$$r_n(\mathcal{F}(\gamma, L)) \le c\sqrt{n}\exp\left\{-\sqrt{r(p)\gamma n}\right\},$$

where $r(p)$ is the unique root of the equation $E(r) = r$.

The main result of the present paper shows that this rate of convergence can be improved.

Theorem 2 *For any* $\varepsilon > 0$*, there exists an estimator* $f^*(x, Y_1, \ldots, Y_n)$ *and a sequential design* $\mathbf{X}_1, \ldots \mathbf{X}_n$ *such that*

$$\sup_{f \in \mathcal{F}(\gamma, L)} \mathbf{E}_f \|f^* - f\| \leq c(\varepsilon)\sqrt{n} \exp\left\{-(1-\varepsilon)\sqrt{2r(p)\gamma n}\right\}, \tag{13}$$

where the constant $c(\varepsilon) > 0$ *does not depend on* n*.*

Note. It follows from the definition of the reliability function $E(R)$ that for sufficiently large p, say $p > p_0$, the root of the equation $E(r) = r$ is given by

$$r(p) = \log \frac{\sqrt{2}}{\sqrt{p} + \sqrt{1-p}}.$$

Theorem 2 will be proved in Section 4. In Section 3 we will discuss lower bounds in the edges recovery.

3 A lower bound

In this section we will prove the following lower bound for the minimax risk $r_n(\mathcal{F}(\gamma, L))$.

Theorem 3 *There exists a constant* $c > 0$ *such that*

$$r_n(\mathcal{F}(\gamma, L)) \geq cLn^{-1} \exp\left\{-\sqrt{2\gamma C(p)n}\right\}, \tag{14}$$

where C(p) is defined by (11).

The proof of this theorem is based on two ideas: ε - capacity of the class $\mathcal{F}(\gamma, L)$ and Fano's inequality. The reader will note that the assumption that f be a periodic function is not essential in this lower bound.

3.1 ε -capacity of the set $\mathcal{F}(\gamma, L)$

Suppose that we have a finite set of functions $f_1, \ldots, f_N,\ N = N(\varepsilon)$ such that:

- $f_i \in \mathcal{F}(\gamma, L),$ (15)
- $\|f_i - f_j\| \geq \varepsilon,\ i \neq j.$ (16)

Lemma 1 *For any* $\varepsilon > 0$*, one can find a family* $f_1, \ldots, f_N$ *satisfying (15–16), with*

$$N \geq N(\varepsilon) = \sqrt{\frac{L}{\varepsilon}} \exp\left\{\frac{1}{2\gamma} \log^2 \frac{L}{\varepsilon}\right\}. \tag{17}$$

Proof. A function $f_i(x) \in \mathcal{F}(\gamma, L)$ can be defined by its Fourier coefficients chosen at will under the conditions spelled out in (3). Let us chose the k-th Fourier coefficient f_k from the ε-net $\mathcal{A}_k$ of points

$$a_{kl} = \varepsilon l, \quad l = -A_k, \ldots, A_k,$$

where

$$A_k = \lfloor L\varepsilon^{-1} \exp(-2\gamma|k|) \rfloor.$$

Obviously non-zero Fourier coefficients in $\mathcal{A}_k$ could be found only for

$$|k| \le W,$$

where

$$W = W(\varepsilon) = \left\lfloor \frac{1}{2\gamma} \log \frac{L}{\varepsilon} \right\rfloor.$$

Let $f_i(x)$ and $f_j(x)$ be any two so defined functions and f_{ik}, f_{jk} be their Fourier coefficients correspondingly. Then by the Parseval identity

$$\|f_i - f_j\|^2 \ge \int_0^1 [f_i(x) - f_j(x)]^2 \, dx = \sum_{k=-W}^{W} (f_{ik} - f_{jk})^2 \ge \varepsilon^2,$$

and (16) is satisfied.

It remains to estimate the cardinality of the so defined ε-net $\{f_i(\cdot)\}$ which, by the multiplication principle, is given by

$$\begin{aligned} \prod_{k=-W}^{W} A_k = \prod_{k=-W}^{W} \frac{L}{\varepsilon} \exp(-2\gamma|k|) = \left(\frac{L}{\varepsilon}\right)^{2W(\varepsilon)+1} \exp\left\{-4\gamma \sum_{k=1}^{W(\varepsilon)} k\right\} \\ = \sqrt{\frac{L}{\varepsilon}} \exp\left\{\frac{1}{2\gamma} \log^2 \frac{L}{\varepsilon}\right\}. \qquad \square \end{aligned} \tag{18}$$

3.2 The Fano inequality

Suppose that the unknown function $f(x)$ is chosen randomly from a given finite set of functions $f_1(x), \dots, f_N(x)$ so that

$$\mathbf{P}\{f = f_k\} = \frac{1}{N}, \quad k = 1, \dots, N.$$

Any estimate $\hat{f}(x) = \hat{f}(x, Y(X_1), \dots, Y(X_n))$ of f, based on the observations (1) and an arbitrary sequential design $\mathbf{X}_1, \dots, \mathbf{X}_n$, could be used then to construct a selection procedure $\tilde{f} \in \{f_1, ..., f_N\}$, by choosing (one of) the closest to $\hat{f}$ elements among $f_1, ..., f_N$, in the sup-norm. Suppose we are interested in the classification error, or the average error probability

$$P_e(\tilde{f}) = \mathbf{P}\left\{\tilde{f} \ne f\right\} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{P}_{f_k} \left\{\tilde{f} \ne f_k\right\} = 1 - \frac{1}{N} \sum_{k=1}^{N} \mathbf{P}_{f_k} \left\{\tilde{f} = f_k\right\}.$$

This performance characteristic may be instrumental in evaluating the overall performance of the estimate $\hat{f}(\cdot)$. Indeed, if the ε-net satisfies (16), then obviously

$$\mathbf{E}_{f_k} \|\hat{f} - f_k\| \ge \mathbf{E}_{f_k} \|\hat{f} - f_k\| \mathbf{1}\left\{\tilde{f} \ne f_k\right\} \ge \frac{\varepsilon}{2} \mathbf{P}_{f_k} \left\{\tilde{f} \ne f_k\right\}.$$

Therefore

$$r_n(\mathcal{F}(\gamma, L)) \geq \inf_{\hat f} \frac{1}{N}\sum_{k=1}^{N} \mathbf{E}_{f_k}\|\hat f - f_k\| \geq \frac{\varepsilon}{2}\inf_{\tilde f} P_e(\tilde f)\,, \tag{19}$$

and any lower bound for the classification error $P_e = P_e(\tilde f)$ yields a lower bound for the minimax risk.

The following lower bound is based on the so called Fano inequality of the Information Theory; cf. Fano (1961), Ch. 6.2, or a more recent monograph Cover and Thomas (1991), Ch. 2.11. It was introduced in the non-parametric statistics by Ibragimov and Khas'minskii, cf. Ibragimov and Khas'minskii (1981), Ch. VII.1. Since the Fano inequality is quite elementary, it is befitting to remind it here.

Lemma 2 *For any classification rule $\tilde f$ and for any sequential design $\mathbf{X}_1, \ldots, \mathbf{X}_N$*

$$-H(P_e) + P_e \log(N-1) \geq \log N - C(p)n, \tag{20}$$

where $C(p)$ is defined by (11). This inequality can be weakened to

$$P_e \geq \frac{\log N - C(p)n - \log 2}{\log N}. \tag{21}$$

Proof. For any given design $\mathbf{X}_1, \ldots, \mathbf{X}_N$, denote the induced marginal distribution of the data $Y = (Y(X_1), \ldots, Y(X_n)) \in \{0,1\}^n$ by

$$\bar{\mathbf{P}}\{\cdot\} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{P}_{f_k}\{\cdot\}\,.$$

The Shannon information contained in the data, about the random element f_k is given by

$$I = \bar{\mathbf{E}} \log \frac{d\mathbf{P}_{f_k}(Y)}{d\bar{\mathbf{P}}(Y)} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{E}_{f_k} \log \frac{d\mathbf{P}_{f_k}}{d\bar{\mathbf{P}}}.$$

Since the function $\log(x)$ is convex on $x \in (0,\infty)$, the Jensen inequality yields

$$\begin{aligned}
-I &= \frac{1}{N}\sum_{k=1}^{N} \mathbf{E}_{f_k} \mathbf{1}\left\{\hat f = f_k\right\} \log \frac{d\bar{\mathbf{P}}}{d\mathbf{P}_{f_k}} + \frac{1}{N}\sum_{k=1}^{N} \mathbf{E}_{f_k} \mathbf{1}\left\{\hat f \neq f_k\right\} \log \frac{d\bar{\mathbf{P}}}{d\mathbf{P}_{f_k}} \\
&\leq (1-P_e)\log\left\{\frac{1}{(1-P_e)N}\sum_{k=1}^{N} \mathbf{E}_{f_k}\mathbf{1}\left\{\hat f = f_k\right\}\frac{d\bar{\mathbf{P}}}{d\mathbf{P}_{f_k}}\right\} \\
&\quad + P_e \log\left\{\frac{1}{P_e N}\sum_{k=1}^{N} \mathbf{E}_{f_k}\mathbf{1}\left\{\hat f \neq f_k\right\}\frac{d\bar{\mathbf{P}}}{d\mathbf{P}_{f_k}}\right\} \\
&= (1-P_e)\log\left\{\frac{1}{(1-P_e)N}\sum_{k=1}^{N} \bar{\mathbf{P}}\left\{\hat f = f_k\right\}\right\} \\
&\quad + P_e \log\left\{\frac{1}{P_e N}\sum_{k=1}^{N} \left[1 - \bar{\mathbf{P}}\left\{\hat f = f_k\right\}\right]\right\} \\
&= (1-P_e)\log\left\{\frac{1}{(1-P_e)N}\right\} + P_e\log\left\{\frac{N-1}{P_e N}\right\} \\
&= -H(P_e) + P_e\log(N-1) - \log N.
\end{aligned} \tag{22}$$

So to prove the Lemma, it remains to establish a corresponding upper bound on the Shannon information I. We have

$$I = \frac{1}{N}\sum_{k=1}^{N} \mathbf{E}_{f_k} \log \frac{d\mathbf{P}_{f_k}}{d\bar{\mathbf{P}}} = \frac{1}{N}\sum_{k=1}^{N}\sum_{y\in\{0,1\}^n} \mathbf{P}_{f_k}\{y\}\log\frac{\mathbf{P}_{f_k}\{y\}}{\bar{\mathbf{P}}\{y\}}$$
$$= \sum_{y\in\{0,1\}^n} \bar{\mathbf{P}}\{y\}\log\frac{1}{\bar{\mathbf{P}}\{y\}} + \frac{1}{N}\sum_{k=1}^{N}\sum_{y\in\{0,1\}^n} \mathbf{P}_{f_k}\{y\}\log \mathbf{P}_{f_k}\{y\}. \tag{23}$$

Using again the Jensen inequality, one obtains

$$\sum_{y\in\{0,1\}^n} \bar{\mathbf{P}}\,\{y\}\log\frac{1}{\bar{\mathbf{P}}\,\{y\}} \le \log \sum_{y\in\{0,1\}^n} 1 = \; n\log 2. \tag{24}$$

It remains to note that the last term in (23) equals $nH(p)$. This is easy to prove using, for any f_k, the representation

$$\mathbf{E}_{f_k}\log \mathbf{P}_{f_k}\{Y\} = \mathbf{E}_{f_k}\sum_{l=1}^{n}\log \mathbf{P}_{f_k}\{Y(X_l)|Y(X_1),...,Y(X_{l-1})\}$$

and noting that although the conditional distribution of $Y(X_l)$ depends on the past, which may effect the choice of the design points $X_s,\ s\le l$ in (1) - it's either $\mathbf{P}_{f_k}\{1\}=p$, or else $\mathbf{P}_{f_k}\{0\}=p$, its entropy $H(p)$ does not depend on the past. This result is well known in the information theory; see e.g. Cover and Thomas (1991), Ch. 8.12. □

Proof of Theorem 3. By Lemma 1, one can choose, for any given $\varepsilon > 0$, a family of functions $f_1,\dots,f_{N(\varepsilon)}$ simultaneously satisfying (15–17). Now choose ε as a root of the equation

$$\log N(\varepsilon) = C(p)n + \log 2e$$

or in view of (17) as

$$\varepsilon \ge L\exp\left\{-\sqrt{2\gamma(C(p)n+\log 2e)}\right\}.$$

The theorem follows now from (21) and (19). □

4 Recovery of analytic functions

4.1 Sequential interpolation and recovery

The following interpolation problem is a quintessential part of the proof of Theorem 2. Suppose first that we want to interpolate $f(x)$ based on the data $y(x)=f(x)$, where x runs through the set of m equidistant points

$$x_{j1} = \frac{j-1}{m}, \quad j=1,\dots,m.$$

For brevity denote this set by $\mathbf{x}_1^m$ and the corresponding data collectively as by $y(\mathbf{x}_1^m)$. According to Proposition 1, the interpolation error is bounded by

$$\inf_{\bar{f}}\sup_{f\in\mathcal{F}(\gamma,L)} \|\bar{f}(y(\mathbf{x}_1^m)) - f\| \le c(\gamma)L\exp(-\gamma m).$$

Now suppose we would like to improve our interpolation, by adding m new interpolation points. Obviously if we define this set of points, say $\mathbf{x}_2^m$, as

$$x_{j2} = \frac{1}{2m} + \frac{j-1}{m}, \quad j = 1, \dots, m$$

then the union of $\mathbf{x}_1^m$ and $\mathbf{x}_2^m$ is again a uniform grid. Hence we get the interpolation error

$$\inf_{\bar{f}} \sup_{f \in \mathcal{F}(\gamma, L)} \|\bar{f}(y(\mathbf{x}_1^m \cup \mathbf{x}_2^m)) - f\| \leq c(\gamma) L \exp(-2\gamma m).$$

What happens next when we try to improve our interpolation of by adding another m points? Suppose we chose this new set, say $\mathbf{x}_3^m$, as

$$x_{j3} = \frac{1}{4m} + \frac{j-1}{m}, \quad j = 1, \dots, m$$

What interpolation error could we achieve now? Heuristically, one would expect that

$$\inf_{\bar{f}} \sup_{f \in \mathcal{F}(\gamma, L)} \|\bar{f}(y(\mathbf{x}_1^m \cup \mathbf{x}_2^m \cup \mathbf{x}_3^m)) - f\| \leq c(\gamma) L \exp(-3\gamma m). \tag{25}$$

This inequality, however, is less trivial, since the interpolation design is no longer equidistant. Traditionally , the interpolation problems of this kind are shunned away in the existing textbooks. The interpolation method outlined next this Section is largely motivated by the theory of so-called *multi-channel sampling*; see Kohlenberg (1953) and a recent exposition in Higgins (1996), Ch. 12, 13.

When the resources permit adding another m interpolation points $\mathbf{x}_4^m$, we would choose them as

$$x_{i4} = \frac{3}{4m} + \frac{i-1}{m}, \quad i = 1, \dots, m.$$

so that the lattices $\mathbf{x}_1^m, \mathbf{x}_2^m, \mathbf{x}_3^m, \mathbf{x}_4^m$ together form again a uniform grid:

$$x_j = \frac{j-1}{2m}, \quad j = 1, \dots, 2m.$$

Obviously, in this case

$$\inf_{\bar{f}} \sup_{f \in \mathcal{F}(\gamma, L)} \|\bar{f}(y(\mathbf{x}_1^m \cup \mathbf{x}_2^m \cup \mathbf{x}_3^m \cup \mathbf{x}_4^m)) - f\| \leq c(\gamma) L \exp(-4\gamma m).$$

After that, we could continue adding m interpolation points at a time. In this way, after k steps we will arrive at k different lattices $\mathbf{x}_s^m,\ s = 1, ..., k$ of points

$$x_{js} = \frac{a_s + j - 1}{m}, \quad j = 1, \dots, m,$$

where, in case $k = 2^u + v,\ u = 0, 1, ...,\ v = 0, ..., 2^u - 1$, we define a_s as

$$a_s = \begin{cases} (s-1)2^{-u}, & s = 1, \dots, 2^u, \\ [2(s - 2^u - 1) + 1]2^{-u-1}, & s = 2^u + 1, \dots, k. \end{cases}$$

Denote

$$\mathbf{S}_k^m = \bigcup_{s=1}^{k} \mathbf{x}_s^m$$

Lemma 3 *For any given integer k there exists an interpolation $\bar{f}(x, y(\mathbf{S}_k^m))$ of the function f based on the data*

$$y(x_i) = f(x_i), \; x_i \in \mathbf{S}_k^m, \quad i = 1, \ldots, km, \tag{26}$$

such that

$$\sup_{f \in \mathcal{F}(\gamma, L)} \|\bar{f}(y(\mathbf{S}_k^m)) - f\| \le c(k, \gamma) L \exp(-\gamma k m).$$

Proof. Obviously, it is sufficient to prove this Lemma in the case when m is even. Our proof below is motivated by the technique of multi-channel sampling. First, define the functions

$$f^s(t) = f\left(t + \frac{a_s}{m}\right), \quad s = 1, ..., k.$$

So we have to recover $f(t)$, $t \in [0, 1]$, from k samples $f^s((j-1)/m)$, $s = 1, \ldots, k$, $j = 1, ..., m$. The underlying idea is based on two obvious principles:

- using $k \times m$ available values $f^s((j-1)/n)$, we calculate estimates $\tilde{f}_l$ of the spectrum of f in the band $(-km/2, km/2)$;
- finally we reconstruct f using the formula

$$\tilde{f}(t) = \sum_{|l| < km/2} \tilde{f}_l \exp(-2\pi i l t). \tag{27}$$

To this end, let $\hat{f}_l^s$ be the empirical discrete Fourier transform of $f^s(t)$ defined as

$$\hat{f}_l^s = \frac{1}{m} \sum_{j=0}^{m-1} f^s\left(\frac{j}{m}\right) \exp\left(\frac{2\pi i j l}{m}\right).$$

Let f_l and f_l^s be the Fourier coefficients of $f(t)$ and $f^s(t)$ respectively. A simple algebra shows that

$$\hat{f}_l^s = \sum_{p=-\infty}^{\infty} f_{l+pm}^s,$$

$$f_l^s = \exp\left(-\frac{2\pi i l a_s}{m}\right) f_l.$$

Substituting the last equation into the previous one, we get

$$\hat{f}_l^s = \sum_{p=-\infty}^{\infty} f_{l+pm} \exp\left(-\frac{2\pi i (l+pm) a_s}{m}\right).$$

Let $l \in [0, n/2]$. Then $\hat{f}_l^s$ can be represented as

$$\hat{f}_l^s = \sum_{p:\, |l+pm| < km/2} f_{l+pm} \exp\left(-\frac{2\pi i (l+pm) a_s}{m}\right) + \sum_{p:\, |l+pm| \ge km/2} f_{l+pm} \exp\left(-\frac{2\pi i j (l+pm) a_s}{m}\right).$$

Note that the last term in this formula is bounded in absolute value by

$$\sum_{p:\,|l+pm|\geq km/2} |f_{l+pm}| \leq CL\exp(-k\gamma m). \tag{28}$$

Therefore we can ignore this term. Out of remaining $km-1$ Fourier coefficients in (28), we can determine those k coefficients $f_{l'}$ with $l' = l \bmod(m)$, as solution of the following linear equations

$$\sum_{p:\,|l+pm|<km/2} \tilde{f}_{l+pm}\exp\left(-2\pi i p a_s\right) = \exp\left(\frac{2\pi i l a_s}{m}\right)\hat{f}_l^s, \quad s = 1, ..., k. \tag{29}$$

Now for any given $l \in [0, m/2]$, we have k equations which uniquely define k Fourier coefficients $f_{l'}$ with $|l'| < km/2,\ l' = l \bmod (m)$. Indeed, since the set of functions $\exp(2\pi i l), |l| < km/2$ is a *Chebyshev system* on [0,1) and therefore is linearly independent on any set of km distinct points in $[0,1)$ (see e.g. Timan (1963), Sect. 2.3), our data uniquely determine $f_l,\ |l| < km/2$ in the special case when all remaining Fourier coefficients $f_l,\ |l| \geq km/2$ vanish. Therefore, the determinant of (29) does not vanish and obviously it also does not depend on m.

This implies that we can recover all the Fourier coefficients $f_{l'}$ with $|l'| < km/2,\ l' = l \bmod (m)$ from (29), with the accuracy $O(\exp(-k\gamma m))$ uniformly in $l \in [0, m/2]$. In order to reconstruct the spectrum f_l on the whole interval $[-km/2, km/2]$ we use the formula

$$f_l = f^*_{-l}.$$

Then uniformly in and $k \in [-km/2, km/2]$,

$$|\tilde{f}_k - f_k| \leq CL\exp(-k\gamma m). \tag{30}$$

Finally by (27) and (30) one obtains

$$\begin{aligned}\sup_t |f(t) - \tilde{f}(t)| &\leq \sup_t \left| \sum_{k=-\infty}^{\infty} f_l \exp(-2\pi i l t) - \sum_{|l|\leq km/2} \tilde{f}_l \exp(-2\pi i l t) \right| \\ &\leq \sum_{|l|\geq km/2} |f_l| + \sum_{|l|\leq km/2} |\tilde{f}_l - f_l| \leq c\exp(-k\gamma m). \quad \square\end{aligned}$$

The following lemma deals with interpolation with errors in the data.

Lemma 4 *For any integer k there exists an interpolation formula $\bar{f}(x, y(\mathbf{S}_k^m))$ based on the data*

$$y(x_i) = f(x_i) + \zeta_i,\ x_i \in \mathbf{S}_k^m, \quad i = 1, \ldots, km,$$

such that

$$\sup_{f\in\mathcal{F}(\gamma,L)} \|\bar{f}(y(\mathbf{S}_k^m)) - f\| \leq c \max_{i=1,\ldots,m} |\zeta_i| k\log(m) + c(\gamma, k)L\exp[-\gamma k m].$$

Proof. The required inequality can be easily obtained by combining the methods used in proving Lemma 3 and Proposition 2. $\square$

4.2 Proof of the Theorem 2

Let us first describe our method of estimation. Our procedure will be carried out in K successive steps where K is a sufficiently large integer; to be precise, one can choose K such that

$$(\varepsilon^{-1} - 1)/2 \leq K < (\varepsilon^{-1} + 1)/2.$$

At the k-th step, an intermediate estimate $\bar{f}_k(x)$ will be constructed and a confidence band B_k for $f(x)$ will be obtained based on $\bar{f}_k(x)$. This confidence band will be used to plan the strategy for the next step, whereupon a more accurate estimate $\bar{f}_{k+1}(x)$ will be found. Each step of the algorithm will consist of three distinctive parts: allocation of new cross-sections, estimation of $f(x)$ at the newly chosen cross-sections, and interpolation of the obtained results. The principles governing all three parts of the procedure are simple.

- At the k-th step, $k = 1, ..., K$, we add $M = \lfloor \alpha\sqrt{n} \rfloor$ new cross-sections allocated at the points of the grid $\mathbf{x}_k^M = \mathbf{x}_k$, following the rules explained in the previous section.
- To estimate $f(x)$ for $x \in \mathbf{x}_k$, we make $N_k = \lfloor \beta_k\sqrt{n} \rfloor$ sequential observations, using the BZ-procedure as described in Section 2, at each of the newly added cross-sections. Denote the thus estimated values $\hat{f}_k(x),\ x \in \mathbf{x}_k$.
- We interpolate all the obtained values $\hat{f}_l(x), x \in \mathbf{x}_l,\ 1 \leq l \leq k$, up to the most recent ones, using the procedure described in Lemma 4.

Since we have at our disposal n observations, we arrive at the following restriction on the number of observations:

$$M \sum_{k=1}^{K} N_k \leq n$$

or equivalently

$$\alpha \sum_{k=1}^{K} \beta_k \leq 1. \tag{31}$$

Note that when the procedure stops after K steps, there will be KM interpolation points accumulated altogether. By Lemma 3, even if the values of $f(x)$ at all chosen cross-sections were known exactly, we would have the interpolation error to the (logarithmic) order $\exp(-K\alpha\gamma\sqrt{n})$. Thus to achieve this order of interpolation in the noisy recovery problem, using the estimated values $\hat{f}_l(x)$, we must guarantee, according to the Proposition 2, that the expected error terms $\mathbf{E} \max_i |\zeta_i| = \mathbf{E}_f \max_{x \in \cup \mathbf{x}_i} |\hat{f}_l(x) - f(x)|$ are to the same (logarithmic) order $\exp(-K\alpha\gamma\sqrt{n})$. To achieve this, recall that the BZ estimates depends on a parameter Δ determining the rate of transmission R; cf. Theorem 1. Thus in each case when the sequential BZ algorithm is applied, we will choose

$$\Delta = \exp(-K\alpha\gamma\sqrt{n}).$$

It remains to specify the parameters α and β_k. We will do so after formally describing each of the steps. In this we will use the shortened notation $B_k = [b_k^-(x), b_k^+(x)]$ for a confidence set of the form

$$B_k = \left\{ f : b_k^-(x) \leq f(x) \leq b_k^+(x),\ 0 \leq x \leq 1 \right\}$$

and denote by

$$\operatorname{diam}(B_k) = \sup_{0\le x\le 1} [b_k^+(x) - b_k^-(x)].$$

Let $b_0^-(x) \equiv 0, b_0^+(x) \equiv 1$, so that $D_0 := \operatorname{diam}(B_0) = 1$.

Step 1. Assuming a priory that $f \in B_0$, carry out $N_1 = \lfloor \beta_1\sqrt{n} \rfloor$ sequential observations using the BZ, specified by the parameter Δ, at each of the $M = \lfloor \alpha\sqrt{n} \rfloor$ cross-sections allocated at the points of grid $\mathbf{x}_1$. Compute thus the estimates $\hat{f}_1(x), x \in \mathbf{x}_1$ and denote by A_1 the following event

$$A_1 = \left\{ \sup_{x\in\mathbf{x}_1} |\hat{f}_1(x) - f(x)| \le \Delta \right\}.$$

Using the procedure described in Lemma 4, interpolate the obtained values $\hat{f}_1(x),\ x \in \mathbf{x}_1$, to obtain $\bar{f}_1(x),\ x \in [0,1]$. Note that given A_1,

$$\|\bar{f}_1 - f\| \le D_1 := c\Delta \log n + c(\gamma,1)L\exp(-\gamma\alpha\sqrt{n}).$$

Therefore the estimate $\bar{f}_1(x)$ leads to the improved confidence band $B_1 = [b_1^-(x), b_1^+(x)]$ where

$$b_1^-(x) = \min\{b_0^+(x), \max[b_0^-(x), \bar{f}_1(x) - D_1]\},$$
$$b_1^+(x) = \max\{b_0^-(x), \min[b_0^+(x), \bar{f}_1(x) + D_1]\}$$

and

$$\operatorname{diam}(B_1) \le D_1.$$

Now assuming that $k-1$ steps have been already implemented ($2 \le k \le K$), we describe the follow-up step.

Step k. At this step, we first construct an estimate $\hat{f}_k(x)$ at the M newly allocated cross-sections $\mathbf{x}_k$. While using again the BZ algorithm, we assume that the unknown function $f(x)$ *a priori* belongs to the confidence band B_{k-1}. Denote by A_k the event

$$A_k = \left\{ \sup_{x\in\mathbf{x}_k} |\hat{f}_k(x) - f(x)| \le \Delta \right\}.$$

Next, using the procedure described in Lemma 4, interpolate all the obtained values $\hat{f}_l(x),\ x \in \mathbf{x}_l$ for $l = 1,...,k$ thus obtaining the updated estimate $\bar{f}_k(x),\ x \in [0,1]$. Note that if $A_1,...,A_k$ hold true, then by Lemma 4,

$$\|\bar{f}_k - f\| \le D_k := c\Delta k \log n + c(\gamma,k)L\exp(-k\gamma\alpha\sqrt{n}). \tag{32}$$

Thus given $A_1,...,A_k$, the unknown function f is contained in the confidence band $B_k = [b_k^-(x), b_k^+(x)]$ given by

$$b_k^-(x) = \min\left\{b_{k-1}^+(x), \max[b_{k-1}^-(x), \bar{f}_k(x) - D_k]\right\},$$
$$b_k^+(x) = \max\left\{b_{k-1}^-(x), \min[b_{k-1}^+(x), \bar{f}_k(x) + D_k]\right\}$$

and

$$\operatorname{diam}(B_k) \le D_k.$$

Note that

$$B_k \subset B_{k-1} \quad \text{and} \quad D_k \le D_{k-1}.$$

Finally, Theorem 1 and the remark following it show that the confidence level of B_k is bounded by

$$\sup_{f\in\mathcal{F}(\gamma,L)} \mathbf{P}_f\left\{ A_k^c \middle| \bigcap_{i=1}^{k-1} A_i \right\} \le \alpha\sqrt{n}\exp\left\{-E((K-(k-1))\alpha\gamma/\beta_k)\beta_k\sqrt{n}\right\}. \tag{33}$$

The algorithm ends with obtaining the final estimate $f^*(x)$, by projecting (if necessary) the estimate $\bar{f}_K(x)$ to the confidence band B_{K-1}:

$$f^*(x) = \max\left\{b_{K-1}^-(x), \min[b_{K-1}^+(x), \bar{f}_K(x)]\right\}.$$

Now we are ready to finish the proof. Denote

$$A = \bigcap_{i=1}^{K} A_i = A_1 \cdots A_K.$$

Note that the complement of A can be represented as

$$A^c = A_1^c + A_1A_2^c + \ldots + A_1\cdots A_{K-1}A_K^c.$$

Also, by the construction of the algorithm, for any $l = 1, ..., K$

$$f^*(x) \in B_{K-1} \subset B_{l-1}$$

whereas $f \in B_{l-1}$ given $A_1 \cdots A_{l-1}$, cf. (32). Therefore given $A_1 \cdots A_{l-1}$

$$\sup_{0\le x\le 1} |f^*(x) - f(x)| \le D_{l-1}.$$

Using (33), the risk of the estimate $f^*(x)$ can be bounded from above as follows

$$\begin{aligned}
\mathbf{E}_f \sup_{0\le x\le 1} |f^*(x)-f(x)| &= \mathbf{E}_f \sup_{0\le x\le 1} |f^*(x)-f(x)| \left[\mathbf{1}\left\{A^c\right\} + \mathbf{1}\left\{A\right\}\right] \\
&\le \mathbf{E}_f\left\{\sup_{0\le x\le 1}|f^*(x)-f(x)| \middle| A\right\} + \mathbf{E}_f\left\{\sup_{0\le x\le 1}|f^*(x)-f(x)| \middle| A_1^c\right\}\mathbf{P}_f\left\{A_1^c\right\} \\
&\quad + \sum_{l=2}^{K} \mathbf{E}_f\left\{\sup_{0\le x\le 1}|f^*(x)-f(x)| \middle| A_1\cdots A_{l-1}A_l^c\right\}\mathbf{P}_f\left\{A_1\cdots A_{l-1}A_l^c\right\} \\
&\le \exp(-K\alpha\gamma\sqrt{n}) + C\alpha\sqrt{n}\sum_{l=1}^{K}\exp\left\{-[(l-1)\alpha\gamma + E((K-(l-1))\alpha\gamma/\beta_l)\beta_l]\sqrt{n}\right\}.
\end{aligned}$$

To obtain a balanced risk formula, in which all the terms are to the same (logarithmic) order, the parameters β_l must satisfy

$$E((K-(l-1))\alpha\gamma/\beta_l) = (K-(l-1))\alpha\gamma/\beta_l, \tag{34}$$

or equivalently

$$\beta_l = (K - (l-1))\alpha\gamma/r(p), \quad l = 1, 2, ..., K. \tag{35}$$

Remind that $r(p)$ was defined as the root of the equation $E(r) = r$. Therefore

$$\mathbf{E}_f \|f^* - f\| \le C(K)\alpha\sqrt{n}\exp\left(-K\alpha\gamma\sqrt{n}\right). \tag{36}$$

Here $C(K)$ stands for a generic constant independent of n.

It remains to determine the parameter α. Substituting (35) and in (31) one obtains

$$1 = \alpha\sum_{k=1}^{K}\beta_k = \frac{\gamma\alpha^2}{r(p)}\sum_{k=1}^{K}(K-(k-1)) = \frac{\gamma\alpha^2 K(K+1)}{2r(p)}.$$

This gives

$$\alpha = \sqrt{\frac{2r(p)}{\gamma K(K+1)}}$$

and by (36) the risk of $f^*(x)$ is bounded from above by

$$\begin{aligned}\sup_{f\in\mathcal{F}(\gamma,L)} \mathbf{E}_f \|f^* - f\| &\le C(K)\sqrt{n}\exp\left\{-\sqrt{\frac{2\gamma r(p)nK}{K+1}}\right\}\\ &\le C(K)\sqrt{n}\exp\left\{-(1-\varepsilon)\sqrt{2\gamma r(p)n}\right\}. \quad \square\end{aligned}$$

Remark 1. The method proposed in the proof of Theorem 2 can be used for the noiseless data. This situation is simpler since one can use the Brent algorithm in order to recover the function in the cross-sections. So in this case the balance condition (34) takes the form

$$K\gamma\alpha = (k-1)\gamma\alpha + \beta_k \log 2.$$

Therefore the situation is analogous to the considered case with $r = \log 2$ and the risk of the procedure bounded from above by $C(K)\sqrt{n}\exp\left\{-(1-\varepsilon)\sqrt{2\log 2\cdot\gamma n}\right\}$. Comparing this with the lower bound from Theorem 3 and noticing that $\lim_{p\to 0} C(p) = \log 2$, we see that our method almost attains (up to arbitrary ε) the lower bound.

Remark 2. We believe the inequality (13) can be improved. In the previous remark we have seen how it could be done for $p = 1$. To do that for arbitrary p, one needs a revision of the method by Burnashev and Zigangirov. Unfortunately their estimator depends strongly on R. This prevents to improve (13). If one could find an algorithm which does not depend on R and for which (12) holds true then it will be possible to show that the upper bound of recovering analytic functions is given by

$$c(\varepsilon)\sqrt{n}\exp\left\{-(1-\varepsilon)\sqrt{2r_0\gamma n}\right\},$$

where

$$r_0 = \min_{0<R\le C(p)}[R + E(R)] = \log 2 - 2\log(\sqrt{p} + \sqrt{1-p}).$$

Notice that $r_0 = 2r$ for sufficiently small p.

5 Appendix

5.1 Proof of Proposition 1

We use the following interpolation formula

$$\bar{f}(x) = \bar{f}(x, y_1, \ldots, y_m) = \sum_{|k| \leq m/2} \bar{f}_k \exp(-2\pi i k x), \tag{37}$$

with empirical Fourier coefficients

$$\bar{f}_k = \frac{1}{m} \sum_{l=1}^{m} y_l \exp(2\pi i k x_l)$$

Denoting for $f_k = \langle f, \exp(-2\pi i k \cdot) \rangle$, we can write

$$\begin{aligned}
\bar{f}_k &= \frac{1}{m} \sum_{l=1}^{m} \sum_{s=-\infty}^{\infty} f_s \exp(-2\pi i s x_l) \exp(2\pi i k x_l) \\
&= \sum_{s=-\infty}^{\infty} f_s \frac{1}{m} \sum_{l=1}^{m} \exp[-2\pi i (s-k) l/m] \\
&= \sum_{s=-\infty}^{\infty} f_s \sum_{p=-\infty}^{\infty} \delta_{s-k-pm} = \sum_{p=-\infty}^{\infty} f_{k+pm}
\end{aligned}$$

where δ_n is the Kronecker delta. Therefore

$$\begin{aligned}
\max_{x \in [0,1]} |\bar{f}(x) - f(x)| &\leq \sum_{|k| > m/2} |f_k| + \sum_{|k| \leq m/2} \left| f_k - \sum_{p=-\infty}^{\infty} f_{k+pm} \right| \\
&\leq \frac{2L}{1 - \exp(-2\gamma)} \exp(-\gamma m) + \sum_{|k| \leq m/2} \sum_{|p| > 0} |f_{k+pm}| \\
&\leq \frac{2L}{1 - \exp(-2\gamma)} \exp(-\gamma m) + \frac{2L}{(\exp(2\gamma) - 1)(1 - \exp(-\gamma m))} \exp(-\gamma m).
\end{aligned}$$

This inequality proves the proposition. □

5.2 Proof of Proposition 2

We can rewrite (37) as

$$\bar{f}(x, y_1, \ldots, y_m) = \sum_{i=1}^{m} f(x_i) \frac{\sin[\pi(x - x_i)m]}{m \sin[\pi(x - x_i)]} + \frac{1}{m} \sum_{i=1}^{m} \zeta_i \frac{\sin[\pi(x - x_i)m]}{\sin[\pi(x - x_i)]}. \tag{38}$$

We have seen that

$$\left| f(x) - \sum_{i=1}^{m} f(x_i) \frac{\sin[\pi(x - x_i)m]}{m \sin[\pi(x - x_i)]} \right| \leq c(\gamma) \exp(-\gamma m).$$

Denote $\max_{i=1,\dots,m}|\zeta_i| = \epsilon$. Using the inequality $\sin(x) \ge 2x/\pi,\ x \in [0, \pi/2]$, the last term in (38) can be estimated very easily

$$\sup_x \left| \frac{1}{m} \sum_{i=1}^m \zeta_i \frac{\sin[\pi(x-x_i)m]}{\sin[\pi(x-x_i)]} \right| \le \frac{\epsilon}{m} \sup_x \sum_{i=1}^m \left| \frac{\sin[\pi(x-x_i)m]}{\sin[\pi(x-x_i)]} \right|$$
$$\le 2\epsilon + \max_{0 \le x \le 1/(2m)} \frac{2\epsilon}{m} \sum_{l=1}^{m/2} \frac{1}{\sin[\pi(l/m - x)]}$$
$$\le 2\epsilon + \frac{2\epsilon}{m} \sum_{l=1}^{m/2} \frac{1}{\sin[\pi(l-0.5)/m]} \le 2\epsilon + \epsilon \sum_{l=1}^{m/2} \frac{1}{l-0.5} \le c\epsilon \log m. \quad \square$$

5.3 Method by Burnashev and Zigangirov

We describe here the sequential procedure proposed in Burnashev and Zigangirov (1974) for estimating an unknown jump point in the binary regression model. It is assumed that we have at our disposal the noisy data

$$Y_i = 1\{x_i \le \theta\} \oplus \xi_i, \quad i = 1, \dots, n.$$

The goal is to estimate the parameter $\theta \in [0, 1]$, assuming that the design points x_i are sequentially chosen. The performance of an estimator $\hat{\theta}_n$ is measured by the error probability

$$\sup_{\theta \in [0,1]} \mathbf{P}_\theta \left\{ |\hat{\theta}_n - \theta| \ge \Delta \right\}.$$

Define the intervals $\Delta_i = [(i-1)\Delta, i\Delta]$, $i = 1, \dots, \Delta^{-1}$ and assuming that θ has the uniform prior distribution on $[0, 1]$, compute the pseudo posterior probabilities

$$\pi_s^k = \frac{\displaystyle\int_{\Delta_k} \prod_{i=1}^s \beta^{1\{x_i \le \theta\} \oplus Y_i} \alpha^{1 - 1\{x_i \le \theta\} \oplus Y_i} d\theta}{\displaystyle\int_0^1 \prod_{i=1}^s \beta^{1\{x_i \le \theta\} \oplus Y_i} \alpha^{1 - 1\{x_i \le \theta\} \oplus Y_i} d\theta},$$

where $\beta = 1 - \alpha$ and

$$\alpha = \begin{cases} \sqrt{1-p}/(\sqrt{1-p} + \sqrt{p}), & 0 \le R \le R_{cr}, \\ \rho, & R_{cr} < R < C(p). \end{cases}$$

The design points x_i are chosen from the points $i\Delta$, $i = 1, \dots, \Delta^{-1}$. Suppose we have already chosen the points $x_1, \dots, x_j$ and we want to find x_{j+1}. This can be done as follows. First find integer $k = k(j)$ such that

$$\sum_{s=1}^k \pi_j^s \le \frac{1}{2}, \quad \sum_{s=1}^{k+1} \pi_j^s > \frac{1}{2}.$$

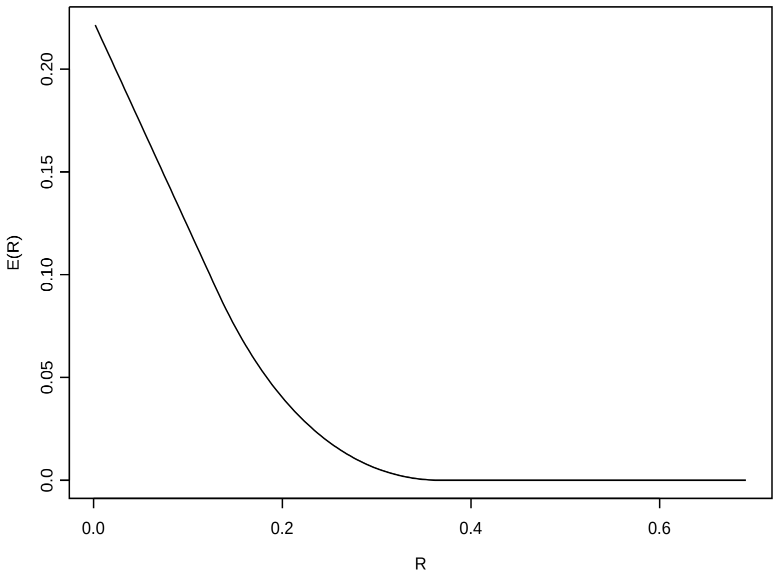

Figure 1: Reliability function $E(R)$ for $p = 0.9$. This function is linear for $0 \le R \le R_{cr} = 0.130812$ and zero for $R \ge C(p) = 0.3680642$.

Then x_{j+1} can be $k\Delta$ or $(k+1)\Delta$. We chose the point $k\Delta$ randomly with probability p_1 and the point $(k+1)\Delta$ respectively with probability $p_2 = 1 - p_1$. The probability p_1 is function of R, p and the values τ_1, τ_2 defined by

$$\tau_1 = \sum_{s=k+1}^{1/\Delta} \pi_j^s - \sum_{s=1}^{k} \pi_j^s, \quad \tau_2 = \sum_{s=1}^{k+1} \pi_j^s - \sum_{s=k+2}^{1/\Delta} \pi_j^s.$$

The probability p_1 is computed for $R \in [0, R_{cr}]$ as

$$p_1 = \frac{\tau_2}{\tau_1 + \tau_2}$$

and for $R \in (R_{cr}, C(p))$ as

$$p_1 = \frac{[1 + (1-2\alpha)\tau_2]^\lambda - [1 - (1-2\alpha)\tau_2]^\lambda}{[1 + (1-2\alpha)\tau_2]^\lambda - [1 - (1-2\alpha)\tau_2]^\lambda + [1 + (1-2\alpha)\tau_1]^\lambda - [1 - (1-2\alpha)\tau_1]^\lambda},$$

where

$$\lambda = \frac{\log[p/(1-p)]}{\log[\rho/(1-\rho)]} - 1.$$

References

[1] Brent, R. (1973). *Algorithms for minimization without derivatives.* Prentice-Hall Series in Automatic Computation, Englewood Cliffs, NJ, USA.

[2] Burnashev, M. V., Zigangirov K. Sh. (1974) An interval estimation problem for controlled observations. *Problems of Information Transmission* **10**, 223–231.

[3] Cover, T. M., Thomas J. A. (1991) *Elements of Information Thoery.* John Wiley & Sons, New York, Toronto.

[4] Fano R. (1961) *Transmission of information. A statistical theory of communications* M.I.T. Press and John Wiley & Sons, Inc. New York, London.

[5] Higgins, J. R. (1996) *Sampling theory in Fourier and Signal Analysis.* Clarendon press. Oxford.

[6] Ibragimov, I. A., Has'minskii, R. Z. (1981) *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

[7] Kohlenberg, A. (1953). Exact interpolation of band-limited functions. *J. Appl. Phys.,* **24**, 1432–1436.

[8] Korostelev, A. (1999) On minimax rates of convergence in image models under sequential design. *Statistics and Probability Letters* **43**, 369–375.