*On the Recovery of Discrete Probability Densities from Imperfect Measurements**

by LUC P. DEVROYE

School of Computer Science, McGill University, Montreal, Canada H3C 3G1

and GARY L. WISE

Department of Electrical Engineering, University of Texas at Austin, Austin, Texas 78712, U.S.A.

ABSTRACT: *The problem of the estimation of a discrete probability density from independent observations is considered. For a wide class of noises, a method is given for estimating a probability density when the measurements are corrupted by additive noise. This method is shown to be consistent, and several bounds on the error are given. An application to the detection of a (nonparametric) random signal is discussed. Finally, the estimation of a probability density is considered where the measurements are noisy and some of the measurements are incorrect. This situation may arise when a machine collecting the data fails part of the time.*

I. Introduction

The need for considering discrete data is often encountered in data communications, digital signal processing, and other areas. In this paper we consider discrete valued random variables, and we are concerned with estimating the discrete probability density function. Measurements are taken, and from these measurements a density function is obtained. However, we assume that the measurements are imperfect. We derive the estimators, establish the appropriate forms of convergence, and supply an abundance of bounds on the errors.

Assume that we can observe $X_1, X_2, \ldots, X_n$, a sequence of independent identically distributed random variables with the unknown discrete probability density f. An obvious way of estimating $f(x)$ is to use the empirical density based on the n observations. However, the estimation problem is complicated if we can only observe X_1+Z_1, $X_2+Z_2, \ldots, X_n+Z_n$, where $X_1, Z_1, X_2, Z_2, \ldots, X_n, Z_n$ are independent random variables and the Z_i's, commonly referred to as noise, have a common known discrete probability density function g. For a wide class of densities g, a method is given to recover

* This research was supported by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant AFOSR-76-3062.

f which is shown to be strongly uniformly consistent, that is,

$$\lim_{n\to\infty} \sup_x |f_n(x)-f(x)|=0$$

with probability one (wp1), where f_n is the estimate of f with just n observations.

In signal detection, we observe $X_1+Z_1, X_2+Z_2, \ldots, X_n+Z_n$ if the signal is present, and $Z_1, Z_2, \ldots, Z_n$ if the signal is absent; and we are asked to decide whether the signal is present or not. A decision procedure is given with the property that, for all nontrivial f and for a wide class of discrete densities g, the probability of making a wrong decision tends to zero exponentially fast as n grows large.

In some applications we are asked to recover f when we observe $Z_1+X_1S_1, Z_2+X_2S_2, \ldots, Z_n+X_nS_n$, where $S_i=0$ if the signal is not present at the ith time instant, and $S_i=1$ if the signal is present. The random variables $S_1, S_2, \ldots, S_n$ are assumed to be independent and identically distributed. Again, g, the density of Z_1, and $\pi_1=P\{S_1=1\}$ are assumed to be known. For this case, the method to recover f generalizes the method that is used when $\pi_1=1$.

II. Properties of the Empirical Density

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random variables with a discrete probability density function f. Assume without loss of generality that f is supported on $\mathbb{Z}$, the set of integers. The empirical density f_n is defined by

$$f_n(x)=\frac{1}{n}\sum_{i=1}^{n} I_{\{X_i=x\}}, \quad x\in\mathbb{Z},$$

where I is the indicator function. Thus, f_n is also a discrete probability density.

We will briefly review some properties of f_n, starting with the pointwise consistency. By the strong law of large numbers [(**1**), p. 239] we know that $f_n(x)\to f(x)$ wpl for all x. In fact, by Hoeffding's inequality (**2**), for all $\varepsilon>0$,

$$P\{|f_n(x)-f(x)|\geq\varepsilon\}\leq 2\exp(-2n\varepsilon^2),$$

from which the strong consistency follows by the Borel–Cantelli lemma [(**1**), p. 228]. This bound is independent of f and x. Since countable unions of null events are null, we immediately have the strong uniform consistency

$$\lim_{n\to\infty} \sup_x |f_n(x)-f(x)|=0 \text{ wp1}.$$

Now we consider some uniform error bounds. If F_n and F are the distribution functions corresponding respectively to f_n and f, then

$$\begin{aligned} P\{\sup_x |f_n(x)-f(x)|\geq\varepsilon\} &\leq 2P\left\{\sup_x |F_n(x)-F(x)|\geq\frac{\varepsilon}{2}\right\} \\ &\leq 2C_1\exp\left[-2n\left(\frac{\varepsilon}{2}\right)^2\right], \end{aligned} \tag{1}$$

by an inequality of Dvoretzky *et al.* (**3**), where C_1 is a universal constant. In Appendix A it is shown that $C_1 < 610.4$. Recently, Singh (**4**) [see also, (**5**)] has shown that

$$P\left\{\sup_x |F_n(x) - F(x)| \geq \beta\right\} \leq 4e^2 n\beta \exp(-2n\beta^2)$$

for $n\beta^2 \geq 1$. This implies that

$$P\left\{\sup_x |f_n(x) - f(x)| \geq \varepsilon\right\} \leq 4e^2 n\varepsilon \exp\left(-n\frac{\varepsilon^2}{2}\right) \tag{2}$$

for $n\varepsilon^2 \geq 4$. Both bounds (1) and (2) are valid for *all* discrete densities f and, by the Borel–Cantelli lemma, each implies that

$$\lim_{n\to\infty} \sup_x |f_n(x) - f(x)| = 0 \text{ wp1}.$$

Now we consider the following measures of distance between f_n and f:

(i) $$\sum_{x=-\infty}^{\infty} |f_n(x) - f(x)|,$$

(ii) $$\left[\sum_{x=-\infty}^{\infty} |f_n(x) - f(x)|^p f(x)\right]^{1/p}, \quad p \geq 1,$$

(iii) $$\sup_x |f_n(x) - f(x)|,$$

and

(iv) $$\sup_{A \in \mathcal{P}(\mathbb{Z})} \left|\sum_{x \in A} f_n(x) - \sum_{x \in A} f(x)\right|,$$

where $\mathcal{P}(\mathbb{Z})$ is the power set of the integers. Notice that in (iii) the supremum is taken over all singleton sets while in (iv) the supremum is over all subsets of the integers. For (i) we have the following result.

Lemma 1

Let $f_n(x)$ be the empirical estimate of the discrete density $f(x)$. Then

$$P\left\{\sum_{x=-\infty}^{\infty} |f_n(x) - f(x)| \geq \varepsilon\right\} \leq K_1 \exp(-K_2 n) \tag{3}$$

where K_1, $K_2 > 0$ depend upon ε and f only.

Proof: Pick $N \geq 1$ such that

$$\sum_{|x|>N} f(x) < \frac{\varepsilon}{6}.$$

Then

$$P\left\{\sum_{x=-\infty}^{\infty} |f_n(x)-f(x)| \ge \varepsilon\right\}$$

$$\le P\left\{\sum_{|x|\le N} |f_n(x)-f(x)| \ge \frac{\varepsilon}{2}\right\} + P\left\{\sum_{|x|>N} |f_n(x)-f(x)| \ge \frac{\varepsilon}{2}\right\}$$

$$\le \sum_{|x|\le N} P\left\{|f_n(x)-f(x)| \ge \frac{\varepsilon}{4N+2}\right\} + P\left\{\sum_{|x|>N} f_n(x) - \sum_{|x|>N} f(x) \ge \frac{\varepsilon}{6}\right\}$$

$$\le (4N+2)\exp\left[-2n\left(\frac{\varepsilon}{4N+2}\right)^2\right] + \exp\left[-2n\left(\frac{\varepsilon}{6}\right)^2\right]$$

$$\le (4N+3)\exp\left[-2n\left(\frac{\varepsilon}{4N+2}\right)^2\right]$$

by Hoeffding's inequality (**2**). *Q.E.D.*

We see that a similar result is true for (iv) when we notice that

$$\sum_{x=-\infty}^{\infty} |f_n(x)-f(x)| = 2 \sup_{A\in\mathscr{P}(\mathbf{Z})} \left|\sum_{x\in A} f_n(x) - \sum_{x\in A} f(x)\right|. \tag{4}$$

The question remains whether an upper bound exists that decreases to zero as n grows large and does not depend upon f. The answer to this question is negative because for fixed n, it suffices to let

$$f(x) = \begin{cases} \dfrac{1}{2n}, & 1 \le x \le 2n \\ 0, & x<1 \quad \text{or} \quad x>2n. \end{cases}$$

Clearly, if $\varepsilon < \frac{1}{2}$, then

$$P\left\{\sup_{A\in\mathscr{P}(\mathbf{Z})} \left|\sum_{x\in A} f_n(x) - \sum_{x\in A} f(x)\right| \ge \varepsilon\right\} \ge P\left\{\frac{n}{2n} \ge \varepsilon\right\} = 1.$$

Distance (ii) has the disadvantage that it weighs $|f_n(x)-f(x)|$ less when $f(x)$ is small. For distribution-free results, (iii) seems the natural choice for the distance between f_n and f. For strong asymptotical results, but not distribution-free, (i) and (iv) can be used as well. The previous development gives explicit bounds relating to the distance measures.

Consider the following Lemma.

Lemma 2

Let $(\Omega, \mathscr{B}, P)$ be a probability space. Let $f_1, f_2, \ldots$ be densities on the integers for each fixed ω, and random variables on $(\Omega, \mathscr{B}, P)$ for each fixed x. We will write $f_n(x,\omega)$ to make the dependency on ω explicit. Let f be a density on the integers. If $f_n(x,\omega) \xrightarrow{n} f(x)$ wpl (in probability) for all x, then

$$\sum_{x=-\infty}^{\infty} |f_n(x,\omega)-f(x)| \xrightarrow{n} 0 \tag{5}$$

wpl (in probability).

The proof is given in Appendix B.

By Lemma 2 we have that

$$\lim_{n\to\infty} |f_n(x)-f(x)|=0$$

wpl (in probability) for $x\in\mathbb{Z}$ implies that

$$\lim_{n\to\infty}\sum_{x=-\infty}^{\infty} |f_n(x)-f(x)|=0$$

wpl (in probability). From (4) this is equivalent to

$$\lim_{n\to\infty}\sup_{A\in\mathcal{P}(\mathbb{Z})}\left|\sum_{x\in A} f_n(x)-\sum_{x\in A} f(x)\right|=0$$

wpl (in probability). This in turn implies that

$$\lim_{n\to\infty}\sup_{x}|f_n(x)-f(x)|=0$$

wpl (in probability). Since

$$\left[\sum_{x=-\infty}^{\infty}|f_n(x)-f(x)|^p f(x)\right]^{1/p}\le\sup_x|f_n(x)-f(x)|,$$

we then have that

$$\lim_{n\to\infty}\left[\sum_{x=-\infty}^{\infty}|f_n(x)-f(x)|^p f(x)\right]^{1/p}=0$$

wpl (in probability). Thus (5) implies that each of the four distance measures also converges to zero in the appropriate sense. Thus we conclude that for *any* sequence f_n of densities for which

$$\lim_{n\to\infty}|f_n(x)-f(x)|=0$$

wpl or in probability, for all $x\in\mathbb{Z}$, each of the four distance measures also converges to zero in the same sense.

III. Estimation in the Presence of Additive Noise

Because of background noise, faulty equipment, or other practical problems, it may not be possible to observe $X_1, X_2, \ldots, X_n$; but instead, we can observe $Y_1, Y_2, \ldots, Y_n$, where

$$Y_i=X_i+Z_i,\quad 1\le i\le n,$$

and $X_1, Z_1, X_2, Z_2, \ldots, X_n, Z_n$ are independent. The Z_i have a common known density g on the integers and the X_i have an unknown density f on the integers which we would like to estimate. The discrete probability density h of the Y_i is given by

$$h(x)=\sum_{y=-\infty}^{\infty} f(x-y)g(y).$$

Assume that we can write

$$f(x)=\sum_{y=-\infty}^{\infty} \xi_y h(x-y)=\sum_{y=-\infty}^{\infty} \xi_{x-y} h(y)$$

for some sequence $\{\xi_i\}$ of real numbers. Resubstitution gives

$$f(x)=\sum_{y=-\infty}^{\infty} \xi_y \sum_{u=-\infty}^{\infty} g(u)f(x-y-u) \tag{6}$$

which should hold for all x and all f.

Lemma 3

Equation (6) is valid for all x and all f if and only if

$$\sum_{y=-\infty}^{\infty} \xi_y g(k-y)=\begin{cases}1 & \text{for} \quad k=0\\ 0 & \text{for} \quad k\neq 0\end{cases}, \tag{7}$$

for all integers k.

Proof: Clearly, if (7) holds, then (6) is valid for all x and all f. Conversely, let $f(0)=1$, and note that

$$\sum_{y,u:y+u=0} \xi_y g(u)=1.$$

Next, let $f(0)=\frac{1}{2}=f(k)$, $k\neq 0$. Then (6) reads, for $x=0$,

$$\begin{aligned}\frac{1}{2}&=\frac{1}{2}\sum_{y,u:y+u=0} \xi_y g(u)+\frac{1}{2}\sum_{y,u:y+u=-k} \xi_y g(u)\\ &=\frac{1}{2}+\frac{1}{2}\sum_{y,u:y+u=-k} \xi_y g(u),\end{aligned}$$

from which (7) follows by the arbitrariness of k. *Q.E.D.*

Deferring for the moment the question of how to determine the ξ_y from g so that (7) is satisfied, we return to the construction of an estimate of f assuming the knowledge of the ξ_y. Let h_n be the empirical density for $Y_1, Y_2, \ldots, Y_n$,

$$h_n(x)=\frac{1}{n}\sum_{i=1}^{n} I_{\{Y_i=x\}},$$

which suggests the following estimate of f:

$$\begin{aligned}f_n(x)&=\sum_{y=-\infty}^{\infty} \xi_y h_n(x-y)\\ &=\sum_{y=-\infty}^{\infty} \xi_{x-y} h_n(y)\\ &=\frac{1}{n}\sum_{i=1}^{n}\sum_{y=-\infty}^{\infty} \xi_{x-y} I_{\{Y_i=y\}}\\ &=\frac{1}{n}\sum_{i=1}^{n} \xi_{x-Y_i}.\end{aligned} \tag{8}$$

Notice that for g with $g(0)=1$ (and thus $\xi_0=1$, $\xi_i=0$, $i\neq 0$), we get back the original empirical estimate of f because

$$\xi_{x-Y_i}=I_{\{Y_i=x\}}.$$

The first question that arises is the question of the closeness of f_n to f. Notice that f_n is not a probability density in general. Of course, $\tilde{f}_n$ defined by

$$\tilde{f}_n(x)=\begin{cases}1, & f_n(x)\geq 1\\ f_n(x), & 0<f_n(x)<1\\ 0, & f_n(x)\leq 0\end{cases}$$

is a strictly better density estimate than f_n. However, we will not further discuss this trivial modification of our estimate. Clearly, f_n satisfies

$$\sup_x |f_n(x)-f(x)|\leq\left(\sum_{y=-\infty}^{\infty}|\xi_y|\right)\sup_x|h_n(x)-h(x)|, \tag{9}$$

$$\sup_x |f_n(x)-f(x)|\leq\left(\sup_y|\xi_y|\right)\sum_{x=-\infty}^{\infty}|h_n(x)-h(x)|, \tag{10}$$

and

$$\sum_{x=-\infty}^{\infty}|f_n(x)-f(x)|\leq\left(\sum_{y=-\infty}^{\infty}|\xi_y|\right)\sum_{x=-\infty}^{\infty}|h_n(x)-h(x)|.$$

Let

$$C=\sum_{y=-\infty}^{\infty}|\xi_y|$$

and

$$D=\sup_y|\xi_y|.$$

Applying some of the results of the previous section, we have the following theorem.

Theorem I

Let $f_n(x)$ be given by (8) and assume that $\{\xi_y\}$ satisfies (7). Then if C is finite, we have

$$P\{\sup_x|f_n(x)-f(x)|\geq\varepsilon\}\leq\begin{cases}2C_1\exp\left[-\dfrac{n\varepsilon^2}{2C^2}\right], & \text{all } n\geq 1\\ \dfrac{4e^2n\varepsilon}{C}\exp\left[-\dfrac{n\varepsilon^2}{2C^2}\right], & \text{all } \quad n\geq\dfrac{4C^2}{\varepsilon^2}\end{cases} \tag{11}$$

and

$$P\left\{\sum_{x=-\infty}^{\infty}|f_n(x)-f(x)|\geq\varepsilon\right\}\leq K_1\exp(-K_2n) \tag{12}$$

where K_1, $K_2>0$ are constants depending upon ε, h, and C. Also, if D is finite, we have

$$P\{\sup_x |f_n(x)-f(x)|\geq \varepsilon\}\leq K_3 \exp(-K_4 n) \tag{13}$$

where K_3, $K_4>0$ are constants depending upon ε, h, and D.

The bounds in (11) are distribution-free but require that $C<\infty$. Eq. (13), which is not distribution-free, requires only that $D<\infty$. The strong result (12) assumes finiteness of C and is not uniform over all densities as we might expect from the remarks of the previous section.

Using the Borel–Cantelli lemma, we obtain the following result.

Corollary 1

Let $f_n(x)$ be given by (8) and assume that $\{\xi_y\}$ satisfies (7). If $D=\sup_y |\xi_y|<\infty$, then

$$\sum_{x=-\infty}^{\infty} |f_n(x)-f(x)| \xrightarrow{n} 0 \text{ wpl.}$$

In the remainder of this section we briefly discuss practical solutions to (7) and give some examples of sequences $\{\xi_y\}$ for some common densities g. The problem of the finiteness of C or D is briefly considered.

Practical Considerations

A solution to (7) can be obtained recursively if g is a single tailed density, that is, if there exists a K such that $g(x)=0$ for all $x>K$ or $g(x)=0$ for all $x<K$. For example, assume that $g(K)>0$ and $g(x)=0$ for all $x<K$. Let $\xi_y=0$ for $y<-K$ and $\xi_{-K}=1/g(K)$. It is easy to see that the $k\leq 0$ equations of (7) hold and that the $k=1$ equation results in

$$\xi_{-K+1}g(K)+\xi_{-K}g(K+1)=0,$$

from which we find ξ_{-K+1}. Solving the $k=2$ equation of (7) gives us ξ_{-K+2} and so on. Clearly, this is probably not the only solution to (7). Consider the following simple example.

Example 1

Let $g(0)=g(1)=\frac{1}{2}$ and $g(x)=0, x\neq 0, 1$. Then (7) results in

$$\xi_{-1}+\xi_0=2$$

$$\ldots=\xi_{-2}+\xi_{-1}=\xi_0+\xi_1=\xi_1+\xi_2=\ldots=0.$$

If $\xi_0=\alpha$ and $\xi_{-1}=2-\alpha$, then all solutions of (7) can be written as

$$\xi_y=\alpha(-1)^y,\quad y\geq 0$$

$$\xi_{-y}=(2-\alpha)(-1)^{y+1},\quad y\geq 1$$

where α is any real number. For this case we note that D is finite while C is infinite.

Example 2

In Example 1 we have $\sup_y |\xi_y| < \infty$, but this is not always the case. If $g(-1) = g(1) = 1/4$ and $g(0) = 1/2$, then it is straight-forward to show that for any solution of (7) we have $\sup_y |\xi_y| = \infty$.

Now we will give solutions for some well known densities g.

Example 3. *Poisson Noise with Parameter* $\lambda > 0$

Let

$$g(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x \geq 0,$$

and it can be verified that a solution to (7) is given by

$$\xi_y = \begin{cases} \dfrac{(-\lambda)^y}{y!} e^{\lambda}, & y \geq 0 \\ 0, & y < 0 \end{cases}.$$

It is easily seen that

$$\sum_{y=-\infty}^{\infty} |\xi_y| = e^{2\lambda} < \infty.$$

In (6) the estimation of a continuous probability density function from measurements corrupted by Poisson noise is considered.

Example 4. *Geometric Noise with Parameter* $\lambda > 1$

If

$$g(x) = (\lambda - 1)/\lambda^{x+1}, \quad x \geq 0,$$

then a solution to (7) is given by

$$\xi_y = \begin{cases} \lambda/(\lambda - 1), & y = 0, \\ -1/(\lambda - 1), & y = 1, \\ 0, & y \neq 0, 1. \end{cases}$$

Example 5. *Binomial Noise with Parameters N and* $p \neq 0, 1$

If

$$g(x) = \binom{N}{x} p^x (1-p)^{N-x}, \quad 0 \leq x \leq N,$$

then a solution to (7) is given by

$$\xi_y = \begin{cases} (-1)^y \dbinom{N+y-1}{y} (1-p)^{-N} \left(\dfrac{p}{1-p}\right)^y, & y \geq 0. \\ 0, & y < 0. \end{cases}$$

We notice that

$$\sum_{y=-\infty}^{\infty} |\xi_y| < \infty$$

if and only if $p<1/2$, and that $|\xi_y| \xrightarrow{y} \infty$ if $p>1/2$ or if $N>2$ and $p=1/2$. Another solution to (7) is given by

$$\xi_{-N-y} = \begin{cases} (-1)^y \binom{N+y-1}{y} p^{-N} \left(\frac{1-p}{p}\right)^y, & y \geq 0 \\ 0, & y<0 \end{cases}$$

and it is easy to see that for each $p>1/2$,

$$\sum_{y=-\infty}^{\infty} |\xi_y| < \infty.$$

To be able to use the results of Theorem I, we have to know whether

$\sup_y |\xi_y| < \infty$ or

$$\sum_{y=-\infty}^{\infty} |\xi_y| < \infty.$$

Can this be done without actually determining the solution to (7)? The answer to this is positive in many cases of practical interest. For example, let $g(x)=0$ for $x \leq 0$ and $x>M$, and assume that $g(1)>0$ and $g(M)>0$. Then (7) reduces to

$$\sum_{y=1}^{M} g(y)\xi_{k-y} = \begin{cases} 1, & k=0 \\ 0, & k \neq 0. \end{cases}$$

Let a prime denote the transpose. Let

$$\xi(k) = [\xi_{k-M}, \ldots, \xi_{k-1}]'.$$

Now consider the equation

$$\xi(k+1) = A\xi(k), \quad k \geq 0 \tag{14}$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ & & \cdots & . & & \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & -\frac{g(M)}{g(1)} & -\frac{g(M-1)}{g(1)} & \cdots & -\frac{g(3)}{g(1)} & -\frac{g(2)}{g(1)} \end{bmatrix} \tag{15}$$

and

$$\xi(0) = \left[0, 0, \ldots, 0, \frac{1}{g(1)}\right]'.$$

The solution $\{\xi_y\}$ we are considering has $\xi_y = 0$ for $y<-1$, and ξ_y is determined by (14) for $y \geq -1$. Define the norm of $V = [v_1, \ldots, v_M]'$ as

$$\|V\| = \sum_{i=1}^{M} |v_i|.$$

Notice that

$$\sum_{k=0}^{\infty} \|\xi(k)\| < \infty$$

if and only if

$$C = \sum_{y=-\infty}^{\infty} |\xi_y| < \infty.$$

Also, $\sup_k \|\xi(k)\| < \infty$ if and only if $D = \sup_y |\xi_y| < \infty$.

Let e_i denote an M dimensional vector with a one in the ith position and zeros elsewhere. Then we have the following result.

Lemma 4

Let A be given by (15). Then

$$\sum_{n=0}^{\infty} \|A^n V\| < \infty$$

for all $V \in \mathbb{R}^M$ if and only if

$$\sum_{n=0}^{\infty} \|A^n e_M\| < \infty.$$

Proof: The sufficiency is obvious. For necessity, notice that we need only show that

$$\sum_{n=0}^{\infty} \|A^n e_i\| < \infty$$

for all $1 \leq i < M$. For e_1 we have

$$\sum_{n=0}^{\infty} \|A^n e_1\| = 1.$$

For $1 < i < M$, we have

$$\begin{aligned}\sum_{n=0}^{\infty} \|A^n e_i\| &= \sum_{n=1}^{\infty} \|A^{n-1} A e_i\| + 1 \\ &\leq \sum_{n=1}^{\infty} \|A^{n-1} e_{i-1}\| + \frac{g(M+2-i)}{g(1)} \sum_{n=1}^{\infty} \|A^{n-1} e_M\| + 1 \\ &\leq \sum_{n=0}^{\infty} \|A^n e_{i-1}\| + \frac{g(M+2-i)}{g(1)} \sum_{n=0}^{\infty} \|A^n e_M\| + 1.\end{aligned}$$

Thus the result follows by induction. *Q.E.D.*

From Lemma 4 and the stability theory of discrete dynamical systems [see, for example, (7)], we have the following result.

Lemma 5

Let $\{\xi_y\}$ be defined through (14). Then

$$C = \sum_{y=-\infty}^{\infty} |\xi_y| < \infty$$

if and only if all the eigenvalues of the matrix A have magnitudes less than one. A sufficient condition for

$$D = \sup_y |\xi_y| < \infty$$

is that all the eigenvalues of A have magnitudes less than or equal to one, and any eigenvalue with magnitude one is a simple zero of the minimal polynomial of A (i.e. corresponds to a Jordan block of order one in the Jordan canonical form of A).

We can also construct a solution in the above situation with $\xi_{-M} = 1/g(M)$ and $\xi_y = 0$ for $y > -M$. Then we can use the above type of argument to establish similar conditions on the solution.

IV. Signal Detection

Assume that we observe $Y_1, Y_2, \ldots, Y_n$, where either $Y_i = X_i + Z_i$ or $Y_i = Z_i$. We assume that $X_1, \ldots, X_n$ are identically distributed and $Z_1, \ldots, Z_n$ are identically distributed. Also, $X_1, Z_1, \ldots, X_n, Z_n$ are independent. As in the last section, we assume that X_1 has an unknown probability density f on the integers and that Z_1 has a probability density g on the integers. In engineering applications, the X_i are often identified with the signal and the Z_i are considered as noise. The goal is to devise a machine to detect whether we are observing a signal with noise superimposed on it, or just noise. We say that f defines a signal if $f(0) \neq 1$.

Assume that a solution $\{\xi_y\}$ to (7) exists for the density g. Now consider the following detection procedure. Let $f_n(0)$ be the estimate of $f(0)$ given by

$$f_n(0) = \frac{1}{n} \sum_{i=1}^{n} \xi_{-Y_i}.$$

Let $\{\varepsilon_n\}$ be a sequence of positive numbers and define the random variable D_n by

$$D_n = \begin{cases} 1 & \text{if} \quad f_n(0) < 1 - \varepsilon_n \\ 0 & \text{if} \quad f_n(0) \geq 1 - \varepsilon_n \end{cases}.$$

If $D_n = 1$, we decide that the signal is present, and if $D_n = 0$, we decide that it is not present. Let E_n be the indicator of an error, that is,

$$E_n = \begin{cases} 1 \text{ if } D_n = 1 \text{ and the signal is not present} \\ 1 \text{ if } D_n = 0 \text{ and the signal is present} \\ 0 \text{ otherwise.} \end{cases}$$

Then we have the following result.

Theorem II

Assume that

$$C = \sum_{y=-\infty}^{\infty} |\xi_y| < \infty$$

where $\{\xi_y\}$ is a solution to (7) for the density g. Let $\{\varepsilon_n\}$ be a sequence of positive numbers such that, as $n \to \infty$,

$$\varepsilon_n \to 0$$

and

$$n\varepsilon_n^2 \to \infty.$$

If $f(0) \neq 1$, then

$$\lim_{n\to\infty} E_n = 0 \text{ in probability}$$

under both hypotheses. If in addition

$$\sum_{n=1}^{\infty} \exp\left[-\frac{n\varepsilon_n^2}{8C^2}\right] < \infty, \tag{16}$$

then

$$\lim_{n\to\infty} E_n = 0 \text{ wpl}$$

under both hypotheses.

Proof: Let N be such that for all $n \geq N$,

$$\varepsilon_n < \frac{1-f(0)}{4}.$$

Now consider $n \geq N$. Let h be the density of $X_1 + Z_1$, and let $\hat{h}$ be the density of Y_1, i.e. $\hat{h} = h$ if $Y_1 = X_1 + Z_1$ and $\hat{h} = g$ if $Y_1 = Z_1$. Let $\hat{h}_n$ be the empirical density for $\hat{h}$. From (9) it follows that if

$$\sup_x |\hat{h}_n(x) - \hat{h}(x)| < \frac{\varepsilon_n}{2C},$$

then

(i) $|f_n(0) - 1| < \frac{\varepsilon_n}{2}$ (and thus $D_n = 0$) if the signal is absent, and

(ii) $|f_n(0) - f(0)| < \frac{\varepsilon_n}{2}$ (and thus $D_n = 1$) if the signal is present.

Thus

$$\begin{aligned} P\{E_n = 1\} &\leq P\left\{\sup_x |\hat{h}_n(x) - \hat{h}(x)| \geq \frac{\varepsilon_n}{2C}\right\} \\ &\leq 2C_1 \exp\left[-\frac{n\varepsilon_n^2}{8C^2}\right] \end{aligned}$$

by (11), regardless of whether $\hat{h} = h$ or $\hat{h} = g$. Therefore, $E_n \to 0$ in probability. The convergence with probability one follows from the Borel–Cantelli lemma. *Q.E.D.*

Thus the described signal detection procedure is asymptotically error-free for a large class of discrete densities g. The theorem applies to all probability densities f with $f(0) \neq 1$. Notice that

$$\lim_{n\to\infty} \frac{n\varepsilon_n^2}{\log n} = \infty$$

is sufficient for (16) regardless of C. Thus, for example, $\varepsilon_n = n^{-\alpha}$ for $0 < \alpha < 1/2$ satisfies the conditions of Theorem II.

V. Recovery of a Density when Some Measurements are Incorrect

We now assume that $Y_1, Y_2, \ldots, Y_n$ is a sequence of independent identically distributed integer-valued random variables with the discrete density $\pi_1 f + \pi_2 g$ where $\pi_1 + \pi_2 = 1$, $\pi_1 > 0$, $\pi_2 > 0$, π_1, π_2, and g are known, and the density f is unknown. Thus, roughly speaking, a portion $\pi_2 n$ of $Y_1, Y_2, \ldots, Y_n$ have the known density g. This situation may occur when a machine or human collecting the observations fails part of the time.

We assume the knowledge of π_1 and π_2 because in general they cannot be estimated from the data unless $f \neq g$, which we do not know. If h is the density of Y_1 and

$$h_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{Y_i = x\}}$$

is the empirical density estimate of $h(x)$, then we can estimate

$$f(x) = \frac{h(x)}{\pi_1} - \frac{\pi_2}{\pi_1} g(x)$$

by

$$f_n(x) = \frac{h_n(x)}{\pi_1} - \frac{\pi_2}{\pi_1} g(x). \tag{17}$$

Since

$$f_n(x) - f(x) = \frac{1}{\pi_1}[h_n(x) - h(x)],$$

we have the following result.

Theorem III

Let $f_n(x)$ be defined by (17). Then

$$P\left\{\sup_x |f_n(x) - f(x)| \geq \varepsilon\right\} \leq \begin{cases} 2C_1 \exp\left[-\dfrac{n\pi_1^2\varepsilon^2}{2}\right], & \text{all } n \geq 1 \\ 4e^2 n\pi_1 \varepsilon \exp\left[-\dfrac{n\pi_1^2\varepsilon^2}{2}\right], & \text{all } n \geq \dfrac{4}{\pi_1^2\varepsilon^2}. \end{cases}$$

Using the Borel–Cantelli lemma and Lemma 1, we have the following Corollary.

Corollary 2

Let $f_n(x)$ be defined by (17). Then

$$\lim_{n\to\infty} \sum_{x=-\infty}^{\infty} |f_n(x)-f(x)|=0 \quad \text{wpl.}$$

Now consider the slightly more complicated model in which we observe $Y_i = X_iS_i + Z_i$, where $\{X_1, Z_1, S_1, \ldots, X_n, Z_n, S_n\}$ are independent integer valued random variables where

(i) the Z_i have a known density g,

(ii) the S_i are $\{0, 1\}$-valued and identically distributed with $\pi_1 = P\{S_1 = 1\} > 0$ known,

(iii) the X_i have a common unknown density f to be estimated.

The model of the data is different from that in the signal detection problem in that now a signal is sometimes present ($S_i = 1$, $Y_i = X_i + Z_i$) and sometimes absent ($S_i = 0$, $Y_i = Z_i$). In the previous part of this section we desired to estimate a mixture component of a density; and now we desire to estimate a convolution component of a mixture component of a density, since the density of Y_1 is $\pi_1(f*g)+\pi_2 g$, where $*$ denotes the convolution operation.

We construct our estimate of f in two steps. First, let h_n be the empirical estimate of h with $Y_1, \ldots, Y_n$:

$$h_n(x)=\frac{1}{n}\sum_{i=1}^{n} I_{\{Y_i=x\}};$$

and let

$$\tilde{h}_n(x)=\frac{h_n(x)}{\pi_1}-\frac{\pi_2}{\pi_1}g(x)$$

be our estimate of $(f*g)$ (x). Following the reasoning of Section III, we then try to recover f by deconvolving as follows, where $\{\xi_y\}$ is a solution of (7) for the density g:

$$\begin{aligned} f_n(x) &= \sum_{y=-\infty}^{\infty} \xi_y \tilde{h}_n(x-y) \\ &= \frac{1}{\pi_1}\left(\frac{1}{n}\sum_{i=1}^{n} \xi_{x-Y_i}\right)-\frac{\pi_2}{\pi_1}\sum_{y=-\infty}^{\infty} \xi_{x-y} g(y). \end{aligned} \tag{18}$$

Notice that

$$f_n(x)-f(x)=\frac{1}{\pi_1}\sum_{y=-\infty}^{\infty} \xi_{x-y}[h_n(y)-h(y)].$$

Thus we have

$$\sup_x |f_n(x)-f(x)| \leq \frac{1}{\pi_1}\left(\sum_{y=-\infty}^{\infty} |\xi_y|\right) \sup_x |h_n(x)-h(x)|$$

and

$$\sup_x |f_n(x)-f(x)| \le \frac{1}{\pi_1} \sup_y |\xi_y| \sum_{x=-\infty}^{\infty} |h_n(x)-h(x)|,$$

which are similar to (9) and (10), respectively. Thus, from Theorem I, we have the following result.

Theorem IV

Let $\{\xi_y\}$ be a solution to (7) for the density g, and let $f_n(x)$ be given by (18). Then if

$$C = \sum_{y=-\infty}^{\infty} |\xi_y| < \infty,$$

we have

$$P\left\{\sup_x |f_n(x)-f(x)| \ge \varepsilon\right\} \le \begin{cases} 2C_1 \exp\left[-\dfrac{n\pi_1^2\varepsilon^2}{2C^2}\right], & \text{all } n \ge 1 \\ \dfrac{4e^2 n\pi_1\varepsilon}{C} \exp\left[-\dfrac{n\pi_1^2\varepsilon^2}{2C^2}\right], & \text{all } n \ge \dfrac{4C^2}{\pi_1^2\varepsilon^2} \end{cases}$$

and

$$P\left\{\sum_{x=-\infty}^{\infty} |f_n(x)-f(x)| \ge \varepsilon\right\} \le K_1 \exp(-K_2 n)$$

where K_1, $K_2 > 0$ are constants depending upon ε, h, π_1, and C. Also, if $D = \sup_y |\xi_y| < \infty$, we have

$$P\{\sup_x |f_n(x)-f(x)| \ge \varepsilon\} \le K_3 \exp(-K_4 n)$$

where K_3, $K_4 > 0$ are constants depending upon ε, h, π_1, and D.

Using the Borel–Cantelli lemma, we have the following Corollary.

Corollary 3

Let $\{\xi_y\}$ be a solution to (7) for the density g, and let $f_n(x)$ be given by (18). If $D = \sup_y |\xi_y| < \infty$, then

$$\lim_{n\to\infty} \sum_{x=-\infty}^{\infty} |f_n(x)-f(x)| = 0 \quad \text{wpl.}$$

References

(1) M. Loève, "Probability Theory", Van Nostrand, New York, 1963.
(2) W. Hoeffding, "Probability inequalities for sums of bounded random variables", *J. Am. Statist. Ass.*, Vol. 58, pp. 13–30, 1963.
(3) A. Dvoretzky, J. Kiefer and J. Wolfowitz, "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator", *Ann. Math. Statist.*, Vol. 27, pp. 642–669, 1956.

(4) R. S. Singh, "On the rate for uniform strong consistency of empirical distributions of independent nonidentically distributed multivariate random variables", *J. Multivariate Analysis*, Vol. 6, pp. 338–342, 1976.
(5) L. P. Devroye, "A bound for the uniform deviation of empirical distribution functions", *J. Multivariate Analysis*, Vol. 7, pp. 594–597, 1977.
(6) G. L. Wise, A. P. Traganitis and J. B. Thomas, "The estimation of a probability density function from measurements corrupted by Poisson noise", *IEEE Trans. Inform. Theory*, Vol. IT-23, pp. 764–766, 1977.
(7) J. L. Willems, "*Stability Theory of Dynamical Systems*", pp. 173–174, John Wiley, New York, 1970.
(8) W. Feller, "An Introduction to Probability Theory and Its Applications", Vol. I., p. 54, John Wiley, New York, 1968.
(9) H. L. Royden, "Real Analysis," p. 270, Macmillan, London, 1968.

Appendix A

In this appendix we prove the following Lemma. The bound that we present results directly from the work of Dvoretzky *et al.*

Lemma

If F is any distribution function on $\mathbb{R}$ and F_n is the empirical distribution function with $X_1, X_2, \ldots, X_n$, a sequence of independent random variables with distribution function F, then

$$P\{\sup_x |F_n(x)-F(x)| \geq \varepsilon\} \leq C_1 \exp(-2n\varepsilon^2)$$

where

$$C_1 = 2\left[1+\frac{32}{\sqrt{(6\pi)}}+\frac{8}{\sqrt{3}}+\frac{4\sqrt{2}\exp(40/9)}{\sqrt{e}}\right].$$

Proof: We will use the notation of Dvoretzky *et al.* [(3), pp. 646–648]. In (3) they establish that

$$1-G_n(r) \leq 2[1-H_n(r)]$$

where r takes values in $(0, \sqrt{n})$. Expression (2.9),

$$1-H_n(r) = \left(1-\frac{r}{\sqrt{n}}\right)^n + r\sqrt{n}\sum_{j=[r\sqrt{n}]+1}^{n-1} Q_n(j, r),$$

can be upper bounded as follows. Notice that

$$\left(1-\frac{r}{\sqrt{n}}\right)^n < \exp(-2r^2)$$

(see (2.11)).

First, consider those j for which $\left|j-\frac{n}{2}\right| \leq \frac{n}{4}$.

We will show that

$$Q_n(j, 0) < c_2 n^{-\frac{3}{2}}$$

for $c_2 = 16/\sqrt{(6\pi)}$. By an approximation of Feller (8) for $n!$ we have

$$Q_n(j, 0) = \binom{n}{j}\frac{j^j}{n^n}(n-j)^{n-j-1}$$

$$\leq \frac{1}{\sqrt{(2\pi)}}\sqrt{\frac{n}{j(n-j)^3}}\exp\left[\frac{1}{12n} - \frac{1}{12j+1} - \frac{1}{12(n-j)+1}\right].$$

Notice that $j(n-j)^3 \geq 3(n/4)^4$. Also,

$$\exp\left[\frac{1}{12n} - \frac{1}{12j+1} - \frac{1}{12(n-j)+1}\right] < \exp\left[\frac{-1}{12j+1}\right] < 1.$$

Thus,

$$Q_n(j, 0) < \frac{16}{\sqrt{(6\pi)}}n^{-\frac{3}{2}}.$$

Next, we know that

$$\frac{1}{n} + \int_0^\infty \exp(-8r^2t^2)\,dt < 1 + \frac{\sqrt{(2\pi)}}{4}$$

and (2.15) holds:

$$r\sqrt{n}\sum{}' Q_n(j, r) < 2c_2\left(1 + \frac{\sqrt{(2\pi)}}{4}\right)\exp(-2r^2)$$

$$= \left(\frac{32}{\sqrt{(6\pi)}} + \frac{8}{\sqrt{3}}\right)\exp(-2r^2).$$

Now consider those j for which $|j - n/2| > n/4$. It follows from the equation at the top of p. 647 that

$$Q_n(j, r) \leq Q_n(j, 1)\exp[-f(n, r, j) + f(n, 1, j)]$$

for $r \geq 1$, where $f(n, r, j)$ is the negative of the exponent in (2.12). From (2.9) it follows that $j \geq r\sqrt{n}$. Thus, it is easily seen that $f(n, 1, j) \leq 40/9$. Define $c = \exp(40/9)$. Therefore, we have that

$$Q_n(j, r) \leq cQ_n(j, 1)\exp[-f(n, r, j)]$$

for $r \geq 1$. Since $f(n, r, j) \geq 2r^2 + r^2/64$, we have that

$$Q_n(j, r) \leq cQ_n(j, 1)\exp(-2r^2)\exp\left(-\frac{r^2}{64}\right).$$

Using the fact that $r\exp(-r^2/64)$ is maximized at $r = 4\sqrt{2}$, we obtain

$$Q_n(j, r) \leq \frac{4c\sqrt{2}}{r\sqrt{e}}Q_n(j, 1)\exp(-2r^2).$$

Thus,

$$r\sqrt{n}\sum{}'' Q_n(j, r) \leq \frac{4c\sqrt{2}}{\sqrt{e}}\sqrt{n}\exp(-2r^2)\sum{}'' Q_n(j, 1)$$

$$< \frac{4c\sqrt{2}}{\sqrt{e}}\exp(-2r^2)$$

from (2.16):

Collecting bounds, we find that, for $r \geq 1$,

$$1 - G_n(r) \leq 2\left[1 + \frac{32}{\sqrt{(6\pi)}} + \frac{8}{\sqrt{3}} + \frac{4\sqrt{2}\exp(40/9)}{\sqrt{e}}\right]\exp(-2r^2). \tag{A1}$$

For $r < 1$, the expression on the right hand side is greater than one, so that (A1) is valid for all r. *Q.E.D.*

Appendix B

In this appendix we present the proof of Lemma 2. For the first part, notice that

$$P\{\omega: \overline{\lim_{n\to\infty}} \sum_{x=-\infty}^{\infty} |f_n(x,\omega) - f(x)| \neq 0\} \leq$$
$$P\{\omega: \sum_{x=-\infty}^{\infty} \overline{\lim_{n\to\infty}} |f_n(x,\omega) - f(x)| \neq 0\} =$$
$$P\left\{\bigcup_{x=-\infty}^{\infty}\left(\overline{\lim_{n\to\infty}} |f_n(x,\omega) - f(x)| \neq 0\right)\right\} \leq$$
$$\sum_{x=-\infty}^{\infty} P\{\omega: \overline{\lim_{n\to\infty}} |f_n(x,\omega) - f(x)| \neq 0\}.$$

Therefore, if $f_n(x,\omega) \to f(x)$ wpl, then

$$\sum_{x=-\infty}^{\infty} |f_n(x,\omega) - f(x)| \to 0 \quad \text{wpl.}$$

For the second part, let U^+ denote $\max\{0, U\}$. Since f_n and f are densities, we have

$$\sum_{x=-\infty}^{\infty} |f_n(x,\omega) - f(x)| = 2\sum_{x=-\infty}^{\infty} [f(x) - f_n(x,\omega)]^+ \leq 2$$

for all n and ω. Thus

$$\sum_{x=-\infty}^{\infty} |f_n(x,\omega) - f(x)| \to 0$$

in probability if and only if [(**1**), p. 158]

$$\int_\Omega \sum_{x=-\infty}^{\infty} [f(x) - f_n(x,\omega)]^+ P(d\omega) \to 0,$$

and by Tonelli's theorem (**9**) this is equivalent to

$$\sum_{x=-\infty}^{\infty} \int_\Omega [f(x) - f_n(x,\omega)]^+ P(d\omega) \to 0.$$

But for every x

$$\int_\Omega [f(x) - f_n(x,\omega)]^+ P(d\omega) \to 0$$

since $0 \leq [f(x) - f_n(x, \omega)]^+ \leq f(x)$ and $f_n(x, \omega) \rightarrow f(x)$ in probability [(**1**), p. 125]. Also,

$$\sum_{x=-\infty}^{\infty} \int_{\Omega} [f(x) - f_n(x, \omega]^+ P(\mathrm{d}\omega) \leq \sum_{x=-\infty}^{\infty} f(x) = 1;$$

so that by the dominated convergence theorem [(**1**), p. 125],

$$\sum_{x=-\infty}^{\infty} |f_n(x, \omega) - f(x)| \rightarrow 0$$

in probability. *Q.E.D.*