ON THE ESTIMATION OF DISCRETE PROBABILITY DENSITIES FROM NOISY MEASUREMENTS

LUC P. DEVROYE
School of Computer Science
McGill University
Montreal, Canada H3C 3G1

GARY L. WISE
Department of Electrical Engineering
University of Texas at Austin
Austin, Texas 78712

ABSTRACT

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 independent additive noise. This method is shown to be consistent, and several bounds on the error are given.

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 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.

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 point-wise consistency. By the strong law of large numbers [1, p.239] we know that $f_n(x) \to f(x)$ wp1 for all x. In fact, by Hoeffding's inequality [2], for all $\varepsilon > 0$,

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

Presented at the Fifteenth Annual Allerton Conference on Communication, Control, and Computing, September 28-30, 1977; to be published in the Proceedings of the Conference.

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

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

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

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

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

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

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

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

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

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

Now consider the following two lemmas.

<u>Lemma 1</u>: Let $f_n(x)$ be the empirical estimate of the discrete density f(x). Then

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

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

$$P\left\{\sum_{x=-\infty}^{\infty} |f_n(x) - f(x)| \geq \varepsilon\right\}$$
$$\leq P\left\{\sum_{|x|\leq N} |f_n(x) - f(x)| \geq \frac{\varepsilon}{2}\right\} + P\left\{\sum_{|x|>N} |f_n(x) - f(x)| \geq \frac{\varepsilon}{2}\right\}$$
$$\leq \sum_{|x|\leq N} P\left\{|f_n(x) - f(x)| \geq \frac{\varepsilon}{4N+2}\right\} + P\left\{\sum_{|x|>N} f_n(x) - \sum_{|x|>N} f(x) \geq \frac{\varepsilon}{6}\right\}$$

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

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

by Hoeffding's inequality [2]. Q.E.D.

Lemma 2: Let $(\Omega,\mathcal{B},P)$ be a probability space. Let $f_1, f_2, \ldots$ be densities on the integers for each fixed ω, and random variables on $(\Omega,\mathcal{B},P)$ 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 $f_n(x,\omega) \overset{n}{\to} f(x)$ wp1 for all x, then

$$\sum_{x=-\infty}^{\infty} |f_n(x,\omega) - f(x)| \overset{n}{\to} 0 \quad \text{wp1.} \tag{3}$$

Proof:

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

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

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

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

Fatou: $\overline{\lim} \sum_x |f_n(x,\omega) - f(x)| \leq \sum_x \overline{\lim} |f_n(x,\omega) - f(x)|$.

Therefore, if $f_n(x,\omega) \to f(x)$ wp1, then

$$\sum_{x=-\infty}^{\infty} |f_n(x,\omega) - f(x)| \to 0 \quad \text{wp1.}$$ Q.E.D.

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

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

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

$$f(x) = \sum_{y=-\infty}^{\infty} \xi_y \sum_{u=-\infty}^{\infty} g(u)\, f(x-y-u) \tag{4}$$

which should hold for all x and all f.

Lemma 3: Eq. (4) is valid for all x and all f if and only if

$$\sum_{y=-\infty}^{\infty} \xi_y\, g(k-y) = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } k \neq 0 \end{cases}, \tag{5}$$

for all integers k.

Proof: Clearly, if (5) holds, then (4) 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 (4) reads, for x = 0,

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

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

from which (5) 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 (5) 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:

$$f_n(x) = \sum_{y=-\infty}^{\infty} \xi_y\, h_n(x-y) = \sum_{y=-\infty}^{\infty} \xi_{x-y}\, h_n(y)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{y=-\infty}^{\infty} \xi_{x-y}\, I_{\{Y_i = y\}} = \frac{1}{n} \sum_{i=1}^{n} \xi_{x-Y_i} . \tag{6}$$

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

$$\sup_x |f_n(x) - f(x)| \leq \left(\sum_{y=-\infty}^{\infty} |\xi_y| \right) \sup_x |h_n(x) - h(x)| \quad , \tag{7}$$

$$\sup_x |f_n(x) - f(x)| \leq \left(\sup_y |\xi_y| \right) \sum_{x=-\infty}^{\infty} |h_n(x) - h(x)| \quad , \tag{8}$$

and

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

Let

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

and

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

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

<u>Theorem 1</u>: Let $f_n(x)$ be given by (6) and assume that $\{\xi_y\}$ satisfies (5). Then if C is finite, we have

$$P\left\{ \sup_x |f_n(x) - f(x)| \geq \varepsilon \right\} \leq \begin{cases} 2\,C_1 \exp\left[-\dfrac{n\,\varepsilon^2}{2\,C^2} \right], & \text{all } n \geq 1 \\ \dfrac{4e^2 n\varepsilon}{C} \exp\left[-\dfrac{n\,\varepsilon^2}{2\,C^2} \right], & \text{all } n \geq \dfrac{4C^2}{\varepsilon^2} \end{cases} \tag{9}$$

and

$$P\left\{ \sum_{x=-\infty}^{\infty} |f_n(x) - f(x)| \geq \varepsilon \right\} \leq K_1 \exp(-K_2 n) \tag{10}$$

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

$$P\left\{\sup_x |f_n(x) - f(x)| \geq \varepsilon\right\} \leq K_3 \exp(-K_4 n) \tag{11}$$

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

The bounds in (9) are distribution-free but require that $C < \infty$. Eq. (11), which is not distribution-free requires only that $D < \infty$. The strong result (10) assumes finiteness of C and is not uniform over all densities.

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

Corollary 1: Let $f_n(x)$ be given by (6) and assume that $\{\xi_y\}$ satisfies (5). If $D = \sup_y |\xi_y| < \infty$, then

$$\sum_{x=-\infty}^{\infty} |f_n(x) - f(x)| \xrightarrow{n} 0 \quad \text{wp1.}$$

In the remainder of this paper we briefly discuss practical solutions to (5) and give some examples of sequences $\{\xi_y\}$ for some common densities g.

Practical Considerations: A solution to (5) 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 $k = 0$ equation of (5) holds 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 (5) gives us ξ_{-K+2} and so on. Clearly, this is probably not the only solution to (5). Consider the following simple example.

Example 1: Let $g(0) = g(1) = 1/2$ and $g(x) = 0$, $x \neq 0, 1$. Then (5) 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 (5) 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 straightforward to show that for any solution of (5) 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 (5) is given by

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

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 (5) 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 (5) is given by

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

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 (5) 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 $p > 1/2$,

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

APPENDIX

In this appendix we prove the following Lemma. The bound that we present results directly from the work of Dvoretzky, Kiefer, and Wolfowitz.

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

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

where

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

Proof: We will use the notation of Dvoretzky, Kiefer, and Wolfowitz [3, pp.646-648]. In [3] they establish that

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

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

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

(see (2.11)).

First, consider those j for which $|j - \frac{n}{2}| \leq \frac{n}{4}$. We will show that

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

for $c_2 = \frac{16}{3\sqrt{6\pi}}$. By an approximation of Feller [7] 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 $j(n-j)^3 \geq 27\left(\frac{n}{4}\right)^4$. Also,

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

Thus,

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

Next, we know that

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

and (2.15) holds:

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

$$= \left(\frac{32}{3\sqrt{6\pi}} + \frac{4}{3\sqrt{3}}\right) \exp\left(-2 r^2\right) .$$

Now consider those j for which $\left|j - \frac{n}{2}\right| > \frac{n}{4}$. It follows from the equation at the top of p.647 that

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

for $r > 1$, 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 $f(n,1,j) \leq 40/9 = c$. Therefore, we have that

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

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

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

Using the fact that $r \exp\left(-\frac{r^2}{64}\right)$ is maximized at $r = \sqrt{2}/16$, we obtain

$$Q_n(j,r) \leq \frac{c}{8r\sqrt{2e}} Q_n(j,1) \exp\left(-2 r^2\right) .$$

Thus,

$$r\sqrt{n}\sum{}'' Q_n(j,r) \leq \frac{c}{8\sqrt{2e}} \sqrt{n} \exp\left(-2 r^2\right) \sum{}'' Q_n(j,1)$$

$$< \frac{c}{8\sqrt{2e}} \exp\left(-2 r^2\right)$$

from (2.16).

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

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

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.

ACKNOWLEDGEMENT

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

REFERENCES

1. M. Loève, Probability Theory. New York: Van Nostrand, 1963.
2. W. Hoeffding, "Probability inequalities for sums of bounded random variables," J. Am. Stat. Assoc., 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. Stat., 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," to appear in J. Multivariate Analysis.
6. G. L. Wise, A. P. Traganitis, and J. B. Thomas, "The estimation of a probability density function from measurements corrupted by Poisson noise," to appear in IEEE Trans. Inform. Theory.
7. W. Feller, An Introduction to Probability Theory and Its Applications, Vol. I. New York: Wiley, 1968, p. 54.