

# IMPROVED ESTIMATION OF DISCRETE PROBABILITY DENSITY FUNCTIONS USING MULTIRATE MODELS

Byung-Jun Yoon and P. P. Vaidyanathan

Dept. of Electrical Engineering  
California Institute of Technology, Pasadena, CA 91125, USA  
E-mail: bjyoon@caltech.edu, ppvnath@systems.caltech.edu

## ABSTRACT

For many decades, the problem of estimating a pdf based on measurements has been of interest to many researchers. Even though much work has been done in the area of pdf estimation, most of it was focused on the continuous case. In this paper, we propose a new model based approach for estimating a discrete probability density function. This approach is based on multirate dsp theory, and it has several advantages over the traditional histogram method. It is shown that this method yields an unbiased pdf estimate with small variance, which is guaranteed to have a smaller estimation error than the histogram. Simulation results are given, which show the merit of the proposed method.

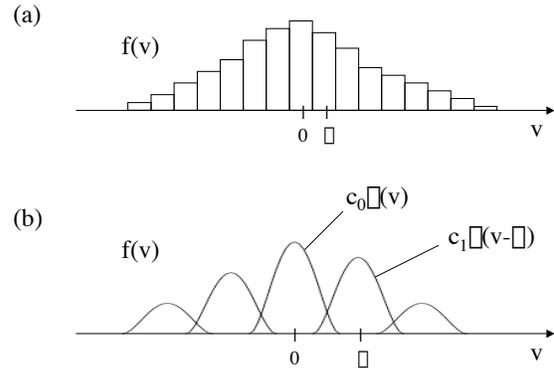
## 1. INTRODUCTION

The problem of density estimation has been widely studied by many researchers in mathematics as well as signal processing communities. The goal is to obtain a good estimate of a pdf  $f(v)$  of a random variable  $v$ , given the observations. The most common way to estimate density functions is the histogram method, though many other methods have been proposed, each with its own advantages. Although histograms may yield reasonable estimates of the original pdf when there are enough observations, it is discontinuous in nature, making it less preferable for estimating continuous random variables. It has been shown that a model based approach has several advantages compared to the histogram method, especially when the number of observations is limited [1], [2].

For example, the kernel based method assumes that the pdf  $f(v)$  can be represented as

$$f(v) = \sum_k c_k \phi(v - s_k, \sigma_k) \quad (1)$$

where  $\phi(v)$  is called the kernel function. It disperses the mass  $c_k$  around the center point  $s_k$ , where  $\sigma_k$  decides the extent to which it will disperse the mass. The kernel function  $\phi(v)$  can be any appropriate positive function, such as a Gaussian, a spline, etc. The preceding model tries to represent the unknown pdf with a linear combination of shifted copies of the fixed function  $\phi(v)$ . With the shifts  $s_k$  and the dispersion factors  $\sigma_k$  typically fixed, the weighting factor  $c_k$  is adjusted based on the measurements of the random variable  $v$ , so that the resulting pdf estimate  $\hat{f}(v)$  approximates the original pdf  $f(v)$  satisfactorily. One advantage of this method is the fact that the resulting pdf estimate  $\hat{f}(v)$  retains some of the properties of the kernel function. For example, if we choose a



**Fig. 1.** (a) Histogram as a special case of kernel based representation when  $\phi(v)$  is rectangular. (b) The pdf representation as a linear combination of shifted versions of the kernel  $\phi(v)$ .

$\phi(v)$  with certain smoothness, the estimate  $\hat{f}(v)$  will also enjoy the same. Let us consider the histogram in Fig. 1(a). This can be considered to be a special case of (1) where  $\phi(v)$  is chosen to be a rectangular pulse,  $\sigma_k$  are fixed so that the width of the pulse is  $\Delta$ , and the shifts are uniform satisfying  $s_k = k\Delta$ . In this case, the mass  $c_k$  will be taken to be proportional to the number of observations that fall in the domain of the  $k$ th pulse  $\phi(v - k\Delta)$ . Generally,  $\phi(v)$  can be chosen such that it is smooth so that we can obtain a smooth pdf estimate. Figure 1(b) shows an example of such a  $f(v)$  with uniform shifts and fixed  $\sigma_k$ . Further discussions on model based methods can be found in many references, e.g. [1], [2], [3].

Even though much work has been done in the area of pdf estimation, most of it was focused on the continuous case. In a recent paper, we proposed a new method for modeling and estimating probability mass functions of discrete random variables [4]. This method is based on multirate dsp models, which takes advantage of well-known results in multirate dsp theory. In this paper, we propose an improved model for discrete pdf estimation, which has several advantages over the traditional histogram method as well the one proposed in [4]. The analogy to the kernel model in (1) will be stated where appropriate. Simulation results are presented in the end that clearly show the advantage of the proposed method.

All notations are as in [5]. Thus  $\downarrow M$  and  $\uparrow M$  represent the  $M$ -fold decimator and expander respectively. Therefore  $[X(z)]_{\downarrow M}$  denotes the  $z$ -transform of the decimated version  $x(Mn)$ , and similarly  $[X(z)]_{\uparrow M} = X(z^M)$  denotes the  $z$ -transform of the expanded version.

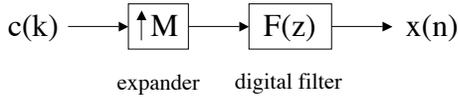
Work supported in part by the ONR grant N00014-99-1-1002, USA.

## 2. MULTIRATE MODEL FOR DISCRETE PDFS

Let us consider a discrete probability density function  $x(n)$  of an integer random variable  $n$ . We assume that this  $x(n)$  can be represented as the output of an interpolation filter  $f(n)$  preceded by an  $M$ -fold expander as in Fig. 2. The input signal  $c(k)$  is the free parameter that is to be adjusted based on the measurements, while  $M$  and  $f(n)$  are fixed. If we let the subspace  $\mathcal{V}_0 = \text{span of } \{f(n - Mk)\}$  where  $k$  is any integer, then  $x(n) \in \mathcal{V}_0$ , and can be written as

$$x(n) = \sum_k c(k)f(n - kM) \quad (2)$$

which is a linear combination of  $f(n)$ ,  $f(n \pm M)$ ,  $f(n \pm 2M)$ , and so on. Notice the analogy to the continuous case in (1).



**Fig. 2.** The basic pdf model.

If both the driving signal  $c(k)$  and the impulse response  $f(n)$  of the interpolation filter are in  $\ell_2$ , the resulting pdf  $x(n)$  also belongs to  $\ell_2$ , hence  $\mathcal{V}_0$  is a subspace of the  $\ell_2$  space<sup>1</sup>. Since this can be viewed as one channel of a  $M$ -channel synthesis filter bank,  $\mathcal{V}_0$  is a proper subspace of  $\ell_2$ . For example, If we choose  $f(n)$  to be a lowpass filter, the resulting  $\mathcal{V}_0$  will be a low frequency subspace. We may choose  $f(n)$  such that  $\mathcal{V}_0$  includes the pdfs that are of our interest. In fact, we can optimize the filter  $f(n)$  for a given class of density functions, and some of the related issues are addressed in [6]. We may also add one or more channels to the model, thereby adding more fine scale components to the probability density function. In this paper, we restrict our attention to the single channel model in Fig. 2, and explain how the estimation procedure works starting from the measurements.

## 3. ESTIMATION OF THE PDF

### 3.1. Estimation Procedure

Let us consider again the pdf model in Fig. 2. Assuming that the probability density function  $x(n)$  can be represented by the output of this model, how can we get the best estimate based on the measurements? In order to answer this question, let us consider a filter  $G(z)$  that satisfies

$$[G(z)F(z)]_{\downarrow M} = 1 \quad (3)$$

This  $G(z)$  is called a biorthogonal partner of  $F(z)$  with respect to  $M$  [7]. One obvious example of such a filter is  $G(z) = 1/F(z)$ . In fact, any  $G(z)$  that can be expressed in the form

$$G(z) = \frac{H(z)}{([H(z)F(z)]_{\downarrow M})_{\uparrow M}}$$

for some  $H(z)$  is a biorthogonal partner of  $F(z)$ , hence the partner is not unique. It is also possible to have an FIR biorthogonal

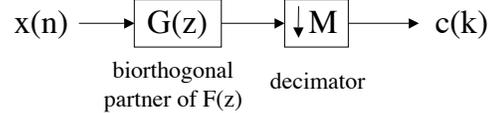
<sup>1</sup>Strictly speaking,  $F(e^{j\omega})$  should be bounded for this.

partner  $G(z)$  under mild conditions on  $F(z)$ . A detailed study of biorthogonal partners can be found in [7].

The importance of biorthogonal partners in estimating the probability density function arises as follows. Let us consider a signal  $x(n)$  that can be represented as in Fig. 2. Therefore, we have

$$X(z) = C(z^M)F(z)$$

From this  $x(n)$ , we can recover the underlying driving signal  $c(k)$  by using a biorthogonal partner  $G(z)$  as in Fig. 3.



**Fig. 3.** Reconstruction of the driving signal  $c(k)$ .

This is not hard to see, since the output of Fig. 3 has the  $z$ -transform

$$\begin{aligned} [G(z)X(z)]_{\downarrow M} &= [G(z)C(z^M)F(z)]_{\downarrow M} \\ &= C(z)[G(z)F(z)]_{\downarrow M} \\ &= C(z) \end{aligned} \quad (\text{From Eq. (3)})$$

hence  $c(k)$  is recovered. Figure 3 shows that  $c(k)$  can be written as

$$c(k) = \sum_n x(n)g(Mk - n) \quad (4)$$

Notice that the signal  $x(n)$  is a probability density function of an integer random variable  $n$ . Therefore the variable  $n$  in (4) should be interpreted as a random variable that is distributed according to  $x(n)$  (instead of as the traditional “time index”). From this point of view,  $g(Mk - n)$  is also a random variable because  $n$  is random, and the right hand side of (4) can be viewed as the expectation of the random variable  $g(Mk - n)$  with respect to  $n$ . Therefore (4) can be rewritten as

$$c(k) = E_n[g(Mk - n)] \quad (5)$$

This kind of interpretation of a signal as the expectation of a random variable naturally appears in almost any non-parametric density estimation scheme [2], [8], [9]. In fact, this plays an important role in the pdf estimation method being proposed in this paper, since this allows us to relate the measurements to the pdf estimate. Assume that we have  $N$  measurements of the random variable  $n$ , and denote them as  $n_i, 0 \leq i \leq N - 1$ . Given these measurements, the expectation in (5) can be approximated by its sample mean as follows:

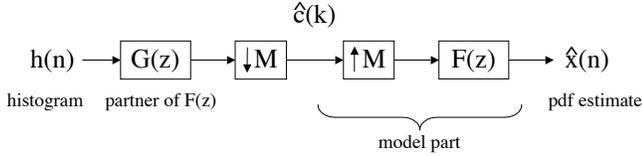
$$\hat{c}(k) = \frac{1}{N} \sum_{i=0}^{N-1} g(Mk - n_i)$$

If we define the signal  $h(n)$  as the relative occurrence of the integer value  $n$  in the measurements  $\{n_i\}$ , we can write  $\hat{c}(k)$  as

$$\hat{c}(k) = \sum_n h(n)g(Mk - n) \quad (6)$$

Since  $h(n)$  is nothing but the histogram obtained from the measurements  $\{n_i\}$ , this means that we can get an estimate of the driving signal  $c(k)$  by feeding the histogram  $h(n)$  to the decimation filter  $g(n)$  and decimating the output by  $M$ , as shown in Fig. 4.

Now that we have the estimate  $\hat{c}(k)$ , this can be used in the original model Fig. 2 to obtain the estimate  $\hat{x}(n)$  of the original pdf. The entire picture is shown in Fig. 4. We may summarize the estimation procedure as follows. Firstly, we assume that the pdf  $x(n)$  of an integer random variable  $n$  comes from the model in Fig. 2. Secondly, we make measurements  $\{n_i\}$  of the random variable  $n$  and construct the histogram  $h(n)$ . Notice that  $h(n)$  is a coarse representation of the original pdf  $x(n)$ , and it need not belong to the subspace  $\mathcal{V}_0$ . Finally, the histogram obtained from above is fed into the system that is shown in Fig. 4 to obtain the estimate  $\hat{x}(n)$ . This pdf estimate belongs to  $\mathcal{V}_0$  as the original pdf.



**Fig. 4.** Estimation of the driving signal  $c(k)$  from the histogram  $h(n)$ , and subsequent estimation of the pdf  $x(n)$ .

Note that the above estimation procedure can be viewed as a projection of the histogram  $h(n)$  onto the subspace  $\mathcal{V}_0$ , where the original pdf  $x(n)$  belongs.

### 3.2. The Choice of $G(z)$

Let us assume that  $F(z)$  and  $M$  are fixed. Since the biorthogonal partner of a filter  $F(z)$  is not unique, the quality of the estimate  $\hat{x}(n)$  may vary depending on the choice of the partner  $G(z)$ . So, the natural question that may arise is how to choose  $G(z)$  in order to obtain the best pdf estimate  $\hat{x}(n)$ , based on the limited number of measurements. To answer this question, let us consider the following. If we choose  $G(z)$  to be the least squares partner of  $F(z)$  with respect to  $M$ , which is defined as

$$G(z) = \frac{\tilde{F}(z)}{([\tilde{F}(z)F(z)]_{\downarrow M})_{\uparrow M}} \quad (7)$$

the projection  $\hat{x}(n)$  becomes an orthogonal projection of  $h(n)$  onto  $\mathcal{V}_0$  [7]. With any other partner  $G(z)$ , the projection  $\hat{x}(n)$  is “oblique” rather than orthogonal. It is shown in [4] that this choice of  $G(z)$  guarantees that the projected signal  $\hat{x}(n)$  is closer to the original pdf  $x(n)$  than the histogram  $h(n)$  is, in  $\ell_2$  sense. In other words, we always have

$$\|h(n) - x(n)\| \geq \|\hat{x}(n) - x(n)\|. \quad (8)$$

Now, suppose that we are going to choose the decimation filter  $G(z)$  to be the LSBP (least squares biorthogonal partner) of  $F(z)$  with respect to  $M$  as in (7). If we consider the denominator  $B(z) = ([\tilde{F}(z)F(z)]_{\downarrow M})_{\uparrow M}$  of  $G(z)$ , it satisfies

$$B(z) = \tilde{B}(z) = B^*(1/z^*).$$

Therefore if  $B(z)$  has a zero at  $z_0$ , then there exists another zero at  $1/z_0^*$ . This can be a problem, since it means that  $G(z)$  cannot have all the poles inside the unit circle, and therefore it cannot be a causal stable filter. One way to get around this problem is to choose  $F(z)$  such that its magnitude square is Nyquist( $M$ ), i.e.

$$[\tilde{F}(z)F(z)]_{\downarrow M} = 1 \quad (9)$$

In this case, the least squares partner  $G(z) = \tilde{F}(z)$ , which can be written as  $g(n) = f^*(-n)$  in the time-domain. We can observe that (9) is equivalent to imposing the orthonormality constraint on the basis functions  $\{f(n - kM)\}$  that span the subspace  $\mathcal{V}_0$ . The interpolation filter  $F(z)$  that satisfies the orthonormality constraint can be designed using one of many known techniques [5].

However, this approach suffers from one disadvantage, namely the fact that the positivity of the output  $x(n)$  may not be guaranteed. This is an important point, when using the estimation process shown in Fig. 4, since the resulting estimate  $\hat{x}(n)$  may not be positive. Consider designing the filter  $F(z)$  such that (9) is satisfied. Then  $f(n)$  will necessarily have negative coefficients unless it has order  $< M$ . Since the projection  $\hat{x}(n)$  consists of a linear combination of shifted copies of  $f(n)$ , it is very likely that  $\hat{x}(n)$  will have some negative coefficients as well. In order to overcome this problem, we may use the model in Fig. 2 to represent the square-root  $x_s(n)$  of the pdf  $x(n)$ , rather than the pdf itself. This square-root pdf model is elaborated in [4], which guarantees that the pdf estimate is positive.

## 4. FIR TRUNCATION OF THE LSBP

Although the square-root model guarantees a positive pdf estimate, it has several shortcomings. In order to get a satisfactory estimate, the sign of the square-root of the histogram has to be adjusted, before it is used in the estimation procedure [4]. The searching process for the optimal signature sequence can be computationally very expensive. Another disadvantage of this approach is the fact that the estimation results are not easy to analyze analytically due to the nonlinearity of the model. In this section, we consider a linear model for representing pdfs, which ensures that the resulting estimate is always positive, and uses only stable and realizable filters in the estimation procedure.

Let us consider again the model in Fig. 2. In order to ensure that the pdf estimate is non-negative, all the coefficients of the filter  $f(n)$  should be non-negative. Now take  $G(z)$  to be the least squares partner of  $F(z)$  as in (7). We know from section 3.2 that unless  $F(z)$  has a filter order  $< M$ ,  $G(z)$  has poles both inside and outside the unit circle, which means that  $G(z)$  cannot be a causal stable filter. However, it is possible to approximate such a filter by an FIR filter by choosing the region of convergence properly, as long as there are no poles on the unit circle [10].

Let us consider an IIR filter  $Q(z)$ , whose poles lie far from the unit circle. Let  $p_{in}$  be the pole with the largest modulus among all poles inside the unit circle, and let  $p_{out}$  be the pole with the smallest modulus among all poles outside the unit circle. If we choose the region of convergence to be the annular region in the  $z$ -plane that satisfies  $|p_{in}| < |z| < |p_{out}|$ , then  $Q(z)$  becomes a noncausal stable filter. Since all poles are far away from the unit circle, the impulse response  $q(n)$  decays very fast. This allows us to approximate this IIR filter by truncating its impulse response as follows.

$$q_L(n) = \begin{cases} q(n) & \text{if } |n| \leq L \\ 0 & \text{otherwise} \end{cases}$$

This corresponds to multiplying  $q(n)$  with a rectangular window, and some other windows (e.g. Kaiser window, Hamming window, etc.) can also be used. Note that  $L$  should be large enough such that most of the energy of  $q(n)$  is confined in  $|n| \leq L$ . Unless there are poles very close to the unit circle, it is possible to approximate  $q(n)$  with a reasonable length  $L$ . For example, it is shown

in [10] that the cubic B-spline filter can be well approximated by a truncated FIR filter of length only five or seven.

Returning to our original interest, let us consider again  $G(z)$ , the least squares partner of  $F(z)$ . Let  $B(z)$  be the denominator of  $G(z)$ . If we choose  $f(n)$  such that  $f(n) \geq 0$  and  $B(z) = [\tilde{F}(z)F(z)]_{\downarrow M \uparrow M}$  has zeros sufficiently apart from the unit circle, it is possible to approximate  $G(z)$  by an FIR filter  $G_L(z)$ , by truncating it using a window function. Using this FIR filter  $G_L(z)$  in place of  $G(z)$  in Fig. 4, we can use a similar estimation procedure as elaborated in section 3.1.

However, one more remark remains to be made regarding the positivity of the pdf estimate. Since the interpolation filter  $f(n)$  is non-negative, it is possible to make the output signal also non-negative by taking a non-negative driving signal  $c(k)$ . So, when modeling the original pdf  $x(n) = \sum_k c(k)f(n - Mk)$ , we can make it a valid pdf by choosing  $c(k) \geq 0$ , and normalizing  $x(n)$  so that it adds up to 1. But when feeding the histogram  $h(n)$  into the system shown in Fig. 4, there is no guarantee that the estimate  $\hat{c}(k)$  will be non-negative for  $\forall k$ . Correspondingly, the orthogonal projection  $\hat{x}(n)$  may not satisfy the non-negativity condition. In order to guarantee that the pdf estimate is non-negative, we simply drop the negative values of  $\hat{x}(n)$  to obtain a positive estimate as follows.

$$\hat{x}_p(n) = \begin{cases} \hat{x}(n) & \text{if } \hat{x}(n) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Note that this  $\hat{x}_p(n)$  may not necessarily belong to  $\mathcal{V}_0$ . Nevertheless, this estimate gets even closer to the original pdf  $x(n)$ , and it can be easily shown that

$$\|h(n) - x(n)\| \geq \|\hat{x}_p(n) - x(n)\| \quad (10)$$

which guarantees that the pdf estimate  $\hat{x}_p(n)$  is always closer to the true pdf  $x(n)$  than the histogram is.

## 5. BIAS AND VARIANCE OF THE PDF ESTIMATES

Since the pdf estimates are based on random observations, the estimates themselves are random variables. Therefore, it is important to understand their statistical properties. Two important properties of a random estimate are the bias and the variance. The bias of an estimate tells us how close it is to the real value ‘‘on the average’’, and similarly, the variance indicates how much an estimate is expected to deviate from the average value [11]. It is desirable to have an estimate that is unbiased and has a small variance at the same time. In the following sections, we’re going to focus on the model in Fig. 2, analyzing its bias and variance, and finally compare them with those of the histogram method.

### 5.1. Histogram Method

Let us first consider the histogram method. Assume that we have  $N$  observations of a random variable  $n$ , where the underlying pdf is  $x(n)$ . The probability that the  $i$ th observation will be  $n_i = n$  is

$$P\{n_i = n\} = x(n)$$

for all  $i = 0, 1, 2, \dots, N - 1$ . Therefore if we let  $y(n)$  be the number of observations in  $\{n_i\}$  that have the value  $n$ ,  $y(n)$  is a binomial random variable with mean  $\mathcal{E}\{y(n)\} = Nx(n)$  and variance  $\text{Var}\{y(n)\} = Nx(n)\{1 - x(n)\}$ . Notice that the histogram can

be written as  $h(n) = y(n)/N$ . Therefore the expectation of  $h(n)$  is

$$\mathcal{E}\{h(n)\} = \frac{1}{N}\mathcal{E}\{y(n)\} = x(n) \quad (11)$$

which shows that the histogram estimate is unbiased. Also from  $h(n) = y(n)/N$ , we get the following variance of  $h(n)$ .

$$\text{Var}\{h(n)\} = \frac{1}{N^2}\text{Var}\{y(n)\} = \frac{1}{N}x(n)\{1 - x(n)\} \quad (12)$$

The variance of the histogram estimate, defined as  $\sum_n \text{Var}\{h(n)\}$  is therefore

$$\mathcal{E}\|h(n) - x(n)\|^2 = \sum_n \text{Var}\{h(n)\} = \frac{1}{N}\{1 - \sum_n x^2(n)\} \quad (13)$$

From (13) we can see that the variance of the estimate decreases as the number of observations  $N$  increases, as expected.

### 5.2. Model Based Method

Let us consider the model in Fig. 4 again. We can write the output as

$$\hat{x}(n) = \sum_l \sum_k h(k)g(Ml - k)f(n - Ml)$$

Therefore, the expectation  $\mathcal{E}\{\hat{x}(n)\}$  can be written as

$$\begin{aligned} \mathcal{E}\{\hat{x}(n)\} &= \sum_l \sum_k \mathcal{E}\{h(k)\}g(Ml - k)f(n - Ml) \\ &= \sum_l \sum_k x(k)g(Ml - k)f(n - Ml) \end{aligned} \quad (14)$$

We can see that the last expression in (14) is the output of Fig. 4 when the input signal is  $x(n)$ . Therefore (14) simply reduces to  $x(n)$ . This proves that the model based method results in an unbiased pdf estimate. It can be shown that the variance of the pdf estimate  $\hat{x}(n)$  is

$$\begin{aligned} \mathcal{E}\|\hat{x}(n) - x(n)\|^2 &= \sum_n \text{Var}\{\hat{x}(n)\} \\ &= \frac{1}{N} \left\{ \sum_k \sum_l x(k + lM)g(-k)f(k) - \sum_n x^2(n) \right\} \end{aligned} \quad (15)$$

Now, let us compare the variances of the two estimates. We can show that the variance of the model based estimate is always smaller than that of the histogram. In order to see this, note that  $\hat{x}(n)$  is the orthogonal projection of  $h(n)$  onto  $\mathcal{V}_0$ . Therefore, we can write  $h(n) = \hat{x}(n) + e(n)$  where  $\hat{x}(n) \in \mathcal{V}_0$  and  $e(n) \in \mathcal{V}_0^\perp$ . So, we have

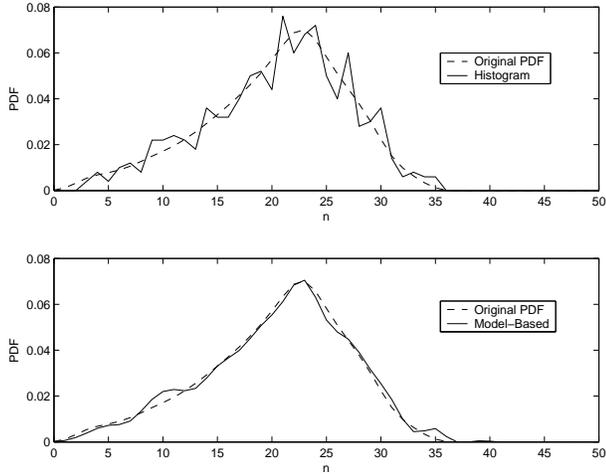
$$h(n) - x(n) = \hat{x}(n) - x(n) + e(n)$$

Since  $x(n) \in \mathcal{V}_0$ , we have  $\hat{x}(n) - x(n) \in \mathcal{V}_0$  and  $e(n) \in \mathcal{V}_0^\perp$ , hence it follows that

$$\|h(n) - x(n)\|^2 = \|\hat{x}(n) - x(n)\|^2 + \|e(n)\|^2$$

Therefore if we subtract the variance of the model based estimate (15) from that of the histogram estimate (13), we get

$$\begin{aligned} \mathcal{E}\|h(n) - x(n)\|^2 &- \mathcal{E}\|\hat{x}(n) - x(n)\|^2 \\ &= \mathcal{E}\|e(n)\|^2 \\ &= \mathcal{E}\|e(n)\|^2 \geq 0 \end{aligned} \quad (16)$$



**Fig. 5.** PDF estimation result. Top plot: the original pdf and the histogram. Bottom plot: the original pdf and the model based pdf estimate.

This shows that the model based pdf estimate in Fig. 4 has a smaller variance than the histogram based estimate. The reduced variance is due to the fact that the pdf estimate  $\hat{x}(n)$  is restricted to  $\mathcal{V}_0$ , which is a proper subspace of  $\ell_2$ .

## 6. SIMULATION RESULTS

In order to demonstrate the ideas in section 4 and section 5, we present an example as follows. We assume  $M = 2$  and use  $F(z) = (1+z)^L/2^L$  with  $L = 6$ . Notice that this filter leads to the 5th order spline function [12].  $G(z)$  is chosen to be the least squares partner of  $F(z)$ , truncated by a rectangular window of length 39. By choosing an appropriate driving signal  $c(k)$ , we obtained a sample pdf  $x(n)$  of length 37. We generated 500 random variables according to  $x(n)$ , and the histogram was obtained from the observations. Then the histogram was fed into the system in Fig. 4 to get the orthogonal projection of the histogram. Finally, negative values in the output of Fig. 4 were dropped, and the result was normalized to get the pdf estimate  $\hat{x}(n)$ . Fig. 5 shows the simulation results. The histogram shown in Fig. 5(top) is quite different from the original pdf, whereas the model based estimate is considerably close to the true pdf, as can be seen in Fig. 5(bottom). The estimation error was

$$\sum_n |h(n) - x(n)|^2 = 0.00151128$$

for the histogram, and

$$\sum_n |\hat{x}(n) - x(n)|^2 = 0.00017605$$

for the model based estimate, which is only about 12% of the error of the histogram. In order to compute the variance of the pdf estimates, we repeated the same estimation procedure as above for 100 times, where 500 random variables were generated in each experiment. The variance, or equivalently the mean square error, of the histogram estimate was

$$\sum_n \text{Var}\{h(n)\} = 0.00191544 \quad (17)$$

and the variance of the model based estimate was

$$\sum_n \text{Var}\{\hat{x}(n)\} = 0.00091091 \quad (18)$$

These values are indeed very close to the theoretical values computed from (13) and (15), which are

$$\mathcal{E}\|h(n) - x(n)\|^2 = 0.00191044, \quad \mathcal{E}\|\hat{x}(n) - x(n)\|^2 = 0.00091044$$

We can observe that the variance of the model based pdf estimate is much smaller than that of the histogram, as expected.

## 7. CONCLUDING REMARKS

In this paper, we proposed a new method for non-parametric estimation of discrete probability density functions. It was shown that the proposed method yields an unbiased estimate with small variance, and that it is guaranteed to have a smaller estimation error than the histogram approach. One of the important issues in this approach is the optimization of the interpolation filter  $f(n)$ , such that the subspace  $\mathcal{V}_0$  includes the pdfs of interest. Another interesting problem is the estimation of a pdf in the presence of noise. These are topics for future research.

## 8. REFERENCES

- [1] Chiu, S. T. "Bandwidth selection for kernel density estimation", *Ann. Stat.*, vol. 19, pp. 1883-1905, 1991.
- [2] Walter, G. G., and Ghorai, J. K., "Advantages and disadvantages of density estimation with wavelets", *Comp. Sci., Stat.*, vol. 24, pp/ 234-243, 1993.
- [3] Walter, G. G. "Density estimation in the presence of noise", *Stat. and Prob. letters*, 41, pp. 237-246, 1999.
- [4] P. P. Vaidyanathan and Byung-Jun Yoon, "Discrete Probability Density Estimation Using Multirate DSP Models", *Proc. 28th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Hong Kong, Apr. 2003.
- [5] Vaidyanathan, P. P., *Multirate systems and filter banks*, Prentice Hall, Inc., 1993.
- [6] Byung-Jun Yoon and P. P. Vaidyanathan, "Non-Parametric Estimation of Discrete Probability Density Functions Using Multirate DSP Models", *Internal Report*, Dept. of Electrical Engineering, California Institute of Technology, May 2003.
- [7] P. P. Vaidyanathan and B. Vrcelj, "Biorthogonal partners and applications", *IEEE Trans. Signal Processing*, vol. 49(5), pp. 1013-1027, May 2001.
- [8] Āencov, N. N., "Evaluation of an unknown distribution density from observations", *Doklady*, (3):1559-1562, 1962.
- [9] Vidakovic, B., *Statistical modeling by wavelets*, John Wiley & Sons, Inc., NY, 1999.
- [10] B. Vrcelj and P. P. Vaidyanathan, "Efficient implementation of all-digital interpolation", *IEEE Trans. Image Processing*, vol. 10(11), pp. 1639-46, Nov. 2001.
- [11] Therrien, Charles W., *Discrete random signals and statistical signal processing*, Prentice Hall, Inc., NJ, 1992.
- [12] Strang, G. and Nguyen, T. *Wavelets and filter banks*, Wellesley-Cambridge Press, MA, 1997.