Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

SUBJECT: THE SPECIFICATION OF AN IDEAL DETECTOR AS A FIRST STEP IN FILTER DESIGN

To: W.K. Linvill

From: W.I. Wells

Date: March 6, 1953

Abstract: We show in this report that the first problem of ideal filter design, that of Detection,[1] is based upon two probability distributions. The first describes the process of measurement, or introduction of noise. The second describes the actual function it is desired to filter. A method for describing both of these statistical processes is given which seems very reasonable and useful in the case of sampled data filters. After the description of the problem, consisting of the specification of these two distribution functions, we give the method of combining them according to the rules of probability theory. This calculation leads to the construction of a probability distribution function involving the variables that it is desired to filter. This distribution function is the desired output of the ideal detector. A complete mathematical analysis is given, and also a simple example, to illustrate the technique.

INTRODUCTION

In this report we will discuss a method for designing a filter on the basis of a statistical analysis of the processes involved. It was shown[1] that the action of any ideal filter can be thought of in two independent steps. The first step is that of Detection. The second step is Selection.

The purpose of the ideal detector is to construct the mathematical function that represents the probability density distribution of the value of the variable being considered. We might illustrate this by assuming a one-dimensional tracking problem where the position coordinate is (x). By a series of measurements we attempt to find the position of this variable. Because of unknown factors and random variables in the process, we can only find a probability density distribution function for (x), that is, a function that gives the probability that the true value of (x) lies in a given small range. The calculation of this probability density distribution function is the desired action of the detector. We will see in detail how this is accomplished, later. But first, we will see what the Selector does.

In general, when one measures a quantity such as (x) above, one intends to take some action based upon these measurements. If this

---

[1] "The Philosophy of Statistical Filter Design," M-1812, W.I. Wells, Jan. 27, 1953.

were not the case, then why make the measurement? The question is then, now that we have the distribution function, how do we decide what specific action to take? This obviously is decided upon by considering the distribution function and also by considering the final desired result of the action. The influence of the desired result of the action is what one associates with the strategy of the situation. Mathematically, it is related to a weighting function or Scanning Function. It is the purpose of the Selection operation to combine the effects of the probability density distribution function with the Scanning Function to make the actual selection of the action to be taken. Only the Detection principles will be investigated in this report.

Detection. As we have stated, the action of the Ideal Detector is to construct the probability density distribution function for the variable being measured. Let us now see what factors must be considered in performing this.

First we might consider the measurements that are made of the variable (call it X). For the type of problem being considered here X is assumed to be some function of time. That is, it may have a different, though unique, value at each different time. Our purpose is to make measurements of X and then try to construct X as well as possible from these measurements. By as well as possible we actually mean that we wish to construct the probability density distribution function for the value of X at each particular time. As a restriction of the problem we will consider only those cases where the measurements are made at discrete intervals of time. This is then a sampled-data system. We will not require in general that the samples be taken at equally spaced times, although this is at present the case of most interest.

When we take these measurements or samples, it is in general not possible to say the measurements that we make are the exact value of X at that time. The reason for this is that all measuring devices are inherently inaccurate to some degree. We find in general that the true value of X, for a large number of measurements, is usually distributed about the measured values. This distribution function is called the distribution function of the "noise." We say that if the measurements are not exact, they are noisy, and the value of the "noise" is randomly distributed according to a distribution function. This function depends upon the particular measuring device. One finds that the measurements made with a sensitive galvanometer are distributed according to a normal distribution function about the true values. In case the values of the samples are quantized the distribution function is a flat distribution over the width of the quantization interval.

The types of noise that may contaminate the data are not restricted to those caused directly by a measurement device. In case the data is passed over some transmission link, noise is invariably added to the signals. This may be of the thermal type or due to some interfering signals. In any case we may still represent this effect by giving the probability density distribution fumction of the noise.

Measurement. For the moment let us work only with the measurement system. We can assume that some function X is to be measured, but nothing about this function is known, á priori. Suppose we take several measurements and then try to reconstruct X. Obviously since no á priori knowledge was given concerning X, it will be impossible to infer anything about X except at the time of each sample. Since each measurement is assumed to be independent of all the others, we find that the joint probability density distribution function of X, at each sample time, is the product of the probability distribution functions for X at each time. As an example, we might suppose that the distribution function for the noise has the form of a normal distribution. That is, the true value of X at the time of the sample is normally distributed about the value of the sample. If we let X_k be the true value of X at the time t equals k and a_k be the value of the sample at this time, we write the distribution of X_k after the reception of a_k as

$$W(X_k/a_k) = \frac{1}{\sqrt{2\pi}\,\sigma_k}\, e^{-\frac{(X_k - a_k)^2}{2\sigma_k^2}} \qquad 1.$$

Then we can also write the product which gives the joint distribution of X for several times after the reception of several samples.

$$W(X_1, \ldots, X_k/a_1, \ldots, a_k) = \prod_{r=1}^{k} \frac{1}{\sqrt{2\pi}\,\sigma_r}\, e^{-\frac{(X_r - a_r)^2}{2\sigma_r^2}} \qquad 2.$$

One notices here that the variance is written as a function of the index, (r). Although most measuring systems' accuracy does not depend upon time, many of them do depend upon the actual range of values being measured. Thus, actually the subscript on the variance might be written so as to show the dependence upon the a_r rather than (r) itself. This detail is easily taken care of in an actual problem. We need only remember for the time being that the variance of each measurement may be different. For instance, in the case of a quantized function the size of the quantization interval may be a function of the value of the variable.

The important thing to notice here is that we are able to write a joint probability density distribution for the values of X at each time a sample is taken. When we do this we have not made use of any other characteristics than those of the measurement system or the noise. In case we knew nothing whatever about X this would be the final output of the detector. Fortunately we often know something about the function X, which enables us to sharpen up this joint distribution.

The Function Representation. We will now discuss the Representation of functions (x) that we will consider. We must be able to describe them in some convenient and accurate way so that the joint distribution found above may be sharpened up and give us a better picture of the function. In particular, if we wish to find out about the function in between the sample times we must know about the general character of X.

The type of systems to be considered are those normally encountered in control problems. Their distinguishing characteristic is their inertia. To cause a motion or change thereof it is necessary to apply a force. This is, of course, obvious, but the reason for pointing it out is that it is the underlying basis for the type of representation that will be used. In inertial systems it is the forces that are important, thus it seems reasonable to describe such a system by describing the values of the forces acting on the system. When we consider one-dimensional motion along the X axis, the applied force is proportional to the second derivative of X. It seems reasonable therefore that if we characterize the second derivative of such intertial systems, we will be able to describe its action accurately. There are other ways of course, however, since this particular method will be seen to be most convenient, we will use it. A slightly different approach to this same representation may be found as follows.

As stated before, we are going to consider only sampled systems. Thus, it is possible to describe the function during each sample interval. This approach leads also to the idea of describing the derivative. Let us suppose that we break X up into sample intervals. Within each interval we notice that the function can be approximated very closely, if the sampled intervals are small, by a polynomial of fairly low order. Suppose that an n'th order polynomial is found to approximate the function X in each interval. The polynomial is different in each interval, but of the same order. Now by differentiation we see that the n'th derivative of our composite function is a constant in each interval. Calling the n'th derivative D^n we see that the value of D^n is likely to be different in each interval but a constant during the interval. Now we ask how this fits with the idea of inertial systems. If the sample period is chosen small enough, a second order polynomial (parabola) will suffice as an approximation in each interval. Then it is the acceleration which is a constant in the interval. Our physical reasoning leads us to suspect that the forces being applied to most physical systems are constant most of the time with changes in value occurring only occasionally.. We might plot a supposed graph of force against time (Fig. 1).

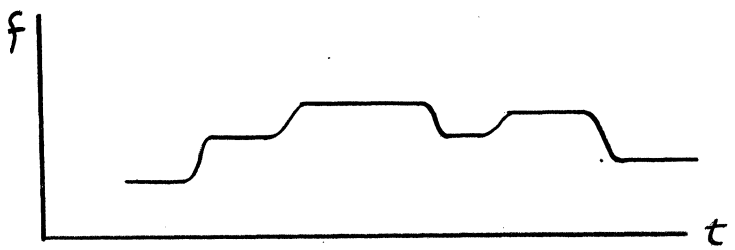

Fig. 1

The changes are not really too abrupt, but the fact that the forces are constant over extended periods checks with our physical reasoning. The approximation that we are making here is that this curve will be approximated by a step function. This is done for convenience; however we must have made sure that the second integral of the step function does give an acceptable approximation to the function X.

In case the function X varies more quickly, it will be necessary to use a higher order polynomial than the second, however the same ideas of constant n'th derivative in the interval carry over directly. In general we will consider an n'th order polynomial approximation in each interval. The actual order needed depends upon the exact problem.

Now that the exact form of the representation has been chosen, we ask: Exactly what quantities must be specified to determine the function X(t)? It is obvious that one of the things we shall require will be the value of D^n, the n' th derivative, in each sample interval. Then in addition to these values we must know the "initial conditions." That is, if we know X and the first n-l derivatives for some interval, then the knowledge of the D^n in each interval enables us to construct X(t).

Since in the case of interest any or all of these quantities will be described statistically, we require the joint probability distribution of X and its first n-l derivatives in one interval and the n'th derivative in each interval. This joint distribution is the complete description of X(t). We write it as follows:

$$W(X_k, D_k^1, \ldots D_k^{n-1}, D_1^n, D_2^n, \ldots D_k^n, \ldots) , \qquad 4.$$

Where D_1^m is the value of the m'th derivative at the beginning of the l'th interval.

At this point we can introduce the concept of stationarity . Suppose all sample intervals are equal for the moment. Now if the process is stationary in the sense we have chosen, we mean that each of the D^n enters the above expression in the same way, except for the effect of the discontinuity at t = 0. This is important merely because we start sampling the function at a finitely remote time. If one will imagine that we have taken a very large number of samples, then each D^n will enter the expression in the same manner. If the process is not stationary, then the D^n will enter the expression in a way that depends upon time. If the process changes very slowly, that is, over a long time the D^n all have approximately the same distribution, we call it quasi stationary. If this is the case and we can find how the variations take place, we may still be able to use the information. In general, however, this paper will deal only with stationary processes. In case the sample intervals are not equal the condition is that the D^n must all enter the expression (4) in the same manner except for the effect of variations in length of interval.

With these concepts and definitions in mind we are ready to see how the detection process combines the above expressions, i.e., analogous

to Eq. 2 and Eq. 4. in order to accomplish its given function. Before launching into the formal solution of this problem it is instructive to get an intuitive feeling for the steps involved by going through a very simple example. In the treatment of the general case, the number of formal manipulations is few but they do not convey the physical reasoning which leads to them. Thus we will give an example first.

Example, Solution A. Instead of beginning with the most general problem, we will imagine this very simple one and work through it to get the main ideas straight. This example, although very simple, contains all of the ideas that are required in the more general treatment. The problem is this:

We are sampling a function X, with a measurement system that introduces random errors that are normally distributed about the true value. The distribution of the true value X_k about the sample value a_k is then:

$$W(X_k/a_k) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(X_k-a_k)^2}{2\sigma^2}} \qquad 5.$$

Further, after examining the process by which X is generated, it is found that the order of the approximating polynomial is two. This means that the acceleration (@) will be assumed a random variable and constant in each sample interval. (@) is assumed to be independent of any past values and to be normally distributed also:

$$W(X_1, D_1', D_1^2) = W(@) = \frac{1}{\sqrt{2\pi}\beta}\, e^{-\frac{(@)^2}{2\beta^2}} \qquad 6.$$

The initial values of X and the derivative, V, the velocity, are assumed to be uniformly distributed. We suppose that we have taken three equally spaced samples (a_1, a_2, a_3) and are interested in finding the joint distribution of the three true values of X (X_1, X_2, X_3).

Even if we were not given Eq. 6, we could write a joint distribution for (X) on the basis of the measurements and the characteristics of the measurement system. This is just the product of terms like Eq. 5. So we write the distribution:

$$W_0(X_1,X_2,X_3/a_1,a_2,a_3) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(X_1-a_1)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(X_2-a_2)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(X_2-a_3)^2}{2\sigma^2}} \qquad 7.$$

It turns out in the following that the coefficients do not play an important part so they will be dropped. One just supposes that a normalizing factor is needed to reduce the total probability to one. This then becomes:

$$W_0(X_1,X_2,X_3/a_1,a_2,a_3) = e^{-\frac{1}{2\sigma^2}\left[(X_1-a_1)^2 + (X_2-a_2)^2 + (X_3-a_3)^2\right]} \qquad 7a.$$

Now we would like to alter this distribution to take into account the fact that we know something about X itself, namely Eq. 6. There are several ways to do this, but since we are interested in the joint distribution of the X's we will proceed as follows. We have in Eq. 7a, a joint distribution of the X's that is independent of any á priori knowledge that we might possess. If we put the á priori knowledge in the form of a joint distribution of the X's, we can multiply these two independent distributions together to get the final joint distribution of the X's.

In order to get Eq. 6 into the form we desire we must express @ in terms of the X's. This is very easily done by the following equations:

$$@_1 = 2(X_1 - X_2 + V_2)$$
$$@_2 = 2(X_3 - X_2 - V_2) \qquad 8.$$

V_2 is the velocity at the beginning of the second sample interval. Any of the velocities could have been used but V_2 was convenient. The reason that this additional variable must be introduced is that one is, in Eq. 8, trying to express three variables, X, in terms of only two new ones, @. The reason there are only two @'s is obvious since there are only two sample intervals involved between three samples. Using equation 8 we can write, from the form of Eq. 6, the joint distribution of the X's based upon our á priori knowledge:

$$W_1(X_1,X_2,X_3,V_2) = \frac{1}{\sqrt{2\pi\beta}}\, e^{-\frac{\left[2(X_1-X_2+V_2)\right]^2}{2\beta^2}} \frac{1}{\sqrt{2\pi\beta}}\, e^{-\frac{\left[2(X_3-X_2-V_2)\right]^2}{2\beta^2}} \qquad 9.$$

Dropping the coefficients as before:

$$W_1(X_1,X_2,X_3,V_2) = e^{-\frac{4}{2\beta^2}\left[(X_1-X_2+V_2)^2 + (X_3-X_2-V_2)^2\right]} \qquad 9a.$$

Now the desired output of the detector is just the product of 7a. and 9a. This is the joint probability density distribution function of X_1,X_2,X_3, after the reception of the three samples, and taking into account the á priori knowledge of the form of X. Thus we have:

$$W(X_1,X_2,X_3,V_2/a_1,a_2,a_3) \qquad 10.$$
$$= e^{-\frac{1}{2\sigma^2}\left[(X_1-a_1)^2 + (X_2-a_2)^2 + (X_3-a_3)^2\right] - \frac{4}{2\beta^2}\left[(X_1-X_2+V_2)^2 + (X_3-X_2-V_2)^2\right]}$$

Although this is the final output of the Detector, one may not realize what all this means until he performs the process of selection. The reason is, that this is a joint distribution in five dimensions and is difficult to visualize. In the case of normal distribution functions it has been shown[1] that the process of selection practically always consists of finding the "most probable" values. To interpret Eq. 10 we will try to find the most probable value for the four variables involved. That is, we will try to find the values of the four variables that are jointly most likely to occur. When the distribution function Eq. 10 is

visualized as a surface in five dimensions, we see that our problem is to find the "highest point" on the surface. That is, the point (X_1, X_2, X_3, V_2) for which W() is maximum. We do this by taking the four partial derivatives with respect to the four variables and setting each equal to zero. This yields four simultaneous equations which in turn are to be solved for the four variables that are the most likely values. The solution to these four equations yields:

$$X_1 = \frac{(2k + 5)a_1 + 2a_2 - a_3}{2(3+k)}$$

$$X_2 = \frac{a_1 + (1+k)a_2 + a_3}{3 + k}$$

$$X_3 = \frac{- a_1 + 2a_2 + (5 + 2k)a_3}{2(3 + k)}$$

$$V_2 = \frac{a_3 - a_1}{2}$$

where $k = \frac{\beta^2}{46^2}$.

The ratio (k) is a convenient measure of the relative effect of the distribution of noise as compared with the distribution of (@). For instance, if (k) is zero, we may infer that (@) is zero. This means that the function X is known to be a perfectly straight line and, hence, if any of the samples vary from a straight line, this must be due to the inaccuracies in the measurement system. The Eqs. 11 in this case give:

$$X_1 = \frac{5a_1 + 2a_2 - a_3}{6}$$

$$X_2 = \frac{a_1 + a_2 + a_3}{3} \qquad 12.$$

$$X_3 = \frac{- a_1 + 2a_2 + 5a_3}{6}$$

$$V_2 = \frac{a_3 - a_1}{2}$$

or since the velocity is always equal to V , we may write:

$$X_3 = X_2 + V_2 = X_1 + 2V_2$$

which is indeed a straight line. Suppose for instance we receive samples that are not on a straight line; let us see what the filter does with them.

Let $a_1 = 1$, $a_2 = 2$, $a_3 = 2$. Then:

$$V = V_2 = 1/2$$

$$X_1 = 7/6 \quad X_2 = 10/6 \quad X_3 = 13/6$$

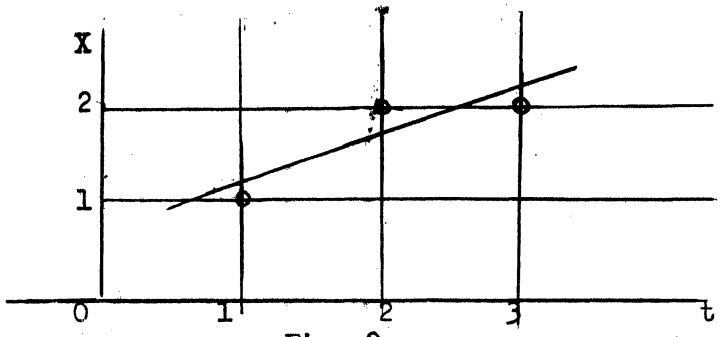

Fig. 2

We see in the plot Fig. 2 that the filter actually passes a perfectly straight line near to the given sample values. Of course, if the samples are themselves on a straight line, the line passes through them exactly.

Now suppose $k = 0$. This is the case of no noise whatever. In this case we would expect to believe the samples as true values and the filter should pass the proper curved line exactly through each sample point. The Eqs. 11 become:

$$X_1 = a_1$$

$$X_2 = a_2 \qquad 13.$$

$$X_3 = a_3$$

$$V_2 = \frac{a_3 - a_1}{2}$$

For the same sample values as above it is easily seen that the filter now passes a curved line through the samples that looks as follows:

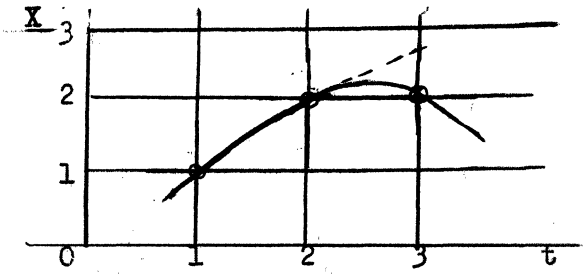

Fig. 3

This is seen to satisfy the condition that that the slope is 1/2 at t = 2.

If k has some intermediate value, this means that there is some noise and also a good likelihood of a change in velocity. In this case one would have great difficulty weighing these two effects by an intuitive method. However, now that we have the equations worked out, it is possible to get the exact answer, one which takes into account the possibilities that the samples may be wrong a certain amount due to noise, and that they may be displaced a certain amount due to the changes in velocity. Suppose we take the case k = 1. Then Eqs. 11 become:

$$X_1 = \frac{7a_1 + 2a_2 - a_3}{8}$$

$$X_2 = \frac{a_1 + 2a_2 + a_3}{4} \qquad 14.$$

$$X_3 = \frac{-a_1 + 2a_2 + 7a_3}{8}$$

$$V_2 = \frac{a_3 - a_1}{2}$$

For the same samples as above these values are:

$X_1 = 9/8 \quad X_2 = 14/8 \quad X_3 = 17/8 \quad V_2 = 1/2$

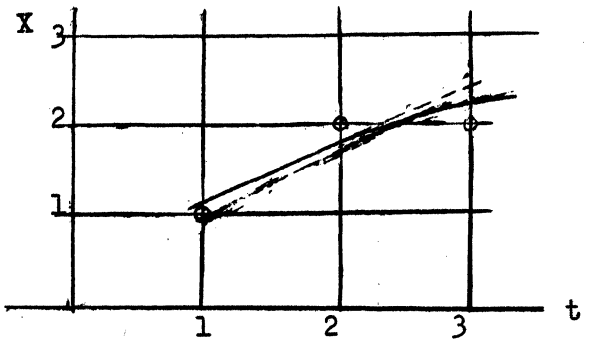

Fig. 4

Here again it is seen that the curve does not actually pass through any of the sample values; however, it is not necessarily a straight line as it was in the first case.

Discussion of Solution A. It will be recalled that we used V_2 as one of the variables, and the statement was made that we could have used any of the velocities. One sees now why that is possible. The solutions that we have obtained are complete in the sense that we can calculate from them any of the other velocities or values of X at intermediate points. Actually the solution that we have obtained is too general for what is needed in control systems or real time problems. In some type of problems it is of interest to solve for the values of X at several different times jointly. If the sequence of values was a code word having a certain meaning, it would be of interest to determine the entire word together. On the other hand, in real time control systems, it is frequently only of interest to determine the value of X at the present time. Actually we do not care

for the exact value of X at some past time, except insofar as it helps to determine the present value of X. In other words, it would have been sufficient to determine the distribution of X_3 alone, without solving the four simultaneous equations for all of the X's. One may accomplish this by integrating Eq. 10 over the variables X_1, X_2, V_2, leaving a distribution involving X_3 only. Now we can differentiate with respect to X_3 and set that equal to zero and determine the most likely X_3. It should not be expected that this value will be the same as found from the joint distribution, although it may be. This seems strange that the most likely value of X_3 should be a different number when two different methods are used to determine it. This is really a question of what particular thing we are trying to determine.

Let us illustrate with a more straightforward problem, that of drawing colored balls from urns. Suppose we have three urns with three different colors of balls in each urn. The ratio of the number of each color in each urn is indicated in the figure.

Urn 1	Urn 2	Urn 3
5/8 Red	0 Red	1 Red
0 Blue	1 Blue	0 Blue
3/8 Green	0 Green	0 Green
1/2	3/8	1/8

Fig. 5

The probability of choosing each urn is the fraction under each urn in the figure. The process is as follows. First, we choose an urn, according to the probabilities given below each urn, and then we choose a ball from that urn according to the probabilities of the colors of the balls within that urn. The first question we could ask is, "What color ball from which urn is most likely to be chosen?" That is, we are asking about the color and the urn, jointly. To determine this we can plot the joint probability distribution as follows:

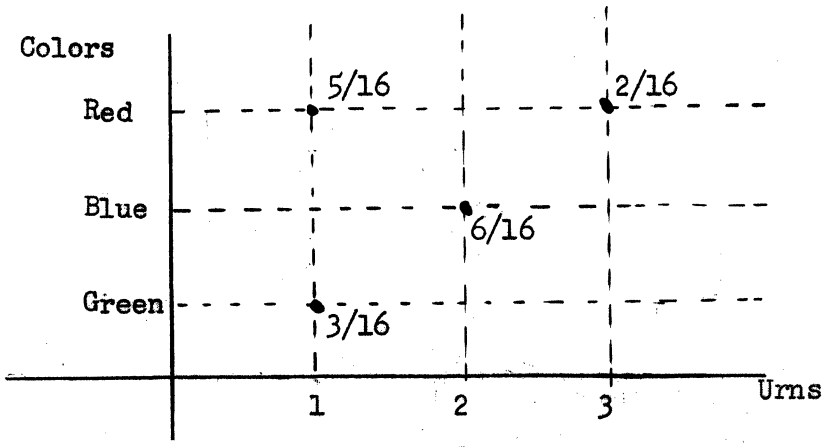

Fig. 6

The numbers next to each point are the probability of occurrence of that particular event. For instance, it is easily seen that the probability of drawing a red ball from the first urn is 1/2 x 5/8 = 5/16. Since the operations of choosing the urn and choosing the ball are independent, their probabilities are multiplied. This plot is the joint probability distribution of colors and urns. We see immediately that the most likely event is that of choosing the second urn and then a blue ball, which occurs with a probability of 6/16.

Next we could ask which urn is most likely to be chosen. This is obviously, from Fig. 5, the first urn which is chosen with a probability of 1/2. To get this from the joint probability distribution, Fig. 6, we sum along the direction of the colors; that is, we integrate over the variable we wish to eliminate. This gives us a one-dimensional distribution:

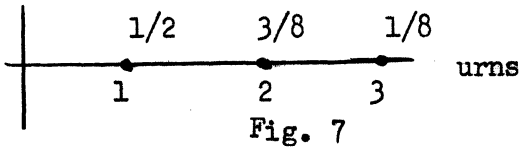

Fig. 7

where we see that our previous result is substantiated.

Now let us ask which color is most likely to be chosen. By summing along the direction of (urns) we again get a one-dimensional distribution:

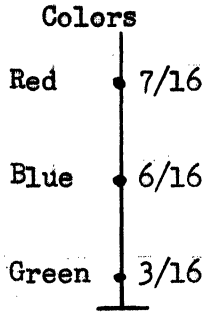

Fig. 8

where we see that the most likely color to be drawn is red with a probability of 7/16. From the result of the last two questions, that the most likely urn is the first and the most likely color is red, one must not then conclude that the most likely urn and color are one and red, for we saw that the correct answer to the joint occurrence was the second urn and the color blue.

From this simple example one sees that the answer he gets depends upon the exact question asked. In our particular example we first asked for the most probable joint distribution of the variables, X_1, X_2, X_3 and V_2. The reason one would desire this joint distribution is if he were actually intending to measure or use them, jointly. If, however, we are only going to use, say X_3, then we would not want to determine the most likely joint occurrence of all variables. From the example of the balls we see that

such a computation would give erroneus answers. In the case of the balls if one were to wager on the color of the next ball drawn, he certainly would not want to bet on the blue one since the red one is most likely. On the other hand if one were wagering on both the urn and the ball, then the blue color would be chosen, along with the second urn. This is the reason we wish to integrate over all unwanted variables in Eq. 10 and get a distribution in the variables in which we are interested. For tracking problems it is usually sufficient to find the joint distribution of the last value of X and of the last value of V. The value of V is important for prediction purposes.

In general, then, we will only retain the joint distribution of the last value of X and V, as our final output from the detector. The operation of selection will then use this distribution. It may turn out in more general cases, as we will see shortly, that the entire joint distribution among all the variables must be kept on hand for computational purposes within the detection process, even though it will be distilled down to the two variables X and V, as a final output. The reason for needing the complete distribution occurs when the values of acceleration in one sample interval depend upon some past values of the variable or its derivatives. If this is so, one must keep the distribution of these past variables on hand in order to be able to determine the distribution of the acceleration in each subsequent sample interval. This will be handled rigorously in a moment, but first we should point out, that while this may cause added complications, the principles that we used in the above example are the same.

Solution B. With the idea in mind that the only quantities in which we are interested are the last values of X and V we can try a slightly different approach to the sample problem. Since we are not interested in keeping the distribution of past values of X we need not even put them in the distributions, as such. It will be recalled that the distributions of the accelerations (@) were transformed into distributions in X. We will not do that this time. Instead we will include the values of (@) and then integrate over them, instead of changing to X and then integrating over the X. The reason for doing it this way is that a rather nice physical picture of the calculation process can be formed by this procedure. Exactly the same steps will be taken but in a different order.

First we take the first sample a_1. The distribution of X_1 about a_1 is the same as Eq. 7a. for only one piece of data; hence

$$W(X_1/a_1) = e^{-\frac{(X_1-a_1)^2}{2\sigma^2}} \qquad 15.$$

Since we have already determined that the final output of the detector will be a joint distribution of X and V, we could interpret the above distribution as that joint distribution where V is as yet uniformly distributed. Suppose we plot this as follows.

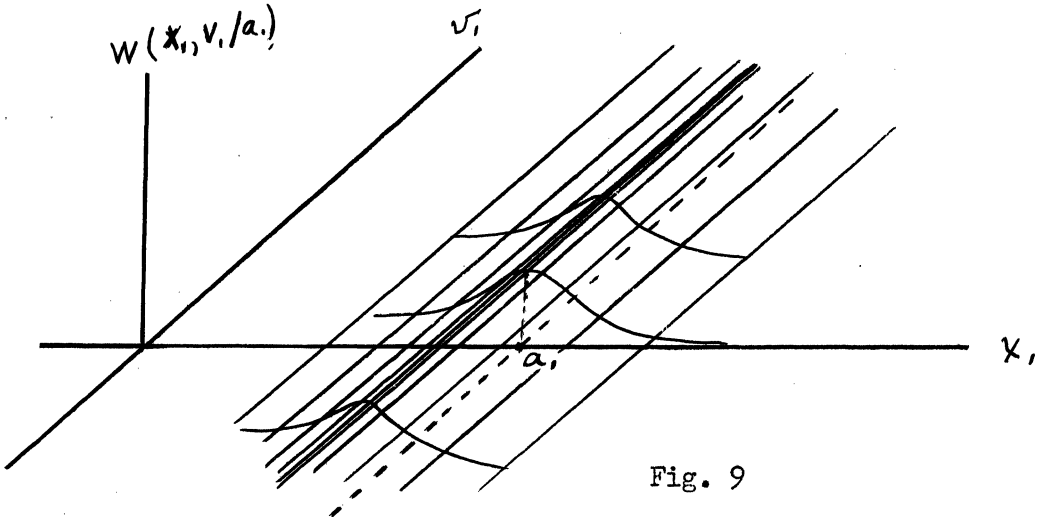

Fig. 9

We visualize this as a plot in three dimensions, the probability distribution $W(X_1,V_1/A_1)$ versus X_1 and V_1 . This is just the distribution obtained from one measurement without knowledge of the process being measured. Suppose now we were to try to predict the future value of X and V. We would have to take two factors into account. First, the object whose X coordinate is being measured may have any velocity, as indicated by Fig. 9. Second, this velocity may change, due to some acceleration, during the next sample interval. Let us handle these two separately.

For the moment let us assume that the acceleration is zero, and ask how the function changes when we try to predict the future value of X and V. Obviously, if the acceleration is zero, the velocity will not change. Also one sees that X will increase by the amount of the initial velocity. Thus to find the distribution function for the future (one sample) values of X_2 and V_2 we need only substitute in Eq. 15, V_2 for V_1 and $X_2 - V_2$ for X_1. Then we have a distribution of X_2 and V_2 for one sample interval in the future under the condition that the acceleration is zero. Regarding Figure 9, it is easily seen that this operation merely slides each V cross section to the right an amount V_1. Thus we get Figure 10.

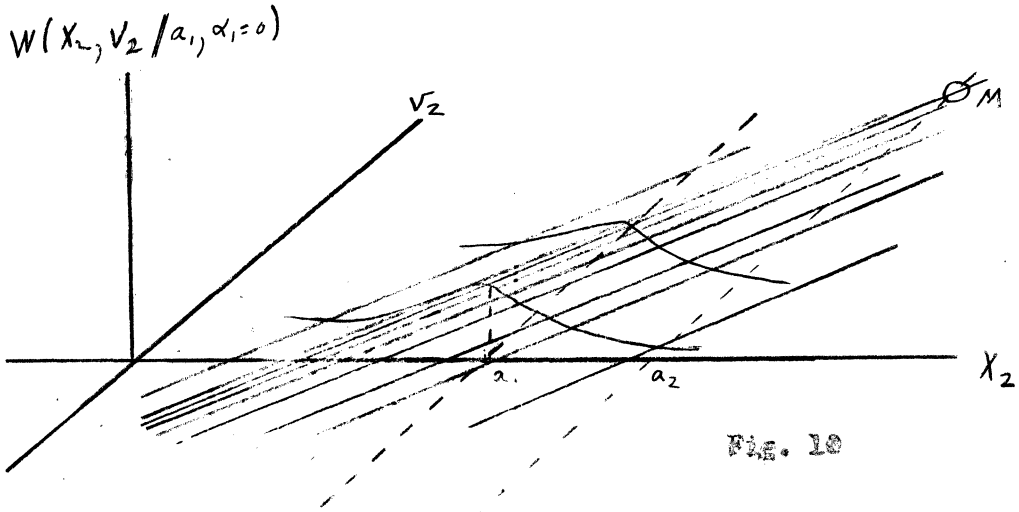

Fig. 10

This just twists the 'mound' to the right. If at this time we found out that the true value of X_2 were (a_2) we could just trace up the line $X_2 = a_2$ and would see that the distribution of V_2 along this line would be normal with the maximum value at $V_2 = a_2 - a_1$.

If the sample value (a_2) were not given to be an exact value of X_2, but were instead distributed similarly to a_1, we would as before multiply the distribution that we have, in Figure 10, by the normal distribution of X_2 and a_2. This would result in a mound as in Figure 11.

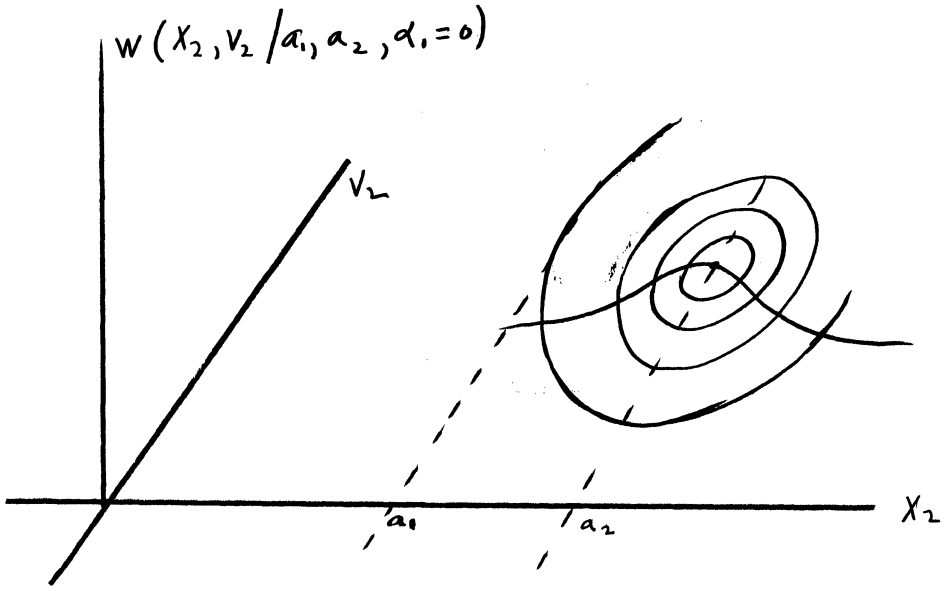

Fig. 11

This is a mound with a normal cross section in all directions $\perp$ to X_2, V_2 and the cross section parallel to X_2, V_2 is ellipsoidal. The peak of the mound is at $X_2 = a_2$ and $V_2 = a_2 - a_1$. This was formed under the assumption of zero acceleration. Let us go back now and see what effect possible accelerations would have on this distribution.

When we made the substitution of variable that led to Figure 10 we included the condition that some initial velocity could have existed at the beginning of the sample interval. Now we have also been given that the accelerations have a normal distribution for each sample interval. Suppose the acceleration is $@_1$. Then instead of V_1 being equal to V_2 we would have to substitute for V_1, $V_2 - @_1$. Also for X_1 we must substitute $X_2 - V_2 - 1/2\ @_1$. Now we could think of the distribution as a function of the three variables X_2, V_2, $@_1$, however, since it is not necessary to retain the information directly pertaining to $@_1$ we merely multiply this distribution by $W(@_1)$, which is given in Eq. 6. Then to clear the expression of $@_1$ we integrate over all $@_1$. We ask now, what has this done to the function of Figure 10? It is seen that whatever effect on the X dimension the effect will be twice as great on the V dimension because of the above substitutions. Actually this process of integration will be seen to be a convolution of $W(@_1)$ with $W(X_2, V_2\ a_1, @_1)$. This convolution has the effect of "smearing" the function along a line $X - 2V = \text{constant}$. To visualize this we think of perhaps rubbing our hand over the function of Figure 10 along the given line in such a way that the top of the function is diminished and the slopes of the sides are stretched out. This is the total effect of allowing the acceleration to be different from zero and normally distributed about zero.

Now again we can receive the second sample a_2 and multiple it as before. The only change in Figure 11 is that the mound is smeared out along the line $X - 2V = \text{constant}$.

After receiving the last sample we again make the same substitution of variable as before and the convolve just as before. We see that each time a new sample is obtained we do the same steps over and over again . In the cases where these processes can be carried out analytically we obtain a recursion relationship which enables us to handle as many samples as desired, indefinitely many, in fact. At each step one can determine by the same methods as before the most probably X and V, if this happens to be the process of Selection chosen.

We see in this process that each time a sample is received the distribution has its sides trimmed down and hence its peak sharpened up. Then as time goes on the smearing action of unknown accelerations tends to flatten out the distribution. From this process one can see how it is possible to determine how often it may be necessary to take samples to insure a distribution meeting certain standards. From this we get ideas of sampling rates as compared with the type of function being measured.

In the ordinary filter, a process of selection follows the detector and chooses a specific action based upon the distribution functions derived by the detector. By examining the distribution function we can describe the quality of the filter, after the selection process is completed.

When the derived process may be carried out analytically we get mathematical expression for the output of the detector. In case the process will not yield to ordinary mathematical manipulations one must have recourse to some type of approximate solution or perhaps numerical solutions. This has not been investigated thoroughly as yet, however.

The solution of the ideal detector problem is the most difficult part of the complete filter design as far as mathematics is concerned. The selection process which follows is usually very simple mathematically, but may be based upon some very subtle considerations. It is impossible to describe this process in much detail except for specific situations. In view of this, we leave the topic out of the general design method and consider the detector design as a large step toward complete filter design.

The General Case In the general case we have only two functions with which to deal. The first is the 'noise' distribution, analogous to the one written for the normal distribution in Eq. 1. We denote this distribution by

$$W_N\,(X_k/a_k) \qquad 16.$$

The second quantity is the distribution describing the process. This is the joint distribution of Eq. 4.

As was pointed out in the example, we are usually only interested in the last values of X and its first n-1 derivatives, rather than all values for all time. Thus, in order to facilitate the calculation of this restricted joint distribution we will rewrite Eq. 4 in such a form that it expressed the joint distribution of all the D^n and of values of X and its first n-1 derivatives in the last time interval. Thus we have:

$$W_P\,(X_k, D_k^1, \ldots D_k^{n-1}, D_1, D_2^n, \ldots D_{k-1}^n) \qquad 17.$$

The problem now resolves itself into this. We receive the first k samples and from the measurement process alone, since the measurements are independent, we may write:

$$W_1(X_1, X_2, \ldots, X_k/a_1, a_2, \ldots, a_k) = \prod_{j=1}^{k} W_N(X_j/a_j) \qquad 18.$$

Now since only the final values of X and its derivatives are of real interest for their own sake, we will, by algebraic manipulations, express each of the X_i in terms of X_k and the derivatives in the k'th interval and the n'th derivative in all of the intervals. Then the above expression may be considered as a joint distribution of the values of X and its first n-1 derivatives in the k'th interval and the n'th derivative in all preceeding intervals, conditioned on the reception of the first k samples, thus:

$$W_1 (X_k, D_k^1, \ldots D_k^{n-1}, D_1^n, D_2^n, \ldots D_{k-1}^n / a_1, a_2, \ldots a_k) \qquad 19.$$

This is the most we can do if the function Eq. 17 is not known. If Eq. 17 is given describing the process we multiply these two functions together since they are independent statements about the same variables. Then to eliminate the D^n, which are not of immediate interest we integrate over all D^n. This leaves us with a joint distribution involving only X and its derivatives in the k'th interval. This is the desired output of the detector.

To sum up the steps we see that:

1. We multiply the distributions (Eq. 16) together that describe the noisy sample values.

2. We make algebraic changes of the variables so as to express the variables with which we want to work.

3. We multiply by the joint distribution (Eq. 17) that describes the process being measured.

4. We integrate over all variables that are not of immediate interest.

We note here, in general, that it is necessary to receive a whole sequence of data before performing any integrations, since the functions may be dependent upon past values. If the integrations were carried out say for k pieces of data, some past information would be destroyed. Then it would be impossible to accurately determine what the exact distribution of the derivatives would be in the k-1 interval. In the special case we worked out first, the distribution of present values of acceleration did not depend upon past values of the other variables and thus we could work the problem step by step as the data arrived. It will be recalled this was the method used the second time (solution B) the problem was discussed, in order to obtain the graphs of the functions, Figure 9, 10 and 11.

Conclusions:

We see that the information needed for the detector design is in two parts. These are the probability density distributions of the function to be measured, and the noise contaminating the data. For the representation chosen, we obtained the complete solution of the detector performance. We see that one may draw inferences as to best sampling rates under certain conditions.

Signed W. I. Wells
W. I. Wells

Approved William K Linvill
W. K. Linvill

WIW/jmc/mrs