Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

SUBJECT: THE PHILOSOPHY OF STATISTICAL FILTER DESIGN

To: W. K. Linvill

From: W. I. Wells

Date: January 27, 1953

Abstract: The application of statistics and probability theory to the design of filters is discussed. The function of a general filter is split into two logical operations. These are called detection and selection. The process of detection is that of separating useful information from noisy data. The problem of selection is that of interpreting this information in the light of criteria that are dictated by the desired purpose of the filter. The role of probability theory is shown to be the foundation of the detection problem and may be of extreme importance in the problem of selection also. This paper solves no practical problems; its only purpose is to clarify the aim and basis of statistical filter design.

I.

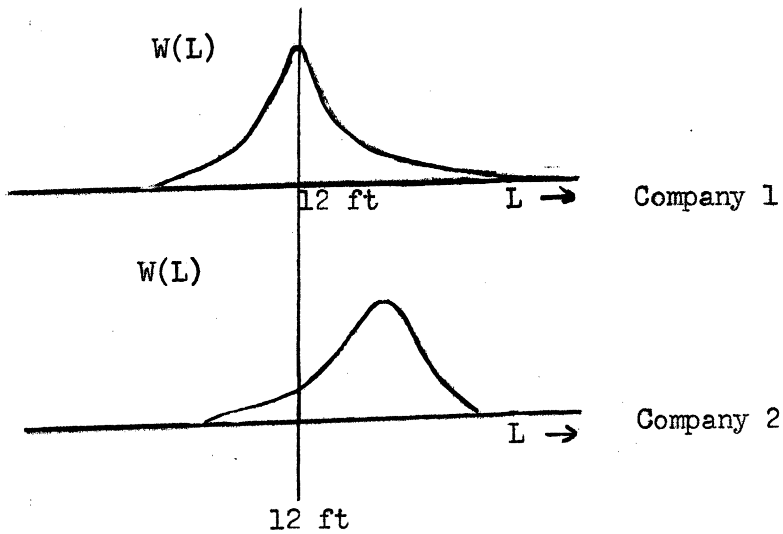
The object of this paper is to give the underlying reasons why statistics and probability theory play such an important part in the design of filters. By the term "filter" we mean any device that is meant to receive data from an outside source and to process this data, for some purpose, and to deliver this processed data to another outside user. In the special case of an electrical filter we have incoming data in the form of a voice wave, for instance. The purpose of the filter may be to reduce the high frequency content of the wave, or to reduce the noise content, and then to deliver the resulting wave to, say, a loudspeaker. In the case of a computer in use as a control element, data is supplied to the computer and the purpose is to perform certain needed calculations with this data and to deliver the results to the controlling elements in the system. As one can see, we are not restricting the term "filter" to linear electrical filters, or any particular special type.

Specifically, we wish to split the function of any filter into two basic functions. Then we wish to show how the ideas of probability theory are related to these two functions. The two basic functions are:

1. The separation of useful information from the data that is supplied from the outside source. (Detection)

2. To use this information, along with some given criterion to accomplish the purpose of the filter. (Selection)

We will illustrate these two functions with an example. Suppose we wish to construct a "filter" that will decide from which company we should buy a certain product. The problem is this. We need, for a certain construction job, steel rods that are exactly twelve feet long. We find that there are two manufacturers that produce these rods as a standard product and the price is the same from each. Now the question is, from which manufacturer should we buy. Obviously, the first step in deciding is to collect information. Suppose we sample the rods from each firm. We find that the steel is exactly the same in each case. The only difference is that, due to slight variations in the cutting machines, neither company produces rods that are exactly twelve feet long. By appropriately sampling the products we determine the distribution of lengths of rods that each company turns out. Suppose these look as follows:



We see that the first company has a fairly narrow distribution, and it is evenly distributed about the twelve foot length. The second company has a somewhat wider distribution, and it is skewed to the lengths longer than twelve feet. These two distribution curves constitute the pertinent information, but obviously we are not finished, since we still have to choose between companies.

In order to make our choice we must look for a desirable criterion. The application of this criterion to the problem of choosing is the second function of the filter. We remarked earlier that the rods had to be exactly twelve feet long. Now if we buy rods that are too short, they cannot be used at all, but, if the rods are too long we can cut them off and only lose the cost of cutting them. Thus, in this case the criterion is to choose in such a way as to buy the least number of rods that are too short. This then resolves itself into deciding which distribution curve has the least area under it to the left of the twelve foot line. The obvious choice must be the second company.

Thus we see that even though the first company has a closer control on its lengths, it could not be chosen because of limitations imposed by our own criterion. We note that the collection of information consisted in determining the distribution curves only, not in making a choice. We also note that the process of making the decision did not affect the manner in which the information was obtained.

The characteristics of this example are common to all filter problems. It is well-known from Information Theory that the best way to collect information from data is to constructe probability distribution curves analogous to those used in the example. The device that performs this function of filtering is called an "ideal detector."[1] The device that performs the second function, that of making the decision which accomplishes the purpose of the filter according to some criterion, will be called the "ideal selector." The reason for the prefix "ideal" is that this will constitute the best that we can do under the conditions of the criteria that we impose. If the information is not collected as well as is possible, then the detector is not ideal, and similarly if the decisions are made in a rough manner, then the selector is not ideal.

We have thus broken the process of filtering in that of detection and selection. Now we wish to examine each of these more closely, especially as they are concerned with probability theory. First we examine the detection procedure.

In the more usual sense, the problem of detection is the problem of separating useful information from given data that is corrupted with noise. The basis of construction of the ideal detector is that the probability distribution functions that are concerned with the quantity being measured are known and likewise that the probability distribution functions of the contaminating noise are known. From these distribution functions one is able to interpret the actual data in such a way as to construct the probability distribution curves of the received information about the quantity being measured. If the á priori (before reception of data) distribution curves are not known, then the process cannot be carried out. There are several reasons why one may not be able to determine these á priori distributions.

First, the distributions may just not be common knowledge and a great deal of collecting of data must be done before they can be determined. This is usually just a question of hard work and a lot of measurements.

If we go into the meaning of the probability distribution curve, we see where another difficulty lies. The probability that an event will happen is measured by watching a process for a long time and counting the number of times an event happens and dividing by the total length of time. Theoretically, the time interval should go to infinity. The question arises, what if a certain event happens frequently for a while

---

[1] R.M. Fano, Notes on Information Theory, M.I.T., Course 6.574, 1952 (not published.)

and then less frequently later on. This is the case of a non-stationary process. In a stationary process, one expects a given event to occur with the same probability at any time. In a non-stationary process one knows the probability that an event will occur is a function of time; that is, it may be more likely today than tonight or tomorrow.

Only if the process is stationary can we talk about á priori distribution functions that are independent of time. Thus, only if the process is stationary can we build a detector whose characteristics do not change with time. If the detector characteristics must change with time, we must either know beforehand how to change them, or must provide a scheme for learning how from the data itself. The process of learning from the data is the subject of much research. If the characteristics change in a known manner with time, the extension to time variable detectors is fairly clear.

When we speak of a stationary process, we evidently must be talking about some particular characteristics of the process, for it is entirely possible for some of the characteristics to be stationary and others non-stationary. For instance, if one is recording the results of a coin tossing game, the probability of a head or tail remains fixed throughout the game. But, if one records the total winnings of one of the players, this is a function that is non-stationary, and in fact its autocorrelation function is non-existent.

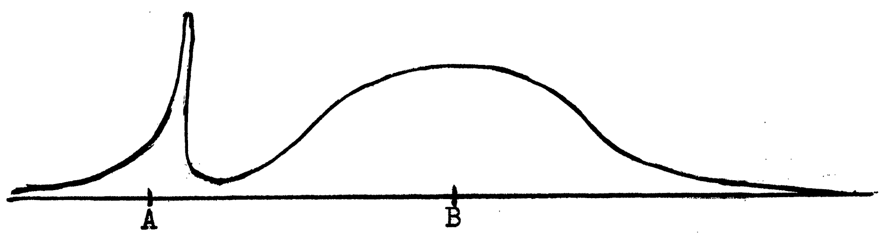
This brings up the question of what is needed for a process to be called stationary. We choose to say that if any characteristic of the process that is useful to the designer is stationary, then the process is called stationary. For instance, one may find several different properties of a process that may be of interest that may or may not be stationary. In this category we do not necessarily restrict ourselves to statistical properties. For instance, if a sine wave is being received, its frequency is constant for all time and thus this process is stationary, in that respect. Some of the properties we look for are:

1. Auto- or cross-correlation functions
2. Frequency
3. Probability distribution of magnitude of the function or one or more of its derivatives
4. Shape of pulses, (as in forms of pulse modulation), etc.

If some of these properties are stationary, we can base fixed detector design on these properties. If the time variation of some of the properties is known, we can base time variable detector design on these properties. If neither of these two possibilities is present, we may attempt to build a "learning" filter if some of the properties are "quasi" stationary. That is, they must be relatively fixed over long enough periods to allow the detector to "learn."

The problems associated with the selector are less well defined. In order for the selector to act, it must have two things at its disposal. First it must have a criterion with which to work. This is a mathematical statement of the purpose of the filter. Next, the probability distribution curves, that are the information, must be supplied by the detector. The only question that the selector has to answer is which value of the variable is most likely to be the most useful to us, under the conditions of the criterion imposed. We illustrate this with an example.

For instance, suppose one is going rabbit hunting. It turns out the grass in the field is tall and one only sees the rabbit when he jumps into the air while running. Thus we receive data on the position of the rabbit that is "sampled." We wish to shoot in such a way that we are most likely to hit the rabbit. As an illustration of how the different criterion may influence the action of the filter, we suppose that by taking all probabilities under consideration, our "detector" decides that the probability distribution of the predicted position of the rabbit looks as follows:



The question is now to decide, where to aim the gun to be most likely to hit the rabbit. The criterion is interpreted as follows. We assume first that we have a gun whose effectiveness is uniform over a certain width. That is, if the gun is a shotgun, it may be uniformly effective over a width of one or two feet depending on the range. If the gun is a rifle, the effective width is more like a half inch, neglecting the size of the rabbit. Now the object is to aim the gun so that its effective width will intercept the most area under the above probability distribution curve; then this maximizes the probability of a hit. It is quite clear that for the small effective area of the rifle one would have to aim at the most probable point, that is, at the highest point (point A), while if one were shooting a shot gun, he would aim at point B to intercept the maximum area. Here we see that the actual decision is influenced not only by the information (distribution curve) supplied by the detector but by other criteria supplied by factors that do not influence the detector at all.

The mathematical formulation of this problem could be done as follows. We devise a function that is a constant over the interval of effectivenss. We wish to find the place to put the center of this function under the constraint that the product of this function with the distribution function will yield a curve which bounds the maximum area.

In some cases it may be that the criterion may be expressed in terms of statistical quantities that may or may not come from the data that is being supplied to the detector. In these cases some similar procedure is indicated. It is seen, however, that in no case does the operation of the selector affect the ideal operation of the detector. The only influence that the selector could have on detector design is in a system where it is hoped to save money, materiel, or complexity in the construction of the detector because the selector is not too critically dependent upon the quality of the information supplied it. Even in these cases, however, it is seen that the process of selection must decrease in quality as the quality of detection drops.

II. Conclusion.

The filter process is broken into two steps. These are detection and selection. The process of detection is done on a probability basis. There is no other way to separate noise from useful information. The process of selection is based upon the information supplied by the detector and on a criterion determined by the purpose of the filter. Probability theory may play a large part in the selection function but it is not necessary in all cases.

If it is desired to build a fixed filter, this filter design must be based on the stationary qualities of the process. If certain of the qualities of the process are non-stationary but vary in a known manner with time, then it is possible to build a time variable filter which depends upon these qualities for its design. If there are qualities that are "quasi" stationary but not known as a function of time, it may be possible to design a filter that is able to "learn" as it goes.

Signed W. I. Wells
W. I. Wells

Approved William K. Linvill
W. K. Linvill

WIW/mrs