CURVES AS PARAMETERS, AND TOUCH ESTIMATION

JOHN W. TUKEY
PRINCETON UNIVERSITY
AND
BELL TELEPHONE LABORATORIES

1. Introduction

The aims of this paper are two-fold: to make clear that certain problems of interpretation which arise in estimating power spectra of stationary time series are entirely analogous to problems of interpretation which arise in carrying out two very elementary and classical statistical procedures, and to develop new concepts of estimation which clarify the interpretation of all three instances. En route, we shall have to give some attention to relation of asymptotic results to practical application.

The nature of the argument with which we are concerned is such that it seems best to begin with two of the most classical situations of statistics, unrestricted regression and the construction of histograms, and to develop the necessary new concepts and attitudes as we progress step by step, rather than to introduce the new formal concepts before motivation and illustration are available.

These concepts are stated, and the nature of the results obtained outlined, in the last section.

2. Regression

A simple regression problem may be described as (i) an attempt to "predict" a y from one or more of the x as well as possible, (ii) an attempt to estimate ave $\{y|x\}$ as well as possible, or (iii) an attempt to estimate some other conditional typical value, such as the conditional median, as a function of x. The differences between these three alternatives are far less significant than the distinction between three other alternatives which cut across the first three:

(1) the functional form to be used is simple, and is prescribed in advance,

(2) the functional form to be used is to be simple, but is not prescribed,

(3) the functional form is not prescribed, and need not be simple, though it is presumably continuous.

Research in part at Princeton University under Contract DA 36-034-ORD-2297, sponsored by the Office of Ordnance Research, U.S. Army.

It is with the last of these, which arises most usually in connection with (ii) or (iii) that we shall be concerned. It is well called *unrestricted* regression, and has been a familiar statistical problem since the early days of Karl Pearson.

3. Regressograms; the first instance

All statisticians who handle data know how to attack the simple case of this situation where y and x are both single real numbers. The x-axis is to be divided into suitable intervals, the mean of all the y-values corresponding to x-values falling in each given interval is to be found, the results are then to be plotted, either as points, each located above the center of the corresponding x-interval, or better as horizontal bars, each extending over the corresponding x-interval. (The heights of the points or bars are of course the corresponding y-means.) We may call the results of plotting bars a *regressogram.*

This is what to do, but why was it done? Can a reasonable aim be made precise enough so that mathematical statistics can study the relative effectiveness of different choices? What is a suitable division of the x-axis into intervals? How does it depend on the number of (x, y) pairs available? These questions deserve some attention.

4. The sampled case

Little has been said about the probability structure which is supposed to give rise to the pairs (x, y). So long as the y follow a definite conditional distribution for each x, the problem makes sense no matter how the x have arisen, whether they are fixed, are a random sample from some distribution, or were picked up from the street.

In the special case where the x are a random sample, one possible description of the aim of the regressogram can be given in terms of classical estimation. The mean of the y associated with a given interval may clearly be taken as estimating the mean over that interval, of the conditional average ave $\{y|x\}$. This description is easily seen to be unsatisfactory as a description of an aim. (Although it is often a very *satisfactory* description of what has been accomplished in the construction of a particular regressogram.) The interval averages "to be estimated" depend on how the x-axis was divided up, yet one of our problems was to compare the results of various ways of dividing up the x-axis. And if any practitioner who has made a regressogram is then given *many* more (x, y) pairs, perhaps 10 or 100 times as many, he will almost surely subdivide the x-axis more finely for the analysis of the more extensive body of data. If the problem is described as estimating interval means of conditional averages or of other conditional typical values, the problem changes far too easily when only the amount of data available is changed. No relatively constant aim has been identified, the problem is not in good shape for formal asymptotic treatment.

If we were to treat the regressogram the way that some treat the estimation

of power spectra, we should attempt to escape by talking about estimates of ave $\{y|x\}$ for individual x_0, and calling the difference between the value of ave $\{y|x\}$ at x_0 and its mean over the interval containing x_0 "bias." Few practicing statisticians would regard the estimation of ave $\{y|x_0\}$ for a single isolated x_0 as at all a reasonable problem, so long as x has some continuous distribution. Such a choice isolates an aim, but it is poorly chosen.

There is a need for something better.

5. Some formalities

A small amount of notation is now desirable. Let us set $g(x) =$ ave $\{y|x\}$, $\Gamma =$ the curve described by $(x, g(x))$, $I_k =$ the kth interval into which the x-axis is divided, $\bar{y}_k =$ mean of the y observed with x in I_k, then the bars which make up the regressogram are the sets

$$C_k = I_k \times (\bar{y}_k) = \{(x, y) \text{ with } x \text{ in } I_k \text{ and } y = \bar{y}_k\}. \tag{1}$$

The formal approach we are about to make needs no assumption that the x are in any way a sample. Thus it is desirable to work conditionally on a specific set of x-values. Let us set

$$\eta_k = \frac{\sum y \text{ for } x \text{ in } I_k}{\text{number of } x \text{ in } I_k} \tag{2}$$

and agree to understand, through section 7, that either "ave" or "average" refers to an average conditional on the given x. Then ave $\bar{y}_k = \eta_k$, and the bar which is the average of the possible bars C_k (each of which lies over the interval I_k) is

$$\Gamma_k = \text{ave } \{C_k\} = I_k \times (\text{ave } \bar{y}_k) = I_k \times (\eta_k). \tag{3}$$

6. Average bars; touch estimation

What can be said about the relation of the average bars Γ_k to Γ? Trivially

$$\min \{g(x)|x \text{ in } I_k\} \leqq \eta_k \leqq \max \{g(x)|x \text{ in } I_k\}. \tag{4}$$

Since $g(x)$ is continuous in I_k, and so must take on all intermediate values there, there is some x' in I_k for which $g(x') = \eta_k$. Expressed geometrically, the average bar Γ_k must have one or more points in common with the curve Γ.

While it might be more precise to say that Γ_k meets or intersects Γ, we shall say that Γ_k touches Γ when we mean that they have a common point or points. (There will be no implication of tangency when this is said.) We do this because "meet estimation" and "intersection estimation" are much less pleasant terms than "touch estimation."

It is thus natural to say that an individual bar C_k is an *unbiased* (in mean) *touch estimate* of the curve Γ. The classical procedure of dealing with unspecified regression leads to unbiased touch estimates. And there are few if any other properties of the classical procedure which hold in general.

7. Confidence procedures

It is natural to go farther and calculate limits, say, $\bar{y}_k \pm ts_k$ associated with the kth interval, where these limits are presumably to be calculated as if the y with x in I_k were a sample from a single distribution. The interval from $\bar{y}_k - ts_k$ to $\bar{y}_k + ts_k$ will be an approximate confidence interval for η_k. Moreover, its true confidence will exceed its nominal confidence in almost all circumstances.

If we were concerned with the conditional median of y rather than with its conditional average, we would have replaced $\bar{y}_k$ by the median of those y whose x fall in I_k. And if we had applied the sign test to those same y, we should have obtained a confidence interval which would surely have had at least the nominal confidence of containing at least one point of the interval from min $\{y|x \text{ in } I_k\}$ to max $\{y|x \text{ in } I_k\}$, and hence have had at least the nominal confidence of containing a value of $g(x)$ for some x in I_k.

Let us write J_k for an interval for which there is nominal confidence that it contains values of $g(x)$ for x in I_k. In practice, this nominal confidence is almost certain to be exceeded by the actual confidence, which depends upon the x and $g(x)$, though this need not always be so when dealing with conditional averages. It is now natural to introduce the blocks

$$D_k = I_k \times J_k = \{(x, y)|x \text{ in } I_k, y \text{ in } J_k\}, \tag{5}$$

for which we have the same confidence that D_k touches Γ, as we have that J_k touches $g(I_k)$ since either is equivalent to J_k containing a value of $g(x)$ for x in I_k.

Whenever this confidence is not merely nominal, it is natural to call D_k a *block touch estimate* with the corresponding *confidence*. Notice carefully that D_k is *not* supposed to surround or contain Γ, but *only to touch* it. If $g(x)$ should oscillate wildly, as well it may under our assumptions, its extremes for x in I_k may lie far above or below D_k for almost every sample.

8. Histograms; the second instance

The "estimation" of a probability distribution by the construction of a histogram is an even more ancient and elementary statistical procedure than unrestricted regression. Many histograms are drawn daily. They are thought to tell us about the probability density functions of the sampled populations. These probability density functions are thought of as smooth (though they may not be so in reality). How is the "estimation" of a continuous probability density function by a histogram, or by something easily developed from a histogram, best described precisely and formally?

The upper margin of each histogram is a bar, which is an unbiased touch estimate of the probability density curve in the sense just made precise. Here the histogram is exactly like the regressogram.

If the actual counted (sample) fraction associated with each histogram interval is replaced by appropriate binomial confidence limits for the population fraction,

each bar can be replaced by a corresponding block. And there will be at least the nominal confidence that each block, separately, touches the probability density curve. Thus we may replace the histogram by a set of block touch estimates with prescribed (or greater) confidence if we desire to do so.

Again, if the probability density curve oscillates wildly, as well it may, there will be no reason for the blocks to contain or surround the probability density curve. They will, in general, be no more or less than touch estimates. The situation is exactly the same as for unrestricted regression.

9. Histograms with bounded errors in x; the third instance

Suppose now that we are trying to study, by histogram methods, the probability density of ξ, where what is observed is $x = \xi + e$, and where e, though unknown for each individual x, is surely known to satisfy $|e| \leqq \delta$. How can we construct something close to a conventional histogram, and state clearly what properties are thus obtained?

We have selected this example because of the analogy with the instance we shall discuss next. Nevertheless, it includes as a special case a situation that could arise in practice, when originally continuous data has been lightly grouped into intervals of length 2δ, and where the grouping intervals are not appropriate histogram intervals.

Let us set

$$
\begin{aligned}
I_k &= k\text{th base interval of the histogram} = (\alpha_k, \beta_k), \\
f_k^+ &= \text{observed fraction of the } x \text{ with } \alpha_k - \delta \leqq x \leqq \beta_k + \delta, \\
f_k^- &= \text{observed fraction of the } x \text{ with } \alpha_k + \delta \leqq x \leqq \beta_k - \delta,
\end{aligned} \tag{6}
$$

for which

$$
\operatorname{ave} f_k^- \leqq P\{\xi \text{ in } I_k\} \leqq \operatorname{ave} f_k^+, \tag{7}
$$

so that the block estimates

$$
D_k = I_k \times J_k, \qquad J_k = \left[\frac{f_k^-}{\beta_k - \alpha_k}, \frac{f_k^+}{\beta_k - \alpha_k}\right] \tag{8}
$$

have an average (block) which surely covers the bar $\Gamma_k = I_k \times (\eta_k)$, where

$$
\eta_k = \frac{P\{\xi \text{ in } I_k\}}{\beta_k - \alpha_k} = \frac{1}{\beta_k - \alpha_k} \int_{\alpha_k}^{\beta_k} f(\xi)\, d\xi, \tag{9}
$$

where $f(\xi)\, d\xi$ is the probability element of ξ. By the mean value theorem and the continuity of $f(\xi)$, we have $\eta_k = f(\xi_0)$ for some ξ_0 in I_k. Hence Γ_k touches the probability density curve, and D_k is an unbiased touch estimate of this curve.

Notice that, while in simpler problems it sufficed to use bars to obtain an unbiased touch estimate, in this instance it was already necessary to use a block.

10. Extension to confidence

If we now turn to one-sided binomial confidence limits, we can replace f_k^+ by a one-sided upper confidence limit $f_k^{+\,+}$ for the corresponding true probability, and replace f_k^- by a corresponding lower confidence limit $f_k^{-\,-}$. The blocks

$$I_k \times \left[\frac{f_k^{-\,-}}{\beta_k - \alpha_k}, \frac{f_k^{+\,+}}{\beta_k - \alpha_k} \right] \tag{10}$$

will then be block touch estimates of the probability density curve of ξ with at least the chosen confidence.

11. The two sources of uncertainty

There are two sources of uncertainty in this example. One comes from the observation of only a finite number of x; the other comes from the lack of definiteness of the relation between x and ξ. In a typical practical instance we should be quite willing to treat the first source of uncertainty asymptotically; but it is quite unlikely that we should want to do this with the second. Either the loose linkage between x and ξ is to be faced up to, or it is to be neglected; asymptotic treatment is not good enough.

The existence of the second source of uncertainty is reflected in a variety of ways. These include:

(1) Our unwillingness to be asymptotic about all sources of uncertainty.

(2) The need to use a block, not just a bar, in obtaining an unbiased touch estimate.

(3) The need to form the blocks of confidence touch estimates by expanding the bar corresponding to the observations in *two* steps.

(4) Failure of these blocks to shrink to bars when an infinite number of obervations are available.

12. Estimating continuous spectral densities; the fourth instance

If we are concerned with the spectrum of a stationary time series which we may assume to have a spectral density $s(\omega)\,d\omega$ with a continuous density function $s(\omega)$, we are facing a situation which clearly resembles the previous situations in very essential respects, especially when the available data consists of a finite number of observations at discrete times. In each of the three cases we have a finite number of observations, and we seek to estimate a curve, about which we are only willing to assume that it is continuous. An infinite number of parameters would be required to specify the curve exactly, while only a finite number of observations are available. Thus direct estimation of each separate parameter cannot be expected. It should not cause surprise if the appropriate ways to treat the four situations should show much similarity, even if the relationship of observations to natural parameters is apparently less direct in the new situation.

13. Set-up and structure

If we have a stretch $\{y_t\}$ of observations at equally spaced times from one realization of the stationary time series, and if we take the equal spacing to be the time unit and the origin of y to be at the mean, then, for all relevant times

$$\text{ave } y_t = 0,$$

$$\text{ave } y_t y_{t+h} = \int_0^\pi \cos h\omega \, s(\omega) \, d\omega, \tag{11}$$

where "ave" indicates an average across the ensemble of possible realizations.

If $q(\omega)$ is an arbitrary polynomial in $\cos \omega$ of degree not greater than the length of the data (not greater than one less than the number of observations in the given stretch), then there will exist a quadratic function, Q of the observations for which

$$\text{ave } \{Q\} = \int q(\omega) s(\omega) \, d\omega. \tag{12}$$

If the time series (stochastic process) is Gaussian, approximations, whose accuracy is usually quite adequate, can be given for the variance of Q, for an equivalent number of (chi-square) degrees of freedom for Q, and hence for confidence limits (individually proportional to Q) for ave $\{Q\}$.

Even if the time series is not Gaussian, several equivalent stretches of data may be available, so that a value of Q can be calculated for each. The resulting sample of Q-values can then be used to set approximate empirical confidence intervals for ave $\{Q\}$.

In many situations, indeed, it is realistic to consider that the length of the available stretches is much more severely limited than the number of such stretches, which can be increased at the cost of money or routine effort.

14. Estimation

If three functions $q(\omega)$, $q^-(\omega)$, and $q^+(\omega)$ are such that, for $0 \leqq \omega \leqq \pi$,

(i) $q^-(\omega) \leqq q(\omega) \leqq q^+(\omega)$

(ii) $0 \leqq q(\omega)$

(iii) $\int q(\omega) \, d\omega = 1$

(iv) $q(\omega) = 0$ outside a subinterval I of $(0, \pi)$

(v) $q^-(\omega)$ and $q^+(\omega)$ are polynomials in $\cos \omega$ of degree less than the length of the data,

and if Q^- and Q^+ are particular quadratic functions of the observations corresponding to $q^-(\omega)$ and $q^+(\omega)$, then

$$\text{ave } \{Q^-\} = \int q^-(\omega) s(\omega) \, d\omega \leqq \int q(\omega) s(\omega) \, d\omega \leqq \int q^+(\omega) s(\omega) \, d\omega = \text{ave } \{Q^+\} \tag{13}$$

and the block $I \times [Q^-, Q^+]$ has an average which covers the bar

$$I \times (\bar{s}) = I \times \left[\frac{\int q(\omega) s(\omega)\, d\omega}{\int q(\omega)\, d\omega} \right]. \tag{14}$$

By a standard mean value theorem, this bar touches the curve $\Gamma = \omega, s(\omega)$.

Thus we have constructed an unbiased touch estimator of the spectral density curve.

In addition to this description, it may well be important to think of the interval $[Q^-, Q^+]$ as estimating the length of I multiplied by the average of $s(\omega)$ over I, where the averaging is according to $q(\omega)$. If this is not at all necessary, or if the importance is quite heuristic (as is usually the case) so that we can allow a blurred image of the behavior of $q^-(\omega)$ and $q^+(\omega)$ over I to serve as a substitute for $q(\omega)$, then we may drop $q(\omega)$ from our structure.

The essentials of the argument above, suitably modified, show that if, for $0 \leqq \omega \leqq \pi$,

(i′) $q^-(\omega)$ is $\geqq 0$ for ω in I and $\leqq 0$ for ω not in I

(iii′) $\int_I q^-(\omega)\, d\omega \leqq 1$

(i″) $q^+(\omega)$ is $\geqq 0$ everywhere

(iii″) $\int_I q^+(\omega)\, d\omega \geqq 1$

(v) $q^-(\omega)$ and $q^+(\omega)$ are polynomials in $\cos \omega$ of degree less than the length of the data,

and if Q^- and Q^+ are quadratic functions of the data corresponding to $q^-(\omega)$ and $q^+(\omega)$ respectively, then $I \times [Q^-, Q^+]$ is an unbiased touch estimate of the spectral density curve.

15. Extension to confidence

If we know how to construct an upper confidence limit Q^{++} for ave $\{Q^+\}$ and a lower confidence limit Q^{--} for ave $\{Q^-\}$, then the block $I \times [Q^{--}, Q^{++}]$ will be a block touch estimate of the spectral density curve with corresponding confidence. This will be possible in at least the two kinds of circumstances mentioned at the close of section 13. The two-step process of attaining a confidence touch estimate in this situation is strikingly similar to that discussed in section 10 above.

The choice of $q^-(\omega)$ and $q^+(\omega)$ will of course now need to be made so as to obtain the best compromise between a small average value for $Q^+ - Q^-$ and highly stable values for Q^- and Q^+.

16. Again two sources of uncertainty

In this instance, just as in the previous one, there are two sources of uncertainty with quite different properties. One comes from the observation of only a

finite total length of all realizations combined; the other comes from the observation of stretches whose lengths are all less than some bound. So long as we observe only stretches of bounded length, even if we observe enough (infinitely many) of them to determine the ensemble of realizations of this length precisely, the corresponding spectrum is *not* precisely determined.

If, for example, we know the ensemble of realizations of length 1000 precisely, we cannot show that the spectrum is not concentrated at any 1000 angular frequencies ω_i that may be chosen, so long as $\cos \omega_i \neq \cos \omega_j$ for $i \neq j$. Analogous results undoubtedly hold for spectra with continuous spectral densities. The uncertainty associated with finite length of stretch is not removed by gathering an infinite amount of data. And reluctance to treat the stretch length as asymptotically infinite is as appropriate in this instance as reluctance to treat the bound on the error as asymptotically zero in the previous instance.

Again, the existence of this second source of uncertainty is reflected in a variety of ways, including all four of those listed at the close of section 11 above.

17. Practical implications

How does the construction of touch estimates of the spectral density curve differ from present practices? If Q_0 is one of the quadratic functions usually associated with the interval I, and if ave $\{Q_0\} = \int q_0(\omega)s(\omega)\,d\omega$ with $\int q_0(\omega)\,d\omega = 1$, it has been the practice to take the interval from $(1 - c)Q_0$ to $(1 + d)Q_0$ as an approximate confidence interval for "a weighted average of the spectral density in I" where c and d depend suitably on the number of equivalent degrees of freedom assigned to Q_0. The proposed procedure is to take the interval from $Q^{--} = (1 - c)Q^-$ to $Q^{++} = (1 + d)Q^+$ as defining the vertical extent of a block touch estimate for the spectral density curve. The result of either procedure can be portrayed by a block in the (ω, s) plane.

When a logarithmic scale is used for $s(\omega)$, the heights of the blocks found by the older approach will be equal to one another. The new blocks will tend to be somewhat less stringent than the old blocks, to extend both somewhat higher and somewhat lower. This increase in uncertainty will become more and more marked as we move to the I whose contribution to the total power is smaller and smaller, reflecting the increased effect upon spectral estimates for such intervals of "leakage," of contributions from $s(\omega)\,d\omega$ for the ω outside of I.

In the case of the old blocks $q(\omega)$ must be a cosine polynomial of bounded degree, consequently it cannot vanish outside I, and w outside I can contribute to both the average and the variability of Q. To date such possibilities have usually been taken account of by mental comparison of the estimated spectrum with $q(\omega)$. The use of either an unbiased block touch estimate or a confidence touch estimate makes specific allowance for "leakage" and should provide answers of greater practical utility.

By introducing the concept of a touch estimate, it has been possible to

formulate what is being done with the new blocks in a way that is consistent with what is almost sure to happen if a sufficiently larger body of data is reanalyzed, namely a reduction in the width of the ω-intervals for which estimates are made. This makes it possible to subject the whole process to effective formal study, without the assumption, often quite doubtful in practice, that the spectral density curve is not only continuous, but quite smooth in the small.

While we have improved our capabilities for spectral estimation both in practice and in theory, in that

(1) We can provide estimates which include explicit allowance for leakage as well as for Gaussian variability, and

(2) we can describe what we are doing in a general framework compatible with an arbitrary amount of data,

we have not decreased either the great value of the recognition that the average value of any quadratic spectral estimate is the integral of an appropriate kernel against the spectrum, or the great value of knowing the kernel, or kernels, involved in whatever specific estimate, or estimates, we are using. Nor have we eliminated the problem of wisely choosing both what to try to estimate and what means to use to do this. Once a set of intervals $\{I\}$ have been chosen as the intervals above each of which we desire to have a block estimate, there will still remain the question of choosing the procedures intended to calculate these blocks. This question has become more ramified. Instead of choosing a $q_0(\omega)$, and then choosing Q_0 so as to be both easy to compute and almost as stable as possible, we now have the opportunity of choosing any $q^-(\omega)$ and $q^+(\omega)$ which meet the requirements of section 14 and then choosing Q^- and Q^+ according to the same principles that we used to choose Q_0. All the familiar considerations will still have to be balanced.

18. Fine structure

The basic complication which estimation of curves must face is fine structure. This probably arises more frequently in the estimation of spectral density curves than in the estimation of regression curves or probability density curves. Fortunately the statistical estimation of spectral density curves has an older brother; physicists have looked at spectra ever since Isaac Newton, and have taken pictures since the time that photography became feasible.

In the optical spectra of physics, spectral lines have fine structure, and hyperfine structure, yet when only a low resolution spectrum can be obtained, one may be fortunate to resolve the individual lines, to say nothing of the fine structure. In numerical spectral analysis, limitations of avilable data always limit resolution, often severely. In many of its applications we may doubt the existence of significant fine structure, but we are not likely to be sure that fine structure is absent. To talk of estimating $s(\omega_0)$ for a particular ω_0 is quite impractical in any such situation.

If there were practical situations where there was a real interest in estimating $s(\omega_0)$, as contrasted to learning what one can about the general run of the spectral density curve, the methodological problem would be severe. Fortunately, and reasonably enough, there seem to be no such situations. The fact that "fine structure" may cause $s(\omega_0)$ to oscillate quite wildly in narrow intervals can, fortunately, be neglected in most practical situations, *so long as* "narrow interval" is interpreted sufficiently strictly.

19. Common features of the instances

The most important common features of the various instances, regressograms, histograms, and spectrograms, are like the two faces of a coin. On the one side is the absence of specified form for the curve and the consequent need of an infinite number of scalar parameters for its complete description, on the other the well-recognized practice of the analyst of data to use finer intervals when he is furnished more data. The body of the coin is the fact of limited resolution: a finite amount of data can tell us only so much, it is unwise to ask it to speak on too many questions.

As statisticians we are used to living with this fact, but not to admitting it. The analysis of variance emphasizes main effects. One reason for this is precisely that there are times when one may learn more by asking a given body of observations fewer questions. But we rarely come out boldly and say "these are the estimated main effects, it is almost certain that there are smaller but meaningful effects in this situation, but with only this much data it is wisest to overlook them." Yet this is very often the position.

In dealing with a curve that may have fine structure, but which it is unwise to force into a few-parameter mold, it is harder to do that which should be done without admitting that one is doing it. If the price of doing right is that right-doing must be admitted, this price should be paid. In every case where we need to estimate a continuous curve which should not be restricted to some few-parameter form, we are going to need, I believe, to use touch estimation, and admit that this is what we are doing.

20. Asymptotic theory

When grasping a difficult problem, all of us find asymptotic theory easier, and it is natural that it should be much used. So far as a particular difficult problem is concerned, asymptotic theory is only useful (i) if it provides a useful approximation to an answer for, at least, certain of the amounts of data that arise in practice, or (ii) it provides useful guidance as to how one would deal with one aspect of a problem if all the other aspects were to be made negligible, perhaps because of the existence of large quantities of data. (Such guidance can of course only be partial, what is actually to be done will have to be chosen

on the basis of guidance drawn from various sources and related to various aspects.)

But if the guidance is to be at all relevant, it must be correctly related to the one aspect. There may be situations where quite unrealistic asymptotic situations give useful guidance, but it will generally be necessary for the asymptotic situation to be realistic for the guidance to be useful. Asymptotic treatments of means for dealing with outliers and other striking values, for example, are likely to be of little use if they do not recognize the inevitable modification of such techniques as the sample size increases and more knowledge becomes available about the nonextreme portions of the distribution. For in such instances the eventual consequence of more and more knowledge is a qualitative change in the technique of analysis. In other situations, the changes introduced by more knowledge need only be quantitative. Perhaps the simplest and clearest instance is Mann and Wald's treatment [1] of the optimum number of intervals for a chi-square goodness of fit test. Touch estimation may in due course prove to be another.

21. Alternatives

We have discussed the estimation of regression curves, probability density curves, and spectral density curves under the sole assumption that the curve considered is continuous. It is natural for many to respond with the reaction: "Isn't this assuming too little, surely in each particular situation it is appropriate to assume somewhat more, the exact nature and amount depending very much upon the situation. By assuming more you will be able to estimate the curves much more precisely." Such a response contains much truth, much temptation, and much danger.

It is certainly true, in each of the general situations discussed, that there is a place for estimating coefficients in a few-coefficient description of regression, distribution, or spectrum, that there are specific situations where the few-coefficient approaches are the best ones to take. But it is important to be clear about the natures of these specific situations. Some of them arise in relatively well-studied areas, where much experience has taught us which few-coefficient models will give satisfactory results. Others might arise where there is so little data that nothing else is reasonable. (These latter instances I believe to be not nearly as frequent as might appear.) In none of these specific situations, however, is it wise to restrict oneself *permanently* to few-constant fits. When an unusually large amount of unusually good data becomes available, or when a new theoretical suggestion produces a re-examination of accumulated data, it is likely to be appropriate to include an unrestricted examination as well as one or more restricted ones.

The temptation is to confine oneself to the relatively precise estimation of possibly irrelevant coefficients, and thereby to lose sight of the data, and of what they might show.

The danger is the loss of the most valuable information that the data might provide, which is most likely to be *qualitative* information, in the search for quantitative characteristics of the data which can be estimated with high precision.

There is surely a place for the few-coefficient analysis, an important place as measured in terms of frequency of use or in terms of contribution to routine measurement. (And these are important terms.) But there is also a place for unrestricted analysis, a place important in terms of new insights or in terms of new qualitative knowledge. We cannot do without either.

22. A terminological solution?

The logical unfortunateness of "touch" and the unsatisfactory sounds and appearances of its logically preferable substitutes were noted above. One possible solution deserves mention. If it turns out, after consideration and use by enough statisticians, that the concepts here discussed under the rubric of "touch estimation" are broadly useful and generally acceptable, then it might be possible to merely delete the word "touch" and speak only of "estimation." For the usual uses of estimation, as applied to one or a few parameters are merely special (if you like, degenerate) cases.

The only change in current usage that would be required would be the need to use some term more specific than "interval estimate" when describing confidence estimation. Even in the simplest cases, it is sometimes necessary to use intervals to carefully estimate (in the point-estimate sense) a single parameter. Thus if a survey using a probability sample yields the results: 65% "yes," 32% "no," 3% no response, the interval "65% to 68%" is the simplest expression of the results which we are sure is unbiased. (This interval is precisely an unbiased touch estimate.) This is "interval estimation" but with no prescribed confidence.

23. Summary of fundamentals of touch estimation

If a curve in a plane is to be estimated, if it is wise to assume the curve continuous, but unwise to restrict it to some few-parameter form, then estimation is best done using the data to determine the size, shape, and location of certain sets in the plane, which are regarded as trying to touch (that is, meet, intersect, overlap), not enclose, the curve. In the particular instances discussed here, the sets are most naturally bars (that is, horizontal line segments) or blocks (that is, rectangles with horizontal and vertical sides), but other shapes may be appropriate in other circumstances.

When the sets are such that average sets (averaged over repetitions on new data) can be easily defined, as is certainly the case for bars or blocks with prescribed horizontal extent, it may be possible to arrange that each average set

will surely touch the curve. In this event the original sets are called unbiased touch estimators of the curve.

In other situations, it may be possible to arrange that there will be at least a certain confidence that a given set will touch the unknown curve. If this is so, the set is called a confidence touch estimator.

Unbiased touch estimators and (often approximate) confidence touch estimators are provided for four situations:

(1) Observations (x, y) with definite unknown conditional distributions of y given x whose typical values (means, medians) define the regression curve to be estimated.

(2) Observations x drawn at random from an unknown distribution, whose probability density curve is to be estimated.

(3) Observations x, where $x = \xi + e$, $|e| < \delta$ given, and the probability density curve of ξ is to be estimated.

(4) Observations y_t made on a stationary time series at equally spaced times, where the corresponding power spectral density curve is to be estimated.

The question of criteria for comparing one touch estimate with another is not discussed, and offers a fertile field for suggestions and trials.

REFERENCE

[1] H. B. MANN and A. WALD, "On the choice of the number of class intervals in the application of the chi-square test," *Ann. Math. Statist.*, Vol. 13 (1942), pp. 306–317.