Fisher Consistency - the Evolution of a Concept:

It's Hard to get it Right the First TIme

Ken Gerow

BU-1022-M Cornell University April, 1989

Abstract

When Fisher first defined consistency in 1922, critical assumptions about restrictions on estimators were omitted from the definition. As a consequence, his definition appears to include both Fisher consistency and probability consistency, definitions which are taken to be distinct today. The evolution of the definition(s) of consistency is studied; the goal being to better understand this inconsistency. To further this aim, the conditions under which probability consistency and Fisher consistency are equivalent, and under which Fisher consistency and unbiasedness are equivalent (which is most of the time) are examined.

§1. Introduction

The development of the definition of consistency as we know it today has, for such a seemingly simple concept, a rather interesting history. When Fisher first defined consistency in 1922, critical assumptions about restrictions on estimators were omitted from the definition. As a consequence, his definition appears to include both what is now known as Fisher consistency (hereafter FC) and what is currently known as consistency (hereafter PC, for probability consistency), definitions which are taken to be distinct today.

The evolution from 1922 to 1956 (in Fisher's writings) of the definition(s) of consistency is studied, the goal being to better understand the nature of this inconsistency. To further this aim, the conditions under which PC and FC are equivalent, and under which FC and unbiasedness are equivalent (which is most of the time) are examined.

Central to this discussion is the use of the multinomial as a model for sampling (iid, sample size n) from any distribution. Relevant notation and a few basic results are presented in §2. It is difficult to appreciate the subtle ways in which the definitions of consistency have evolved, in particular the shortcomings of the early definitions, unless one has an understanding of the contemporary usage of FC and PC. Current definitions and some initial observations about FC and PC are discussed in §3.

Then, having provided a perspective from which to view the history, in §4 we review the pertinent literature to study the evolution of the definitions of FC and the (now) usual definition of consistency, namely PC. When it is appropriate with respect to the telling of the history, we include, as asides, discussion of comparative details of FC and PC. Finally, we discuss the lessons to be drawn from this history in §6.

§2. Mathematical Background.

When sampling (with finite sampling size) from a distribution $F_X(x)$, the resultant data are, because of limits on measurement precision, clearly discrete, no matter whether $F_X(x)$ is discrete or continuous. Thus a multinomial model (with theoretical bin probabilities determined by $F_X(x)$) is reasonable in that setting. We quote Basu (1988, p. 23):

"I hold firmly the the view that this contingent and cognitive universe of ours is in reality only finite and, therefore, discrete. In this essay we steer clear of the logical quicksands of 'infinity' and the 'infinitesimal'. Infinite and continuous models will be used in the sequel, but they are to be looked upon as mere approximations to the finite realities."

Taking the sampling procedure "to the limit" is potentially tricky; however, even in the limit, it is reasonable (in the sense of "reasonable: possessing of a large reality quotient") to suppose that a multinomial model still holds.

It is fairly easy to imagine the sample size increasing without limit; it is quite another thing to imagine shrinking the bin size (by improving the measurement process) infinitely. In the limit (as sample size increases to infinity) we may have a particular continuous distribution $f_X(x)$ which we wish to use to determine the multinomial probabilities, but it's reasonable to assume that we always have a discrete random variable to deal with. Consider the following example.

Suppose the measurement precision of our observations is δ; in other words, in taking observations, we map $[x-\delta, x+\delta)$ to x. Then our multinomial bins have, as centers, the set $\left\{\cdots, x-2\delta, x-\delta, x, x+\delta, x+2\delta, \cdots\right\}$, for any $x \in \mathbb{R}$.

If one presumes, for instance, that the true underlying distribution is normal with mean μ and variance σ^2, then one is in fact sampling from a multinomial distribution with cell probabilities determined by

$$\text{P[we observe x]} = \text{P}[\, x-\delta \leq x < x+\delta] = \int_{\frac{x-\delta-\mu}{\sigma}}^{\frac{x+\delta-\mu}{\sigma}} \phi(t)dt,\ \forall\left(x \in \mathbb{R}^1\right),$$

where $\phi(\cdot)$ is the standard normal pdf.

§2.1 **Empirical distribution functions.**

Without loss of generality, we let X be a non-negative, integer valued random variable from an arbitrary distribution $F_X(x)$. Define c(x) to be the number of occurrences of x in an iid sample of size n from $F_X(x)$. Then the empirical cdf may be written as

$$F_n(x) = \sum_{i=0}^{\infty} \frac{c(i)}{n} I_{[-\infty,\, x]}(i).$$

Further, the empirical pdf is just

$$f_n(x) = \frac{c(x)}{n}, \text{ defined for all non-negative integers.}$$

For example, consider the following random sample of n = 10 iid observations from a discrete U[1, 10] distribution.

$$\underline{x} = (4, 3, 2, 9, 5, 8, 5, 3, 8, 6).$$

$$= (2, 3, 3, 4, 5, 5, 6, 8, 8, 9), \text{ when ordered.}$$

Then, for instance,

$$F_{10}(5) = \sum_{i=0}^{9} \frac{c(i)}{10} I_{[2,\, 5]}(i) = 0 + 0 + 0.1 + 0.2 + 0.1 + 0.2 = 0.6 \ ,$$

and

$$f_{10}(5) = \frac{c(5)}{n} = 0.2 \ .$$

An important asymptotic feature of iid sampling from any arbitrary $F_X(x)$ is the result given by the Glivenko-Cantelli Theorem, which we state here without proof (for details, see Rohatgi, 1976, p. 300):

If $F_n(x)$ is the empirical cdf corresponding to an iid sample of size n from $F_X(x)$, then $F_n(x)$

converges uniformly to $F_X(x)$. That is, for $\epsilon > 0$,

$$\lim_{n \to \infty} P\left\{ \sup_{\infty < x < \infty} \left| F_n(x) - F_X(x) \right| > \epsilon \right\} = 0.$$

Thus, in the sequel, we may make reference to $F_n(x)$ converging to $F_X(x)$ as n goes to ∞ and be assured that this convergence is guaranteed.

§2.2 The disguises of a statistic

When Fisher discussed consistency, he usually considered only statistics which were functionals of the empirical cdf or pdf. We commonly think of statistics as functions of the observations. To aid in the switch from this mindset to the Fisherian one, we offer the following discussion and example.

First, we note that it is appropriate to view the (empirical or true) pdf (or, equivalently, the cdf) as a vector. Then, considered as a functional, the statistic maps points in an infinite vector space to points on the real line. For instance, considered as a vector, the pdf $f_X(x)$ (of an integer valued non-negative random variable) is the vector

$$(\pi_0, \pi_1, \pi_2, \pi_3, \cdots),$$

while the cdf is

$$\left(\pi_0, \sum_0^1 \pi_i, \sum_0^2 \pi_i, \sum_0^3 \pi_i \cdots\right).$$

For the sample of 10 observations from U[1, 10] considered above, the empirical pdf is

$$\begin{aligned} f_{10}(x) &= \left(0, 0, p_2, p_3, \cdots, p_8, p_9, 0, 0, \cdots\right) \\ &= \left(0, 0, 0.1, 0.2, 0.1, 0.2, 0.1, 0, 0.2, 0.1, 0, 0, \cdots\right), \end{aligned}$$

while the empirical cdf is

$$F_{10}(x) = \left(0,\, 0,\, p_2,\, \sum_2^3 p_i\,, \cdots,\, \sum_2^8 p_i,\, 1,\, 1,\, \cdots\right)$$
$$= \left(0,\, 0,\, 0.1,\, 0.3,\, 0.4,\, 0.6,\, 0.7,\, 0.7,\, 0.9,\, 1,\, 1,\, \cdots\right).$$

For example, let us examine a common statistic, the sample mean. We will show it in its usual guise as a function of the observations (1), then transform it into a function of the the empirical pdf (2), and finally present it as a function of the empirical cdf (3).

Without loss of generality, we restrict attention to a random variable which takes on non-negative integer values. As above, we define c(x) to be the number of occurrences of x in the sample. Then

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$= \sum_{i=0}^{\infty} i \cdot \frac{c(i)}{n}$$

$$= \sum_{i=0}^{\infty} i \cdot f_n(i) \tag{2}$$

$$= \sum_{i=0}^{\infty} \left(1 - F_n(i)\right) \tag{3}$$

The steps to get from (2) to (3) are somewhat tedious, but not difficult. The formulation of the sample mean as a functional of the empirical cdf is a standard result. See, for instance, Mood, et al, 1974, pp. 64-65.

§3. Fisher Consistency and Probability Consistency: an introduction.

§3.1 Definitions

Currently, Fisher consistency is a little used concept, appearing mainly in historically oriented papers such as this one or, for example, Savage (1976), or Feinberg, et al (1980). We begin with current definitions of FC and PC. Geisser (1980, p. 60) defines FC in terms which are readily accessible:

> "A statistic is consistent (Fisher consistent) if, when calculated from the whole population, it is equal to the parameter describing the probability law. Thus for every n let $F_n(x)$ be the empirical distribution function and let $T_n = g(F_n(x))$ be an estimator of θ. Then T_n is Fisher consistent (F.C.) for θ if $g(F(x|\theta)) = \theta$."

With respect to the definition of PC, we would now say, mathematically, that the sequence of estimators T_n of a parameter θ, based on a sample of n is consistent (in probability) for θ if for an arbitrary $\epsilon>0$,

$$P\Big(|T_n - \theta|>\epsilon\Big) \to 0 \text{ as } n \to \infty.$$

To see quickly that these definitions are in fact different, consider the statistic $T = \overline{x}$

$+ \frac{1}{n}$. This statistic is PC, but not FC in the current setting, where X is a discrete random variable. The main reason for this is that T is not strictly a functional of the empirical df, but a functional of that and sample size. FC specifically precludes consideration of such statistics. Note, however, that if we were entertaining a mythical continuous random variable, then n would be the number of jumps in the empirical cdf (since ties couldn't happen), and thus a function of it.

§3.2 Initial Observations

There are several features of these definitions worth noting. Fisher consistency is more than an asymptotic result. Thus, when it is applicable, FC is a stronger condition than PC. It demands virtuous behaviour even in the finite sample setting. Of course, the main weakness with PC is that it is primarily an asymptotic result; it is useless for the development of a theory of small samples. For instance (with, say x_i iid $N(\theta, 1)$, where $\theta \gg 0$, suppose we define $T_n = 0$ for $n = 1, 2, \cdots, 10^{10000}$; thereafter define $T_n = \overline{x}$. Then T_n is PC, but not very virtuous in any practical sense.

We mentioned above, with reference to FC, the phrase, "when it is applicable". In order for a statistic to be FC, two conditions must apply. First, it must be that we are doing iid sampling from some distribution. In other words, there must indeed be an $F(x|\theta)$ to plug into the function $g(\cdot)$ in order to be able to determine consistency. Thus one weakness of FC is that it is restricted to iid sampling. It cannot be used, for instance, in situations such as studies of time series, where observations are not independent. For further discussion, see Savage (1976).

The second condition which must be satisfied in order for FC to apply is that the statistic must be an explicit functional of the empirical cdf (or, equivalently, the empirical pdf). That is why, for instance, $T = \overline{x} + \frac{1}{n}$ fails to be FC. This condition is crucial to the finite setting strength of FC, since it restricts attention to statistics which are well behaved in finite samples. FC defines as inconsistent a wide range of statistics whose finite sample properties are not desirable.

In sum, PC has wider applicability, but FC is a stronger finite sampling condition.

§4. History and evolutions of definitions

Fisher discusses two definitions of consistency, the first of which (chronologically) bears his name today. In the definition numbering in what follows, definitions 1.1, 1.2, $\cdots$, are various wordings of the definition of FC (showing its evolution), while definitions 2.1, 2.2, $\cdots$, refer to the evolution of the definition of PC. When it is convenient, we will refer to FC as definition 1 and PC as definition 2 (meaning, in both cases, the respective collections of definitions).

As we proceed through the history, we'll occasionally pause for side discussions on various relationships between FC and PC (and unbiasedness) which will, hopefully, illuminate the unfolding of the history. The history presented here is in large part found in L. J. Savage's 1976 paper, "On Rereading R. A. Fisher".

§4.1 Beginnings

Historically, the first definition went like this (**definition 1.1**) (Fisher, 1922, p. 309):

> "A statistic satisfies the criterion of consistency, if, when it is calculated from the whole population, it is equal to the required parameter."

Later in the same paper, he describes a consistent estimator thusly (p. 316) (**definition 1.2**):

> "That when applied to the whole population the derived statistic should be equal to the parameter".

What does, "when applied to the whole population", mean, anyway? Fisher meant that the consistency criteria can be applied in one of two ways. One is to compute $g(f_X(x))$, where $f_X(x)$ is the true pdf one is sampling from. The other is to observe the sequence T_n. For T_n

to be consistent, the limiting value of the sequence T_n should be the required parameter. By one criterion (the $g(f_X(x))$ one), consistency isn't just a limit property of a sequence of estimators, as we think of it today. The second criterion does indeed imply that consistency is a property of a sequence of estimators.

Left unsaid at this point (see Kallianpur and Rao, 1955) was that the statistic T is explicitly a continuous functional of the empirical pdf: $T = g(f_n)$. We shall see later how this hidden assumption plays a critical role in the evolution of the concept of consistency.

The next definition he used is more recognizable as the definition of consistency (PC) we are familiar with today (definition 2.1) (Fisher, 1925, p. 702):

> "A statistic is said to be a consistent estimate of any parameter, if when calculated from an indefinitely large sample it tends to be accurately equal to that parameter."

At this point, however, he still would not have distinguished between FC and PC. He was, from his perspective, simply discussing consistency. The phrase, "from an indefinitely large sample", operationally was meant to include use of either of the two criteria described above. Compare this definition to a current definition of PC and of FC (Geisser's definition in §3, for instance).

As noted above, it is difficult to differentiate between definitions 1.1 and 2.1. In fact, for many years, Fisher took his two definitions to be describing the same thing. Thirteen years after the first definition appeared, Fisher had this to say about consistency (1935, p. 42: note his clever definition of inconsistency!) (**definitions 2.3 and 1.3**, respectively):

> "First, we may distinguish consistent from inconsistent estimates. An inconsistent estimate is an estimate of something other than that which we want an estimate of. If we choose any process of estimation, and imagine the sample from which we make our calculations to increase without limit, our estimate will usually *tend*, in the special sense in which that word is used in statistics, to a limiting value, which is some function of the unknown parameters. Our method is then a consistent one for estimating this particular parametric function ⟦PC⟧[1], but would be inconsistent for estimating any different function. The limiting value is easily recognised by inserting for the frequencies in our sample their mathematical expectations ⟦FC⟧[1]."

Fisher's definition of consistency here still includes both PC and FC. There are in fact circumstances under which FC $\Leftrightarrow$ PC. This is an appropriate place for a side excursion to explore the properties of FC and PC in a little more detail. In particular, we explore the conditions under which FC $\Leftrightarrow$ PC (§4.2) and under which FC $\Leftrightarrow$ unbiasedness (§4.3).

[1]⟦·⟧ represents an addition by the present author to the quote

§4.2 Coincidence of FC and PC:

Many of our commons statistics are such that they are both FC and PC. We demonstrate this by the following three propositions.

Proposition 4.2.1. If the estimator of θ, the statistic T_n, is a continuous functional of the empirical cdf $F_n(x)$, then FC and PC coincide.

Proof:

To see this, suppose that $T_n = g(F_n(x))$ is FC. FC says that $T_\infty = g(F_X(x)) = \theta$. Note that, by continuity, $F_n(x)$ close to $F_\infty(x)$ $(= F_X(x))$ implies T_n close to T_∞. Recall the Glivenko-Cantelli Theorem of §2.1: $F_n(x)$ converges uniformly to $F_\infty(x)$. Thus, as n goes to ∞, $F_n(x)$ converging to $F_\infty(x)$ implies T_n converges to $T_\infty = \theta$, by which we see that T_n is PC. In the other direction, suppose that T_n is PC. Then, by definition, as n goes to ∞, T_n converges in probability to θ. But, as n goes to ∞, $F_n(x)$ converges to$F_\infty(x)$. Continuity then implies that T_n converges to T_∞, so $T_\infty = \theta$, and T_n is FC.

□

Many of our commonly used statistics are in fact continuous functionals of the empirical distribution function, hence are such that, if they are FC, they are also PC, and vice versa. The following two propositions lend credence to this statement.

First, we need some notation and definitions to carry this discussion. Consider two random variables X and X^* of sample sizes n and n^* from distributions $f_X(x)$ and $f_X^*(x)$, respectively. Without loss of generality, we assume that X and X^* may take on positive integer values only.

Define

$$m = \min\left\{x_{[1]}, x^*_{[1]}\right\}, \text{ and}$$

$$M = \max\left\{x_{[n]}, x^*_{[n^*]}\right\}$$

Define

$$\|f_X\| = \sum_{i=0}^{\infty}\left|f_X(i)\right| .$$

This = 1, for f_X itself; of greater interest is

$$\|f_X\text{-}f_X^*\| = \sum_{i=0}^{\infty}\left|f_X(i) - f_X^*(xi)\right|$$

$$= \sum_{i=m}^{M}\left|f_X(i) - f_X^*(i)\right|$$

Suppose we are considering a statistic of the form

$$T = \sum_{i=1}^{n} w(i)f_n(i)$$

$$= \sum_{i=0}^{\infty}\Big(c(i)w(i)\Big)f_n(i),$$

namely a linear combination of the frequencies.

Proposition 4.2.2: If the statistic can be written as a linear combination of the observed frequencies, then it is a continuous functional of the empirical distribution function.

Proof:

Given $\epsilon > 0$, let $\delta = \frac{\epsilon}{w_*}$, where w_* is defined to be $\max_{i \,\varepsilon\, [m,\, M]} \left\{ w(i) \right\}$.

Suppose $\left\| f_n - f_n^* \right\| \leq \delta$.

Further, define $\delta_i = \left| f_n(i) - f_n^*(i) \right|$ so that $\sum_{i=m}^{M} \delta_i = \delta$.

Then

$$|T - T^*| = \left| \sum_{i=m}^{M} w(i)\Big(f_n(i) - f_{n}^{*}{}_{*}(i) \Big) \right|$$

$$\leq \sum_{i=m}^{M} w(i) \left| f_n(i) - f_{n}^{*}{}_{*}(i) \right|$$

$$= \sum_{i=m}^{M} w(i) \delta_i,$$

$$\leq \sum_{i=m}^{M} w_* \delta_i,$$

$$= \epsilon,$$

so T is continuous.

□

The following proposition shows that many common statistics can be written as a linear combinations of the frequencies, so that, for these statistics, the above proposition holds.

Proposition 4.2.3: A statistic which is a sum of an arbitrary function evaluated at each observed x can be written as a linear combination of the frequencies. (Fisher (1935) asserted that this is true if the function in question is continuous, but we see here that it needn't be.)

Proof:

$$T = \sum_{i=1}^{n} g(x_i), \text{ for arbitrary } g(\cdot),$$

$$= \sum_{i=0}^{\infty} c(i)g(i)$$

$$= \sum_{i=0}^{\infty} f_X(i)\left(n \cdot g(i)\right),$$

a linear combination of the observed frequencies, as asserted.

□

What we have shown here is that if the statistic is a continuous functional of the empirical pdf, then FC ⇔ PC. Then followed two propositions which demonstrated that many of our usual statistics are such that they are continuous functionals of the empirical pdf, hence are such that FC ⇔ PC.

In fact, it is easy to see that, for instance, maximum likelihood estimators are always FC, since they are explicitly functionals of the likelihood function, which is just an arbitrarily scaled pdf (see Rao, 1962, p. 83 for further discussion).

§4.3 Coincidence of FC and unbiasedness

This discussion is built around an example and discussion in Fisher (1935, pp 142-144).

Let F_X be the cdf of a multinomial with some (perhaps infinite) number of classes. Let π_i be the probability of occurrence of class i ($\sum \pi_i = 1$), where the π_i's are a (known) function of the (potentially vector valued) parameter Π. Let the observed proportions from a sample of size n be $p_i = \frac{n_i}{n}$ ($\sum n_i = n$). Suppose the parameter of interest is $\theta = g(\Pi) = \sum w_i \pi_i$. We use as our estimator $T_n = \sum w_i p_i$.

Two observations may be made. First, T_n is FC. Second, since it is a linear combination of the frequencies (with the correct weights!), T_n is unbiased for θ. Further, as seen in §4.2, functions of the form $\sum c(x_i)$ can be reformulated as a linear combination of the frequencies p_i and thus are both FC and PC. Now we see that these functions are also unbiased for the appropriate linear combination of the π_i's (recall Proposition 4.2.3).

Several authors (see §3) have pointed out that Fisher meant to restrict his discussion to estimators which are linear functions of the observed frequencies, hence are both FC and unbiased. He later relaxed this condition, so that, now, FC and unbiasedness are not equivalent. Consider the following example (Fisher, 1956, p. 143):

Suppose $\theta = \exp\left(\sum w_i \pi_i\right)$ is the parameter of interest. Then $T_n = \exp\left(\sum w_i p_i\right)$ is FC, but biased for θ. In a related case, though, $\ln(T_n)$ is both FC and unbiased for $\ln(\theta)$.

§4.4 Consistent Consistency

Kallianpur and Rao (1955) report (from conversations with Fisher) that Fisher had in mind only analytic functions of the frequencies. Thus, it seemed that he intended, at least at first, to restrict (for consistency purposes) consideration to estimators for which FC $\Leftrightarrow$ PC. They say (p. 333):

> "It, thus, clear ⟦that⟧ by consistency Fisher had in mind both the properties of the statistic tending to a limit in probability and the limiting value being attained by the statistic as the expected value of the frequencies."

Eventually Fisher differentiated between FC and PC by making explicit and relaxing his assumptions about the form of the estimator. At this point, he decided that consistency should be what we now call Fisher consistency. Effectively, he discarded PC as a useful definition of consistency in favour of FC. From Fisher (1956, p. 144) (definitions 1.4 and 2.4, respectively):

> "A Consistent Statistic may then be defined as: a function of the observed frequencies which takes the exact parametric value when for these frequencies their expectations are substituted. This definition is

applicable with exactitude to finite samples.

A much less satisfactory definition has often been used, namely that the probability that the error of estimation exceeds in absolute value any value ϵ, or, symbolically,

$$\Pr\left\{|T - \theta| > \epsilon\right\},$$

shall tend to zero as the size of the sample is increased, for all positive values ϵ, however small."

In the discussion preceding this quote, Fisher has given up some of the restrictions he previously had placed upon estimators (in order, it seems, to be able to differentiate between FC and PC more easily). In particular, he gives an example in which the estimator is FC, but is not a linear combination of the observed frequencies (see §4.3 discussion of FC and unbiasedness). Hence, by 1956, the definitions of FC and PC were clearly distinguishable.

Godambe (1976), citing Fisher (1935, pp. 45-46, and 1956, pp. 142-143) asserts that Fisher used the concept of linear estimating equations to define FC. The definitions Fisher used (definitions 1.3 and 1.4) do not bear this out. Further, in the 1956 discussion, Fisher used as an example of FC an estimator which was not a linear estimating equation (the $\theta = \exp(\sum w_i \pi_i)$ example).

The definition of FC has been formalized by Rao (1962, p. 82) as (**definition 1.5**):

"A statistic is said to be FC for a parameter θ if

(1) it is an explicit function of the sample distribution function S_n, (or the observed proportions $[p_1, \cdots, p_k]$ in the case of a multinomial), and (2) the value of the function reduces to θ identically when S_n is replaced by the true distribution function $F(\theta)$, (or the true proportions $[\pi_1(\theta), \cdots, \pi_k(\theta)]$ in the case of [[a]] multinomial."

Another, more recent formalization is offered by Savage (1976, p. 454) (definition 1.6):

"A FC estimate is mathematically a functional defined on distributions that coincides with the parameter to be estimated on the family of distributions governed by the parameter. Employed as an estimate, this functional is applied to the empirical distribution of a sample of n independent drawings from an unknown distribution of the family."

For a definition of FC which is in terms slightly more concrete that those of Savage, we recall the definition of Geisser (1980, p. 60):

"A statistic is consistent (Fisher consistent) if, when calculated from the whole population, it is equal to the parameter describing the probability law. Thus for every n let $F_n(x)$ be the empirical distribution function and let $T_n = g(F_n(x))$ be an estimator of θ. Then T_n is Fisher consistent (F.C.) for θ if $g(F(x|\theta)) = \theta$."

§6. Summary

When one is a beginning student of statistics, it is easy to become intimidated by reading journal articles where (hopefully!) results are presented in a fairly polished form. In particular, it is often difficult to see the true methods by which the results were obtained, since the required evidence (piles of scrap paper, broken pencils, etcetera) often has been sent out with the trash.

Consistency (meaning PC) is a relatively simple idea, one students are commonly introduced to at the senior undergraduate level. It is illuminating to realize that it took, at least as far as the published evidence indicates, Fisher 34 years to polish the definitions of consistency to their present form. In the early goings, he left unstated critical assumptions about the form of estimators under discussion. It appears that Fisher himself was not aware of the importance of these assumptions for quite some time. When he did become cognizant of the consequences of these assumptions, he quickly relaxed them sa as to better differientate between "his" form of consistency (FC) and what he considered to be a much poorer definition , namely probability consistency.

As a student of statistics (and we all are, to varying degrees, and at different stages), we may draw a heartening moral from this story. One shouldn't expect in general that new ideas, however simple they seem, will necessarily resolve themeselves easily or become polished quickly. Be patient, and keep on plugging.

Bibliography

Basu, D. (1988). In Ghosh, J. K., (editor). Statistical Information and Likelihood: a Collection of Critical Essays by Dr. D. Basu. Lecture Notes in Statistics 45. Springer-Verlag, New York.

Feinberg, S. E. and D. V. Hinkley, eds. (1980). R. A. Fisher: an Appreciation. Springer-Verlag Lecture Notes in Statistics 1.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London, Series A, v. 222, pp. 309-368. Paper 10 in CMS[2].

Fisher, R. A. (1925). Theory of statistical estimation. Preceedings of the Cambridge Philosophical Society, v. 22, pp. 700-725. Paper 11 in CMS[2].

Fisher, R.A. (1935). The logic of inductive inference. Journal of the Royal Statistical Society, v. 98, pp. 39-82. Paper 26 in CMS[2].

Fisher, R. A. (1950). Contributions to Mathematical Statistics. Wiley, New York.

Fisher, R. A. (1956). Statistical Methods and Scientific Inference. Oliver and Boyd, Edinburgh.

---

[2]CMS is Fisher's 1950 book: "Contributions to Mathematical Statistics".

Geisser, S. (1980). Basic theory of the 1922 mathematical statistics paper. In Feinberg, S. E. and D. V. Hinkley, eds. R. A. Fisher: an Appreciation. Springer-Verlag Lecture Notes in Statistics 1.

Godambe, V. P. (1976). Discussion of "On Rereading R. A. Fisher". Annals of Statistics, v. 24, pp. 73-101.

Kallianpur, G. and C. R. Rao (1955). On Fisher's lower bound to asymptotic variance. Sankhyā: series A, v. 15, pp. 331-342.

Mood, A. M., Graybill, F.A., and D. C. Boes (1974). Introduction to the Theory of Statistics. McGraw-Hill Book Company, New York.

Norden, R. H. (1972-73). A survey of maximum likelihood estimation. International Statistics Review, v. 40, pp. 329-354, v. 41, pp. 39-58.

Rao, C. R. (1962). Apparent anomalies and irregularities in maximum likelihood estimation. Sankhyā: The Indian Journal of Statistics: Series A, v. 24, pp. 73-101.

Rohatgi, V. K. (1976). An Introduction to Probability Theory and Mathematical Statistics. Wiley, New York.

Savage, L. J. (1956). On rereading R. A. Fisher. The Annals of Statistics, v. 4 no. 3, pp 441-500.