

# Chapter 19

## Other Inference Methods and Concepts

### 19.1 Fiducial Probability and Inference

R.A. Fisher introduced the concept of fiducial probability and used it to develop fiducial inference. Counterexamples though the years have led to its lack of use in statistics.

*Example.* Consider a deck of  $N$  cards numbered  $1, 2, \dots, N$ . One card is drawn at random, its number denoted by  $U$ . Then

$$\mathbb{P}(U = u) = \frac{1}{N} \quad u = 1, 2, \dots, N \quad (19.1)$$

Suppose now that we add an unknown number  $\theta$  to  $U$ . We are not told the observed value of  $U$ ,  $u_{obs}$ , or the value of  $\theta$  but we are told the observed value,  $t_{obs} = u_{obs} + \theta$ , of the total  $T = U + \theta$ . Note that we could see  $t_{obs}$  if and only if one of the following outcomes occurred:

$$(U = 1, \theta = t_{obs} - 1), (U = 2, \theta = t_{obs} - 2), \dots, (U = N, \theta = t_{obs} - N)$$

1. Given the value of  $t_{obs}$  there is a one-to-one correspondence between the values of  $U$  and  $\theta$ . If we knew  $\theta$  then we could determine the value of  $u_{obs}$ .
2. If we do not know the value of  $\theta$  then observing  $T = t_{obs}$  will tell us nothing about  $u_{obs}$ .
3. Thus the state of uncertainty regarding  $u_{obs}$  will be the same after the observation of  $t_{obs}$  as it was before.

Therefore we assume that (19.1) holds, and we can write

$$\mathbb{P}(\theta = t_{obs} - u) = \mathbb{P}(U = u) = \frac{1}{N} \quad u = 1, 2, \dots, N$$

which we call the **fiducial probability distribution** of  $\theta$ .

*Example.* Assume that  $X \stackrel{d}{\sim} N(\theta, 1)$  and define  $U = T - \theta$ . Then  $U \stackrel{d}{\sim} N(0, 1)$ . If we observe  $T = t_{obs}$  then  $t_{obs}$  arises from a pair of values ( $U = u, \theta = t_{obs} - u$ ) so that given  $t_{obs}$  there is a one-to-one correspondence between the possible values of  $U$  and  $\theta$ . Again, since  $\theta$  is unknown we will learn nothing about which value of  $U$  occurred. Thus we may assume that  $U \stackrel{d}{\sim} N(0, 1)$  even after  $T = t_{obs}$  is observed. Thus we can calculate (fiducial) probabilities about  $\theta$  by transforming them into probability statements about  $U$ , i.e.,

$$\theta \leq y \iff U \geq t_{obs} - y$$

so that

$$\mathbb{P}_{\mathcal{F}}(\theta \leq y) = \mathbb{P}(U \geq t_{obs} - y) = \Phi(y - t_{obs})$$

where  $\Phi(w) = \int_{-\infty}^w e^{-z^2/2}/\sqrt{2\pi} dz$ . Note that such probability statements are the same as if we treated  $\theta$  as a random variable with a normal distribution with mean  $t$  and variance 1. Thus the fiducial distribution of  $\theta$  is  $N(t_{obs}, 1)$ . The fiducial density is the derivative with respect to  $\theta$ , i.e.,

$$p_{\mathcal{F}}(\theta; t_{obs}) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\theta - t_{obs})^2}{2}\right\}$$

Kalbfleisch lists sufficient conditions for the fiducial argument to apply:

1. A single real-valued parameter.
2. A minimal sufficient statistic  $T$  exists for  $\theta$ .
3. There is a pivot  $U(T, \theta)$  such that
  - (i) For each  $\theta$ ,  $U(t, \theta)$  is a one-to-one function of  $t$
  - (ii) For each  $t$ ,  $U(t, \theta)$  is a one-to-one function of  $\theta$

These assumptions which are compounded when we move to more than one parameter mean that the scope of fiducial inference is severely limited. In fact it is largely ignored in most modern treatments of statistics.

Suppose that  $X_1, X_2, \dots, X_n$  are iid each  $N(\mu, \sigma^2)$  where  $\sigma^2$  is known. Then, whatever the value of  $\mu$ , we know that for every value of  $\alpha \in [0, 1]$  we can find an upper  $100(1 - \alpha)\%$  confidence interval for  $\mu$ , namely

$$\bar{X} + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$$

which we know has the property that

$$\mathbb{P} \left\{ \mu \leq \bar{X} + z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\} = 1 - \alpha$$

considered as a function of the random variable  $\bar{X}$ . Fisher noted that the left hand-side has all of the properties of a distribution function defined over the parameter space and thus proceeded to define the fiducial distribution of  $\mu$  for fixed  $\bar{X} = \bar{x}_{obs}$  and its derivative as the fiducial density of  $\mu$  given  $\bar{x}_{obs}$ . Thus we can speak of the fiducial probability,  $\mathbb{P}_{\mathcal{F}}(A)$  that  $\mu \in A$  calculated as

$$\mathbb{P}_{\mathcal{F}}(\mu \in A) = \int_A \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{\sqrt{n}(\mu - \bar{x}_{obs})^2}{2\sigma^2} \right\} d\mu$$

As another example let  $X_1, X_2, \dots, X_n$  be iid each uniform on  $(0, \theta)$ . Then  $Y = \max \{X_1, X_2, \dots, X_n\}$  is the minimal sufficient statistic which has distribution function

$$F_Y(y; \theta) = \mathbb{P}_{\theta}(Y \leq y) = [\mathbb{P}_{\theta}(X \leq y)]^n = \left[ \frac{y}{\theta} \right]^n$$

It follows that

$$\mathbb{P} \left( \frac{Y}{\theta} \leq y \right) = \mathbb{P}(Y \leq y\theta) = y^n$$

### 19.1.1 Good's Example

Suppose that  $X$  is a random variable with density function

$$f_X(x; \theta) = \frac{\theta^2(x+a)e^{-x\theta}}{a\theta+1} \quad \text{where } a > 0, \theta > 0, x \geq 0$$

which has distribution function

$$\begin{aligned} F(x; \theta) &= \int_0^x f(t; \theta) dt \\ &= \int_0^x \frac{\theta^2(t+a)e^{-t\theta}}{a\theta+1} dt \\ &= \frac{\theta^2}{a\theta+1} \int_0^x (t+a)e^{-t\theta} dt \end{aligned}$$

Now note that

$$\begin{aligned}
 \int_0^x (t+a)e^{-t\theta} dt &= \int_0^x te^{-t\theta} dt + a \int_0^x e^{-t\theta} dt \\
 &= -\frac{te^{-t\theta}}{\theta} \Big|_0^x + \frac{1}{\theta} \int_0^x e^{-t\theta} dt + a \int_0^x e^{-t\theta} dt \\
 &= -\frac{xe^{-x\theta}}{\theta} + \frac{1+a\theta}{\theta} \int_0^x e^{-t\theta} dt \\
 &= -\frac{xe^{-x\theta}}{\theta} + \frac{1+a\theta}{\theta^2} [1 - e^{-x\theta}]
 \end{aligned}$$

It follows that

$$F_X(x; \theta) = 1 - e^{-x\theta} \left[ 1 + \frac{x\theta}{a\theta + 1} \right]$$

The fiducial density of  $\theta$  is the derivative of  $F_X(x; \theta)$  with respect to  $\theta$  or

$$\mathcal{F}_x(\theta) = \frac{x\theta e^{-x\theta}}{(a\theta + 1)^2} [a + (a+x)(1+a\theta)]$$

Suppose we now observe, independently of  $X$ , another random variable  $Y$  with density

$$f_Y(y; \theta) = \frac{\theta^2(y+b)e^{-y\theta}}{b\theta + 1} \quad \text{where } b > 0, \theta > 0, x \geq 0 \text{ and } b \neq a$$

If we use the fiducial density based on  $X$  as a ‘‘prior’’ for  $\theta$  and combine it with the density for  $Y$  the resulting posterior would be

$$\mathcal{P}_{XY}(\theta; x, y) = \mathcal{F}_x(\theta) f_Y(y; \theta)$$

which is equal to

$$\frac{x\theta e^{-x\theta}}{(a\theta + 1)^2} [a + (a+x)(1+a\theta)] \frac{\theta^2(y+b)e^{-y\theta}}{b\theta + 1}$$

If fiducial probabilities behaved like true probabilities it should make no difference whether we observed  $X$  or  $Y$  first. If we used the fiducial density for  $\theta$  based on  $Y$  and combined it with density of  $X$  the resulting posterior would be

$$\mathcal{P}_{YX}(\theta; y, x) = \mathcal{F}_y(\theta) f_X(x; \theta)$$

which is equal to

$$\frac{y\theta e^{-y\theta}}{(b\theta + 1)^2} [b + (b + x)(1 + b\theta)] \frac{\theta^2(x + a)e^{-x\theta}}{a\theta + 1}$$

The laws of probability would say that these two expressions should be equal. They are clearly not. So fiducial probabilities are not compatible with Bayes theorem and hence are not warranted.

### 19.1.2 Edward's Example

Consider two hypotheses  $\theta = +1$  or  $\theta = -1$  and suppose that there are two possible outcomes of a random variable  $X$ ,  $X = +1$  or  $X = -1$ . The probability model for  $X$  is

$$\mathbb{P}_{\theta=+1}(X = x) = \begin{cases} p & x = +1 \\ q & x = -1 \end{cases} ; \quad \mathbb{P}_{\theta=-1}(X = x) = \begin{cases} q & x = +1 \\ p & x = -1 \end{cases}$$

Since when  $\theta = +1$ ,  $X = +1$  with probability  $p$ , and when  $\theta = -1$ ,  $X = -1$  with probability  $p$ , we have that

$$\mathbb{P}(X\theta = +1) \text{ with probability } p.$$

This statement is true in general and when  $X = +1$  is equivalent to the statement

$$\mathbb{P}(\theta = +1|X = +1) = p$$

which is the fiducial probability statement about  $\theta$  following from observing that  $X = +1$ .

Thus starting with no prior information and performing an experiment which is uninformative we arrive at a statement of probability for  $\theta$ .

Suppose now that  $p = q = \frac{1}{2}$ . Then

- (i) There is no information about  $\theta$  a priori
- (ii) The observation of  $X$  is totally uninformative about  $\theta$
- (iii) The likelihood ratio for comparing  $\theta = +1$  to  $\theta = -1$  is 1 indicating that, based on the observation, we have no preference for  $\theta = +1$  vs  $\theta = -1$

However we find, using the fiducial argument, that

$$\mathbb{P}_{\mathcal{F}}(\theta = +1) = \frac{1}{2} \text{ whatever the value of } X$$

Thus we have another example which casts doubt on the veracity of the fiducial argument.

## 19.2 Confidence Distributions

Recently there has been a great deal of research on confidence distributions which reduce, in many cases, to fiducial distributions, but which are solidly in the frequentist camp.

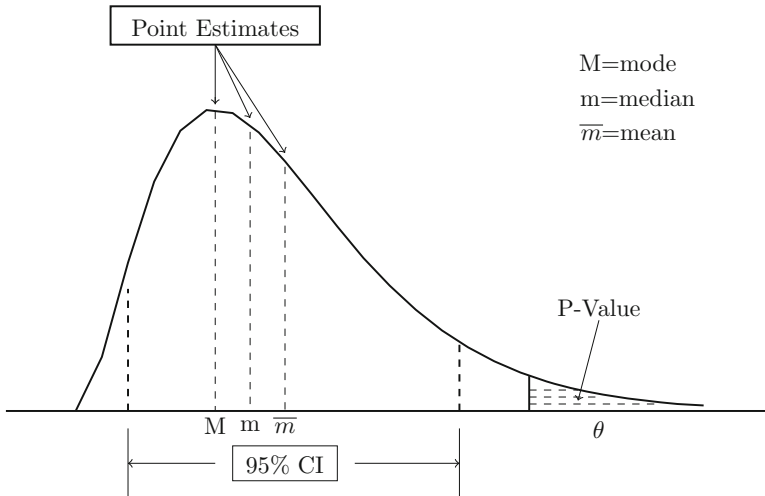
**Definition 19.2.1.** A function  $C_n(\theta; \mathbf{x}_n : \mathcal{X} \times \Theta \mapsto [0, 1]$  is called a **confidence distribution** if

- (i)  $C_n(\theta; \mathbf{x}_n)$  is a distribution function over  $\Theta$  for each fixed  $\mathbf{x}_n \in \mathcal{X}$
- (ii) At the true parameter point  $C_n(\theta_0, \mathbf{x}_n)$  as a function of  $\mathbf{x}_n \in \mathcal{X}$  has a uniform distribution over  $[0, 1]$

$C_n(\theta; \mathbf{x}_n : \mathcal{X} \times \Theta \mapsto [0, 1]$  is an asymptotic confidence distribution if (ii) is satisfied only asymptotically.

The paper by Xie and Singh [55] provides a useful review of the ideas. The following graph shows the unification of frequentist ideas using the confidence density:

Confidence Density Example



### 19.2.1 Bootstrap Connections

If  $\hat{\theta}$  is an estimator of  $\theta$  let the bootstrap estimator be  $\hat{\theta}^*$ . When the asymptotic distribution of  $\hat{\theta}$  is symmetric then the sampling distribution of  $\hat{\theta} - \theta$  is estimated by

the bootstrap distribution of  $\widehat{\theta} - \widehat{\theta}^*$ . In this case an asymptotic confidence distribution is given by

$$C_n(\theta) = 1 - \mathbb{P}(\widehat{\theta} - \widehat{\theta}^* \leq \widehat{\theta} - \theta | \mathbf{x}) = \mathbb{P}(\widehat{\theta}^* \leq \theta | \mathbf{x})$$

the bootstrap distribution of  $\widehat{\theta}$ .

### 19.2.2 Likelihood Connections

If we normalize a likelihood function  $\mathcal{L}(\theta; \mathbf{x})$  so that the area under the normalized likelihood function is 1, i.e., we form

$$\mathcal{L}^*(\theta; \mathbf{x}) = \frac{\mathcal{L}(\theta; \mathbf{x})}{\int_{\Theta} \mathcal{L}(\theta; \mathbf{x}) d\theta}$$

then, under certain conditions, we obtain an asymptotic normal confidence distribution.

Similarly, under the usual regularity conditions for maximum likelihood, we can use the normalized profile likelihood as an asymptotic confidence distribution for a parameter of interest.

It is also possible to construct approximate likelihoods using confidence distributions. Efron considered a confidence density  $c(\theta; \mathbf{x})$  for the parameter of interest. He then considered doubling the data set by introducing another data set, considered independent of the first but having exactly the same data  $\mathbf{x}$ . Then construct the confidence density  $c(\theta, \mathbf{x}, \mathbf{x})$  based on the doubled data set using the same confidence intervals to define the density. Then the implied likelihood function is

$$\mathcal{L}_{imp}(\theta; \mathbf{x}) = \frac{c(\theta; \mathbf{x}, \mathbf{x})}{c(\theta; \mathbf{x})}$$

### 19.2.3 Confidence Curves

Birnbaum introduced the idea of a confidence curve to summarize confidence intervals and levels of tests in one curve. In terms of confidence distributions the confidence curve is given by

$$CC(\theta) = 2 \min\{C_n(\theta; \mathbf{x}), 1 - C_n(\theta; \mathbf{x})\}$$

Thus a confidence distribution is simply a combinant such that for each  $\mathbf{x}_n \in \mathcal{X}$  it is a distribution function as  $\theta$  varies over  $\Theta$  and for fixed  $\theta = \theta_0$  it has a uniform distribution as  $\mathbf{x}_n$  varies over  $\mathcal{X}$ .

*Example.* Let

$$X_i \stackrel{d}{\sim} N(\mu, \sigma^2) \quad i = 1, 2, \dots, n$$

be independent with  $\sigma^2$  known. Then

$$C_n(\mu; \bar{x}_{obs}) = \Phi\left(\frac{\mu - \bar{x}_{obs}}{\sigma/\sqrt{n}}\right) \quad \text{where} \quad \Phi(y) = \int_{-\infty}^y \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

is a confidence distribution for  $\mu$ .

## 19.3 P-Values Again

Given the widespread importance of genomics P-values are now used more than ever. It is important to remember that P-values are observed values of a random variable and hence have intrinsic variability.

### 19.3.1 Sampling Distribution of P-Values

The P-value is defined, for a test statistic  $T$ , with distribution function  $F_{H_0}(t)$  assuming the null hypothesis is true as

$$\mathbb{P}_{H_0}(T \geq t_{obs}) \quad \text{where } t_{obs} \text{ is the observed value of } T$$

Note that the P-value is given by  $1 - F_{H_0}(t_{obs})$  and can be considered as an observed value of the random variable  $Y = 1 - F_{H_0}(T)$ . The distribution function of  $Y$  assuming the null hypothesis is true is

$$\begin{aligned} F_Y(y) &= \mathbb{P}_{H_0}(Y \leq y) \\ &= \mathbb{P}_{H_0}\{1 - F_{H_0}(T) \leq y\} \\ &= \mathbb{P}_{H_0}\{F_{H_0}(T) \geq 1 - y\} \\ &= 1 - \mathbb{P}_{H_0}\{F_{H_0}(T) \leq 1 - y\} \\ &= 1 - \mathbb{P}_{H_0}\{T \leq F_{H_0}^{-1}(1 - y)\} \\ &= 1 - F_{H_0}\{F_{H_0}^{-1}(1 - y)\} \\ &= 1 - (1 - y) \\ &= y \end{aligned}$$



for  $0 < y < 1$ . That is, the P-value under the null hypothesis has a uniform distribution. This fact has been known for decades, but P-values are not reported with a standard error as other statistics are.

If the alternative hypothesis is true assume that  $T$  has a distribution function  $G_H(t)$ . Then

$$\begin{aligned} F_Y(y) &= \mathbb{P}_H(Y \leq y) \\ &= \mathbb{P}_H \{1 - F_{H_0}(T) \leq y\} \\ &= \mathbb{P}_H \{F_{H_0}(T) \geq 1 - y\} \\ &= 1 - \mathbb{P}_H \{F_{H_0}(T) \leq 1 - y\} \\ &= 1 - \mathbb{P}_H \{T \leq F_{H_0}^{-1}(1 - y)\} \\ &= 1 - G_H \{F_{H_0}^{-1}(1 - y)\} \end{aligned}$$

Under suitable regularity conditions the density of  $Y$  under the alternative hypothesis is

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} \\ &= \frac{d[1 - G_H \{F_{H_0}^{-1}(1 - y)\}]}{dy} \\ &= -g_H \{F_{H_0}^{-1}(1 - y)\} \frac{1}{-f_{H_0} \{F_{H_0}^{-1}(1 - y)\}} \\ &= \frac{g_H \{F_{H_0}^{-1}(1 - y)\}}{f_{H_0} \{F_{H_0}^{-1}(1 - y)\}} \end{aligned}$$

Note that for the observed value of  $T$ ,  $t_{obs}$ , we have  $y = 1 - F_{H_0}(t_{obs})$  and hence the density evaluated at the observed value of  $T$ ,  $t_{obs}$ , is given by

$$\frac{g_H \{F_{H_0}^{-1}(1 - y)\}}{f_{H_0} \{F_{H_0}^{-1}(1 - y)\}} = \frac{g_H \{F_{H_0}^{-1}[F_{H_0}(t_{obs})]\}}{f_{H_0} \{F_{H_0}^{-1}[F_{H_0}(t_{obs})]\}} = \frac{g_H(t_{obs})}{f_{H_0}(t_{obs})}$$

the likelihood ratio!

P-values have been under heavy fire in the last few years for overstating the importance of effects observed in clinical and other investigations [48]. The results above due to Donahue [12] and others have been used by Boos and Stefanski [5] to explain why P-values overstate the conclusions of studies. The results of Goodman [21] are also relevant. The bottom line appears to be that use of P-values is not the way to present the evidence from studies that rely on statistical analysis to report their conclusions.

## 19.4 Severe Testing

Severe testing is a concept claimed to be useful for post-data inference. The ideas are best illustrated through an example and I follow the example in Mayo and Spannos [33]. Suppose that  $X_1, X_2, \dots, X_n$  are iid each  $N(\mu, \sigma^2)$  where  $\sigma = 2$ , that  $n = 100$ , and we choose  $\alpha = 0.025$  for a one-sided test of

$$H_0 : \mu \leq 12 \text{ vs } \mu > 12$$

Since, under  $H_0$ ,  $\bar{X}_n \stackrel{d}{\sim} N(12, 2/10)$ , we reject using the Neyman-Pearson theory if

$$d(\mathbf{x}_{obs}) = \frac{\sqrt{n}(\bar{x}_{obs} - 12)}{2} \geq 1.96 \iff \bar{x}_{obs} \geq 12.4$$

Suppose now that we observe  $\bar{x}_{obs} = 11.8$ . Then we would not reject  $H_0 : \mu = 12$ . If a value of  $\mu$  equal to 12.2 was deemed to be of scientific or substantive importance we can ask the question do we have evidence that  $\mu < 12.2$ ? Mayo suggests calculating the severity with which  $\mu < 12.2$  passes the test. The severity with which  $\mu = 12.2$  passes the test in cases where  $H_0$  is accepted is defined in this situation as

$$\begin{aligned} \mathbb{P}_\mu(\bar{X} > \bar{x}_{obs}) &= \mathbb{P}_{\mu=12.2} \{ \bar{X} > 11.8 \} \\ &= \mathbb{P}_{\mu=12.2} \left\{ \frac{\sqrt{100}(\bar{X} - 12.2)}{2} > \frac{\sqrt{100}(11.8 - 12.2)}{2} \right\} \\ &= \mathbb{P} \{ Z > -2 \} \\ &= 0.977 \end{aligned}$$

Note that the power of the test at 12.2 is

$$\begin{aligned} \mathbb{P}_{\mu=12.2} \{ \bar{X} > 12.4 \} &= \mathbb{P}_{\mu=12.2} \left\{ \frac{\sqrt{100}(\bar{X} - 12.2)}{2} > \frac{\sqrt{100}(12.4 - 12.2)}{2} \right\} \\ &= \mathbb{P} \{ Z > 1 \} \\ &= 0.159 \end{aligned}$$

Mayo and Spannos define the **attained power** in this situation as

$$\mathbb{P}_\mu \{ \bar{X} > \bar{x}_{obs} \}$$

so that the severity with which  $\mu$  passes the test is simply the attained power at  $\mu$  when the observed outcome leads to acceptance of the null hypothesis.

Suppose now that the null hypothesis is rejected. In the example suppose that  $\bar{x}_{obs} = 12.6$ . The null hypothesis that  $\mu = 12$  is rejected. Again, assuming that  $\mu = 12.2$  is of scientific or substantive interest, do we have evidence of a value of  $\mu$  of scientific interest? Mayo and Spannos suggest calculating the severity of  $\mu > 12.2$  defined by

$$\begin{aligned} \mathbb{P}_{\mu} \{ \bar{X} \leq \bar{x}_{obs} \} &= \mathbb{P}_{\mu=12.2} \{ \bar{X} \leq \bar{x}_{obs} \} \\ &= \mathbb{P}_{\mu=12.2} \{ \bar{X} \leq 12.6 \} \\ &= \mathbb{P}_{\mu=12.2} \left\{ \frac{\sqrt{100}(\bar{X} - 12.2)}{2} \leq \frac{\sqrt{100}(12.6 - 12.2)}{2} \right\} \\ &= \mathbb{P} \{ Z \leq 2 \} \\ &= 0.977 \end{aligned}$$

Note that in the case of a hypothesis which is rejected the severity is simply 1 minus the attained power.

Severe testing does a nice job of clarifying the issues which occur when a hypothesis is accepted (not rejected) by finding those values of the parameter (here  $\mu$ ) which are plausible (have high severity) given acceptance. Similarly severe testing addresses the issue of a hypothesis which is rejected by finding those values of the parameter  $\mu$  which are plausible (have high severity) given rejection. Note that severity is a function of the test,  $T$ , the hypothesis,  $H$ , and the observed data,  $\mathbf{x}_{obs}$ . Thus it is inherently a random variable and the standard results on  $p$ -values and their distributions apply to severity as well. Also note that conventions need to be established for when severity is judged to be high.

Finally note that most of the existing examples implicitly seem to require a monotone likelihood ratio so that members of the exponential family are included, but whether or not other distributions are covered under the existing theory is unknown.

## 19.5 Cornfield on Testing and Confidence Intervals

The following quotes by Jerry Cornfield (1966) indicate that the problems with frequentist statistics have been known for a long time.

Cornfield defines the  $\alpha$ -postulate as “All hypotheses rejected the same critical level have equal amounts of evidence against them.”

The following example will be recognized by statisticians with consulting experience as a simplified version of a very common situation. An experimenter, having made  $n$  observations in the expectation that they would permit the rejection of a particular hypothesis, at some predetermined significance level, say 0.05, finds that he has not quite attained this critical level. He still believes that the

hypothesis is false and asks how many more observations would be required to have reasonable certainty of rejecting the hypothesis if the means observed after  $n$  observations are taken as the true values. He also makes it clear that had the original  $n$  observations permitted rejection he would simply have published his findings. Under these circumstances it is evident that there is no amount of additional observations, no matter how large, which would permit rejection at the 0.05 level. If the hypothesis being tested is true, there is a 0.05 chance of it having been rejected after the first round of observations. To this chance must be added the probability of rejecting after the second round, given failure to reject after the first, and this increases the total chance of erroneous rejection to above 0.05. In fact as the total number of observations in the second round is indefinitely increased the significance approaches 0.0975 ( $=0.05 + 0.95 \times 0.05$ ) if the 0.05 criteria is retained. Thus no amount of additional evidence can be collected which would provide evidence against the hypothesis equivalent to rejection at the  $P = 0.05$  level and adherents of the  $\alpha$ -postulate would presumably advise him to turn his attention to other scientific fields. The reasonableness of this advice is perhaps questionable (as is the possibility that it would be accepted). In any event it does not seem possible to argue seriously in the face of this example that all hypotheses rejected at the .05 level have equal amounts of evidence against them.

The confidence set yielded by a given body of data is the set of all hypotheses not rejected by the data, so that the relation between hypothesis test and confidence limits is close. In fact the confidence limit equivalent of the  $\alpha$ -postulate is "All statements made with same the confidence coefficient have equal amounts of evidence in their favor." That this may be seen no more reasonable the  $\alpha$ -postulate is suggested by the very common of inference about the ratio of two normal means. The most selective unbiased confidence set for the unknown ratio has the following curious characteristic: for every sample point there exists an  $\alpha > 0$  such that all confidence limits with coefficients  $\geq 1 - \alpha$  are plus to minus infinity. But to assert that the unknown ratio lies between plus and minus infinity with confidence coefficient of only  $1 - \alpha$  is surely being overcautious. Even worse, the postulate asserts that there is less evidence for such an infinite interval than there is for a finite interval about a normal mean, but made with coefficient  $1 - \alpha'$  where  $\alpha' < \alpha$ . The  $\alpha$ -postulate cannot therefore be considered anymore reasonable than it is for hypothesis testing.

It has been proposed by proponents of confidence limits that this clearly undesirable characteristic of the limits on a ratio be avoided by redefining the sample space so as to exclude all sample points that lead to infinite limits for given  $\alpha$ . This is equivalent to saying that if the application of a principle to given evidence leads to an absurdity then the evidence must be discarded. It is reminiscent of the heavy smoker, who, worried about the literature relating smoking to lung cancer, decided to give up reading.