行政院國家科學委員會專題研究計畫　成果報告

Post data 推論在一些統計問題上的應用(III)

計畫類別：個別型計畫
計畫編號：NSC94-2118-M-029-006-
執行期間：94 年 08 月 01 日至 95 年 07 月 31 日
執行單位：東海大學統計學系

計畫主持人：王堯弘

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 95 年 6 月 27 日

UNDERSTANDING THE FIDUCIAL DISTRIBUTION

Y.H. Wang and Hsiao-fen Chang

Department of Statistics, Tunghai University

Department of Accounting, Aletheia University

June 26, 2006

Abstract

The purpose of this paper is to give a systematic explanation of the fiducial distribution. Unlike many of the writings on this topic, it is written with the conviction that the controversial fiducial argument does make sense. We review some literatures on the fiducial argument, give a general definition of the fiducial distribution and propose a method of constructing the fiducial distribution for continuous parameter space. As demonstrations, we present some applications of the fiducial distribution in solving some problems in current statistical inference.

1. INTRODUCTION

Let us start with a simple example of "post-data probability". If X is a continuous random variable with an unknown median θ, then

$$\Pr(X < \theta) = \Pr(X > \theta) = 1/2. \tag{1}$$

The above statement says that it is equally likely that the value of the observable random variable X, when realized, is on the left or on the right of the unobservable parameter θ. Now suppose that $X = 3.45$ is observed, can we still claim that

$$\Pr(3.45 < \theta \mid X = 3.45) = \Pr(3.45 > \theta \mid X = 3.45) = 1/2? \tag{2}$$

Most people would find the statement (2) odds, because unlike (1) which can be mechanically calculated using the distribution of X, (2) can not. We need a little bit of patience here. Doesn't (2) simply say that if X is observed to be 3.45, then with even chance the unobservable parameter θ is either on the left or on the right of 3.45? Which is just a slight rewording of the statement (1). The difficulty here is that in (2) the statements "$3.45 < \theta$" and "$3/45 > \theta$" involve no random elements, and therefore their "probabilities" can not be calculated in the conventional manner. We shall show later in this paper that there do have a way to calculate the probabilities in (2) using a "posterior" distribution of θ conditional on $X = 3.45$. This posterior distribution is the fiducial distribution which we shall explain in detail in this paper.

1.1. Fisher's "Inverse Probability"

William Sealy Gosset changed the course of statistics when he published the paper Student (1908) in which he derived the t-distribution. That paper offered for the first time in statistics history a small-sample exact inference on a parameter of a population. What he was possibly not aware of was that his paper became the cause of a great mystery in statistics, unrelated to the t-distribution, concerning construction of the posterior distributions of unknown parameters in the absence of prior distributions.

A short paragraph in that paper reads:

> "If two observations have been made and we have no other information, it is an even chance that the mean of the population will lie between them."

In modern notation and a slightly more general fashion, let X_1 and X_2 be a random sample of size 2 from a continuous population with median θ and $Y_1 = \min\{X_1, X_2\}$ and $Y_2 = \max\{X_1, X_2\}$ be the corresponding order statistics. Denote y_1 and y_2 to be realizations of Y_1 and Y_2. Then what Gosset meant is that

$$\Pr(y_1 \leq \theta \leq y_2) = 1/2. \tag{3}$$

Since θ is the median of a continuous distribution, we have

$$\Pr(Y_1 \leq \theta \leq Y_2) = 1/2. \tag{4}$$

The probability (4) holds for all continuous distributions with the median θ. Since Gosset was specifically dealing with the normal (θ, σ^2) population, θ is the mean as well as the median, (4) certainly holds for all variance $\sigma^2 > 0$. As in the previous example, the problem is the meaning of the word "probability" in (3). If θ is a random variable with a prior distribution of its own, then it can be interpreted as a posterior probability conditional on $Y_1 = y_1$ and $Y_2 = y_2$. But in the context of the paper, θ is clearly a constant. The usual interpretation of "probability" that $\{y_1 \leq \theta \leq y_2\}$ is fishy, if not impossible.

But Fisher obviously saw that there could be a plausible interpretation of the "probability" in (3) even in the absence of a prior distribution of θ. He published his idea in Fisher (1930) in which he introduced the term "fiducial " into statistics and presented an argument, later known as "fiducial argument", to construct a "posterior distribution" of an unknown parameter θ sans prior distributions. The title of his paper is "Inverse Probability". The term "inverse probability" in modern terminology is a synonym of "post-data or posterior probability". The terms such as fiducial inference, fiducial probability, fiducial interval, etc. are derivatives of the fiducial argument. Even the widely accepted frequentist (pre-data) confidence interval, generally attributed to Jerzy Neyman, was initially intended as a clarification and development of fiducial argument. In Fisher (1939 and 1945) he revealed that Gosset's short paragraph was the origin of his fiducial argument. (See also Edwards (1983).)

While the pre-data probability (4) is true, unfortunately, in Section 3, we shall show that the post-data probability (3) is false. In other words, Fisher started his fiducial argument with the misinterpretation of the post-data probability (3) .

The fiducial argument has since become a very controversial issue in statistics. That " ... the fiducial inference as put forward by R.A. Fisher is not so much a theory as a collection of examples." (Buehler (1983).) "Fiducial inference stands as R.A. Fisher's one great failure." (Zabell (1992).) It is "no more than a misconception born out of an early mistake." (Neyman (1961).) "Among R.A. Fisher's many important contributions to statistics the fiducial argument has had a very limited success and is now essentially dead." (Pedersen (1978).) "It seems never to occur to Fisher that he could be wrong. ... the possibility that the two ideas that ... he valued above all else, fiducial inference and the fundamental theorem of natural selection, were both wrong simply did not enter his mind." (Kempthorne (1983).) We can go on and on to quote many more adverse comments and criticisms on the fiducial argument, but it does not serve any useful purpose.

On the other hand, not everybody is negative about the fiducial argument. There are many serious attempts to make something out of it and especially many keen interest on the philosophical aspect of it, notably Fraser (1961 and 1968), Jeffreys (1932 and 1961), Barnard (1963), Kyburg (1963 and 1974), Dempster (1963, 1964, 1966 and 1968), Hacking (1965), Edwards (1976 and 1983), Pedersen (1978) , Seidenfeld (1979), Buelher (1983), and most recently Wang (2000).

We are deeply convinced that the fiducial argument does make sense and is a worthwhile topic for further study. It took human beings many hundred years to come up with a universally accepted definition of "probability" in 1933 by Kolmogorov. So, instead of trying to retrospectively read and reread Fisher's mind and to critically expose inconsistency and incoherence in his argument, why don't we move ahead to see if we can make it works! We don't have to be confined to his original idea and thinking. We can modify, revise or even expand it. The other two types of statistical inference, Bayesian and frequentist, are both not impeccable. The Bayesian inference is perpetually plagued by the problem of choosing the proper prior distributions of the unknown parameters. We doubt that one day the Bayesians would be ever able to come up with a universally acceptable method of choosing the right priors for the unknown parameters. And the frequentists would always have, among others, the problems of post-data interpretation of pre-data inference and of evaluating coverage probability of a random confidence set of a parameter in the absence of pivotal quantity involving the parameter of interest. (See the next section.) The fiducial inference offers a nice alternative to the other two types of inference and can fill in the gaps in them. (See Wang (2000) for some illustrations.) "I don't understand yet what fiducial probability does. We shall have to live with it a long time before we know what it's doing for us. But it should not be ignored just because we don't yet have a clear interpretation." Fisher confessed to L. J. Savage later in his life. (Savage (1964).) (For definition of pivotal quantity, see Definition 16.)

2. SOME PROBLEMS WITH CURRENT STATISTICAL INFERENCE

Let us examine some problems with the current pre-data inference. We shall show in Section 4 that fiducial argument offers an alternative or even the correct solution to each of the problems.

2.1. Post-data Evaluation of Pre-data Probability

Besides the two post-data probabilities (2) and (3) , let us consider another example along the line. A statement on page 225 in Lehmann (1986) reads: "Suppose for example that X is distributed $N(\theta, 1)$, and consider the confidence interval

$$X - 1.96 < \theta < X + 1.96$$

corresponding to confidence coefficient $\gamma = 0.95$. Then the random interval $(X-1.96, X+1.96)$ will contain θ with probability 0.95. Suppose now that X is observed to be 2.14. At this point, the earlier statement reduces to the inequality $0.18 < \theta < 4.10$, which no longer involves any random element. Since the only unknown quantity is θ, it is tempting (but not justified) to say that θ lies between 0.18 and 4.10 with probability 0.95." "To attach a meaningful probability to the event $0.18 < \theta < 4.10$ *requires* (the italics is ours) that θ be random." He then proceeded to explain what the Bayesian inference does.

This is a problem of the post-data evaluation of a pre-data probability. What is the "probability" that $0.18 < \theta < 4.10$ conditional on $X = 2.14$, given that X is normally distributed? In Lehmann's opinion, this question is meaningless and can not be answered, unless θ is a random variable. In other words, the Bayesian approach is the only proper way out. We shall demonstrate that it is perfectly all right and justified to say that "θ lies between 0.18 and 4.10 with probability 0.95", *without requiring that θ be random.* We shall use the fiducial distribution to be constructed later to compute the probability

$$\Pr(0.18 < \theta < 4.10 \mid X = 2.14) \tag{5}$$

and show that it *is* equal to 0.95. In other words, the fiducial distribution can come to the rescue in this situation.

2.2. Frequentist Confidence Intervals

From frequentist point of view, a confidence set $C(X_1, ..., X_n)$ for a parameter θ is a random set. This random set is usually obtained by inverting the acceptance region of testing a point-null hypothesis $H_0 : \theta = \theta_0$ against the two-sided alternative $H_1 : \theta \neq \theta_0$ or a one-sided null hypothesis $H_0 : \theta \leq \theta_0$ against the one-sided alternative $H_1 : \theta > \theta_0$. The confidence set obtained by the test-inverting method have many problems: One is that it may not be an interval, even in a situation where an interval is most appealing. Another is that it may not contain the maximum likelihood estimator $\widehat{\theta}$ of θ. The other is that if a pivotal quantity involving the parameter θ does not exist, then the evaluation of the coverage probability $\Pr(\theta \in C(X_1, ..., X_n))$ poses a big problem. All these problems are no problem from the fiducialist point of view. (See Wang (2000) for some details.)

Let us elaborate a little bit more on the last point - the evaluation of the coverage probability of a random confidence set in the absence of a pivotal quantity. One of the most well-known approximate confidence interval, the Wald interval, for the binomial success probability θ is

$$I_n(\mathbf{X}) = \left(\widehat{\theta} - z_{\alpha/2} \sqrt{\frac{\widehat{\theta}\left(1-\widehat{\theta}\right)}{n}}, \widehat{\theta} + z_{\alpha/2} \sqrt{\frac{\widehat{\theta}\left(1-\widehat{\theta}\right)}{n}} \right). \tag{6}$$

where $\widehat{\theta} = \overline{X}$, is the sample mean of a Bernoulli (θ) random sample $\mathbf{X} = (X_1, ..., X_n)$. We know that $\lim_{n\to\infty} \Pr(\theta \in I_n(\mathbf{X})) = 1 - \alpha$, but in practice n is never infinity. For finite n, how to evaluate the coverage probability (6) has been a subject of many recent studies. (See Brown et al (2001 and 2002), Wang (2000) and the references cited therein.) The difficulty is that in the binomial distribution there is no pivotal quantity involving θ and the coverage probability $\Pr(\theta \in I_n(\mathbf{X}))$ is a function of θ fluctuating up and down between 0 and 1, even for large values of n, for all confidence coefficients $1-\alpha$. Brown et al (2001) had many detailed graphical displays of the behaviors of the coverage probabilities $\Pr(\theta \in I_n(\mathbf{X}))$ for different values of θ, n and α, and concluded that "the chaotic coverage properties of the Wald interval are far more persistent than is appreciated, ..., common textbook prescriptions regarding its safety are misleading and defective in several respects, and cannot be trusted". We shall see later that the fiducial inference offers an alternative solution to this problem and the fiducial distribution can be used to construct confidence intervals for θ and to evaluate the coverage probabilities of any existing intervals obtained by any means, with or without existence of pivotal quantities.

2.3. Inconherence of the P-value

A *p-value* is a post-data probability, conditional on the observed value of a test statistic, when testing a null hypothesis against an alternative. In applications, it is a simple device to determine if a null hypothesis H_0 is to be rejected. Despite its widespread use in applications, it is known that the current *p-value* theory is incoherent. (See Sackrowitz and Samuel-Caln (1999) and Schervish (1996).) Let us give a simple illustration here.

Consider the normal $(\theta, 1)$ population with the sample mean $\overline{X}$ as the test statistic to test a one-sided null hypothesis $H_0 : \theta \leq 0$ against the alternative $H_1 : \theta > 0$, the *p-value* π is generally defined as

$$\pi_1(\overline{x}) = \Pr\left(\overline{X} \geq \overline{x}\right) = 1 - \Phi(\overline{x}). \tag{7}$$

For testing a point null hypothesis $H_0 : \theta = 0$ against the two-sided alternative $H_1 : \theta \neq 0$, it is defined as

$$\pi_0(\overline{x}) = 2\Pr\left(\overline{X} \geq \overline{x}\right) = 2\left(1 - \Phi(\overline{x})\right), \tag{8}$$

where in (7) and (8) $\overline{x}$ is the observed value of $\overline{X}$ and Φ is the standard normal c.d.f. Since the *p-value* π is continuous in $\overline{x}$, for all $0 < \alpha < 1/2$, we can find

$\overline{x}_0$ such that $\Pr\left(\overline{X} \geq \overline{x}_0\right) < \alpha$ and $2\Pr\left(\overline{X} \geq \overline{x}_0\right) > \alpha$. For these values of $\overline{x}_0$, we would reject $H_0 : \theta \leq 0$ and accept $H_0 : \theta = 0$ simultaneously at level α. But $\{\theta : \theta = 0\} \subseteq \{\theta : \theta \leq 0\}$, the incoherence of the *p-value* is evident.

What is wrong? Because *p-value*, being a post-data probability conditional on the observed value of a test statistic, clearly belongs to the domain of the post-data inference, but has been wrongly treated in the literature as a pre-data inference devise. (Evidently, *p-value* theory comes from the concepts of "the power function" of hypotheses testing.) We shall explain how to rectify it in Section 4, after we introduce the fiducial distribution in the next section.

Remark I

Out of many problems in current statistical theory, we have just chosen three simple ones for later demonstration of the applications of the fiducial distribution to be constructed in the next section. There are many more complicated ones, which we do not plan to get into them at this point, for instance, the Neyman-Pearson Lemma. This Lemma provides a method of deriving a most powerful (MP) critical region for testing a simple null hypothesis $H_0 : \theta = \theta_0$ against a simple alternative $H_1 : \theta = \theta_1 \neq \theta_0$. For simplicity, suppose the population is normal $(\theta, 1)$ and a single observation X is used to test $\theta_0 = 0$ against a fixed point $\theta_1 > 0$, this Lemma leads to the MP level-.05 critical region $C = \{x : x \geq 1.645\}$, for all $\theta_1 > 0$. If $X = 1.5$ is observed, then $\theta_0 = 0$ is accepted, disregarding the value of θ_1. This decision is evidently debatable, because if $\theta_1 = 2$ then $X = 1.5$ is much more likely coming from the population with $\theta = 2$ rather than from that with $\theta = 0$. Similarly, if $X = 1.8$ is observed, then $\theta = 0$ is rejected. But if $\theta_1 = 4$, with $X = 1.8$ the choice between $\theta = 0$ and $\theta = 4$ should not be 4, but 0. (See Wang (2001) for more details.) There are many more flaws attributable to the logic of the Neyman-Pearson Lemma. We believe that one recourse to resolve this dilemma is the fiducial inference.

3. THE FIDUCIAL DISTRIBUTION

3.1 Scope of the Problem

We shall confine ourselves to one dimensional sample space Ω and one dimensional parameter space Θ. Also we shall restrict ourselves to the continuous parameter space only and leave the discrete case in a future report. We assume the continuous parameter space is an interval $\Theta = [\theta_1, \theta_2]$, with one or both limits may be open and/or infinite. For example, for the correlation coefficient of a bivariate distribution $\theta = \rho$, $\Theta = [-1, 1]$, for the scale parameter variance $\theta = \sigma^2$, $\Theta = (0, \infty)$, for the location parameter $\theta = \mu$, $\Theta = (-\infty, \infty)$, for the Bernoulli success probability $\theta = p$, $\Theta = [0, 1]$, for the parameter θ of the uniform $(0, \theta)$ distribution with $0 < \theta < \infty$, $\Theta = (0, \infty)$.

Denote $\mathbf{X} = (X_1, ..., X_n)$ to be a random sample from a certain population and $Y = t(\mathbf{X})$ a sufficient statistic for θ with cumulative distribution function $F(y \mid \theta) = \Pr(Y \leq y \mid \theta)$, where θ is the parameter of concern. For simplicity, we shall assume that Y has no nuisance parameter. We do not require that F be continuous in y. In other words, we shall consider both the continuous and the discrete sample spaces.

The distribution function $F(y \mid \theta)$ plays a central role in fiducial argument. The whole theory is based on the interdependence relation between y and θ through the function F. An ideal distribution F should possess two conditions. One is that it is a monotone function in θ and the other is that it has range in $(0, 1)$ inclusively, for all $Y = y$ fixed. (See Savage (1976).) In Fisher (1930), he considered F a decreasing function in θ and tacitly assumed it has range in whole $(0, 1)$, for all $Y = y$ fixed. Many earlier writers tended to go around with Fisher. Stone (1983) believes that requiring F have range in whole $(0, 1)$ is a "nontrivial condition". (We found that a few distributions fail the second condition.) If these two conditions are met, then for fixed $Y = y$ and p with $0 < p < 1$, the equality $F(y \mid \theta) = p$ leads to a *unique* solution for $\theta = \theta(p, y)$. Therefore, if $Y = y$ is fixed, then the relation between θ and p is one-to-one. (This is a key point in fiducial argument emphasized from the beginning in Fisher (1930).) We shall come back to investigate these two conditions in more details later in this section.

3.1.1 Definitions and Restrictions.—

Let us give a definition of fiducial distribution as follows:

Definition 1 *The fiducial distribution of a parameter θ is a posterior distribution of the parameter conditional on an observed value of the statistic $Y = y$ in the absence of a prior distribution of θ.*

There are several requirements stressed by Fisher and many other writers on fiducial argument.

One essential requirement is "the complete absence of information *a priori* about the parameter in question." (Edwards (1983).) It was Fisher's means "to arrive at the equivalent of posterior distributions in a Bayesian argument without the introduction of prior distributions." (Savage (1976).) This requirement was not clearly stated and emphasized by Fisher in his early writings. It was "stipulated for the first time in Fisher (1956)." (Pedersen (1978).) We shall adhere to this requirement.

Another "principal condition which Fisher laid down was that a proper fiducial distribution must depend on the data only through sufficient statistics." (Dempster (1964).) This condition is certainly theoretically appropriate to have "because otherwise there would be a multitude of fiducial distributions (for each parameter) each corresponding to some other statistic which used less than all the information". (Edwards (1976).) Fisher originally specified that Y "is the estimate found by the method of maximum likelihood". (Fisher (1930).) But in many of his later writings, he consistently used the term "sufficient statistic". What happens when no sufficient statistic exists? Fisher's own opinion was that "the whole of information (supplied by the data) could be utilized in the form of *ancillary* information". (See Fisher's contribution to the discussion in Neyman (1934).) In other words, when there is no sufficient statistic, a recourse is the conditional inference. When making inferences on θ, it certainly makes sense to require Y be sufficient statistic. Therefore we shall restrict Y to be a sufficient statistic for θ. (A statistic $Y = t(\mathbf{X})$ is sufficient for a parameter θ if the joint distribution of $\mathbf{X} = (X_1, ..., X_n)$ conditional on $Y = y = t(\mathbf{x})$ is independent of the parameter θ.)

In his original paper Fisher had required that both the sample space and the parameter space be continuous, and that $F(y \mid \theta)$ be differentiable with respect to both y and θ. And many earlier writers tended to go along with him. (See Neyman (1934), Lindley (1958) and Edwards (1983).) We do not find this requirement necessary. In fact, in Wang (2000), the binomial distribution $F(y \mid \theta) = \sum_{i \leq y} \binom{n}{i} \theta^i (1-\theta)^{n-1}$, which is continuous in θ, but not in y, was used to define a posterior distribution of θ conditional on y and it worked out just fine. In this paper, we shall stick to the continuous parameter space. The details for the discrete parameter space shall be separately reported later.

The last requirement is the existence of a pivotal quantity. (Fisher (1941), see also Edwards (1983).) There are many distributions in which pivotal quantities do not exist, for example all the distributions with continuous sample space and discrete parameter space. We have found that it is not necessary to stick to this requirement.

By the dimension of a sample space, we mean the range of Y, not that of the population. For example, in a bivariate population, the sample correlation coefficient

$Y = R$ defined by

$$Y = R = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}, \tag{9}$$

from a bivariate random sample $\{(X_1, Y_1), ..., (X_n, Y_n)\}$ has sample space $\Omega = [-1, 1]$. For two independent random samples $X_1, ..., X_n$ and $Y_1, ..., Y_m$, the statistic

$$Y = \frac{S_x^2}{S_y^2} \tag{10}$$

has sample space $\Omega = (0, \infty)$. (S_x^2 and S_y^2 are the sample variances of $X's$ and $Y's$, respectively.)

Similarly, by the dimension of a parameter space, it refers to the dimension of the parameter of concern, not that of the parameters in the population. For example, in a bivariate distribution, the correlation coefficient

$$\theta = \rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{11}$$

has parameter space $\Theta = [-1, 1]$. For two independent populations, the ratio of two variances

$$\theta = \frac{\sigma_x^2}{\sigma_y^2} \tag{12}$$

has parameter space $\Theta = (0, \infty)$.

3.2 Constructing a Fiducial Distribution

How can we invent a posterior distribution for the parameter θ conditional on $Y = y$ without a prior distribution of θ? An example immediately pops up in mind is the normal (θ, σ^2) distribution. We take $Y = t(\mathbf{X}) = \overline{X}$, the sample mean. Since $(y - \theta)^2$ is invariant when y and θ are interchanged. Then the function g defines by

$$g(\theta \mid y) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-n(\theta - y)^2/2\sigma^2}, \text{ for } \theta \in R, \ y \text{ fixed}, \tag{13}$$

is an instant probability density function of θ for all given $Y = y$, which is the density of the normal $(y, \sigma^2/n)$ distribution.

Another example is the exponential distribution $f(y \mid \theta) = \frac{1}{\sigma} e^{-(y-\theta)/\sigma}$, for $y > \theta > -\infty$, $\sigma > 0$ is fixed. Switching the roles of y and θ, we get

$$g(\theta \mid y) = \frac{1}{\sigma} e^{(\theta - y)/\sigma}, \text{ for } -\infty < \theta < y, \ y \text{ fixed.}$$

Again g is a perfect density of θ, given $Y = y$.

There are many more examples which work nicely and the posterior density g can be obtained by manipulating the roles of θ and y in the density f. But in general, this approach does not work. For example, the binomial mass function $f(y \mid \theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$, the Poisson mass function $f(y \mid \theta) = e^{\theta} \theta^y / y!$ and the exponential distribution with density $f(y \mid \theta) = \frac{1}{\theta} e^{-y/\theta}$, switching the roles of θ and y in f does not produces posterior densities.

The next definition defines a relation between the parameter θ and the statistic Y which is useful in constructing fiducial distributions.

Definition 2 *A family of distributions* $\{F(y \mid \theta); \theta \in \Theta\}$ *is stochastically increasing if* $\theta_1 > \theta_2$ *implies* $F(y \mid \theta_1) \leq F(y \mid \theta_2)$ *for all* y *and* $F(y \mid \theta_1) < F(y \mid \theta_2)$ *for some* y. *Similarly, a family of distributions* $\{F(y \mid \theta); \theta \in \Theta\}$ *is stochastically decreasing if* $\theta_1 > \theta_2$ *implies* $F(y \mid \theta_2) \leq F(y \mid \theta_1)$ *for all* y *and* $F(y \mid \theta_2) < F(y \mid \theta_1)$ *for some* y.

From the parameter space point of view, a family of distributions $\{F(y \mid \theta); \theta \in \Theta\}$ is stochastically increasing if the distribution $F(y \mid \theta)$ is a decreasing function in θ for all fixed y and stochastically decreasing if the distribution $F(y \mid \theta)$ is an increasing function in θ for all fixed y.

Fisher (1930) suggested that if a distribution $F(y \mid \theta)$ is stochastically increasing, define

$$G(\theta \mid y) = 1 - F(y \mid \theta) \tag{14}$$

as the posterior distribution of θ conditional on the observed value $Y = y$. And its density θ is

$$g(\theta \mid y) = G'(\theta \mid y) = -\frac{\partial}{\partial\theta} F(y \mid \theta), \tag{15}$$

the derivative of G. (Of course, he did not forget to define the density f of Y as $f(y \mid \theta) = \frac{\partial}{\partial y} F(y \mid \theta)$, in the continuous sample space case.) Evidently, as we have pointed out earlier, he thought that all distributions are stochastically increasing, differentiable with respect to y and θ and taking values in $(0, 1)$ inclusively. We shall see later that it is not always so.

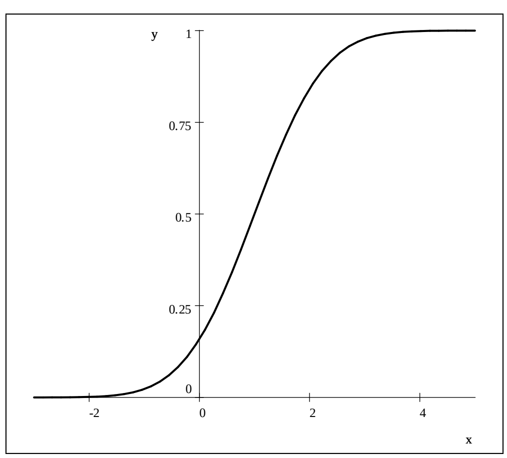
How could one perceive that $G = 1 - F$ in (14) defines a distribution for the unknown θ for all fixed y? It works for all distributions with θ a location or a scale parameter case, (and more)! We believe that it is one of the many Fisher's ingenious discoveries in Statistics.

Let us see two examples of stochastically increasing distributions F and their corresponding $G = 1 - F$:

Example 3 *The normal $(\theta, 1)$ distribution with $Y = \overline{X}$, the sample mean. Then $F(y \mid \theta) = \int_{-\infty}^{y} \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-n(x-\theta)^2/2} dx$ is stochastically increasing so*

$$G(\theta \mid y) = \int_{y}^{\infty} \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-n(x-\theta)^2/2} dx.$$

The graph of $G(\theta \mid 1)$, with $n = 1$, is shown below:

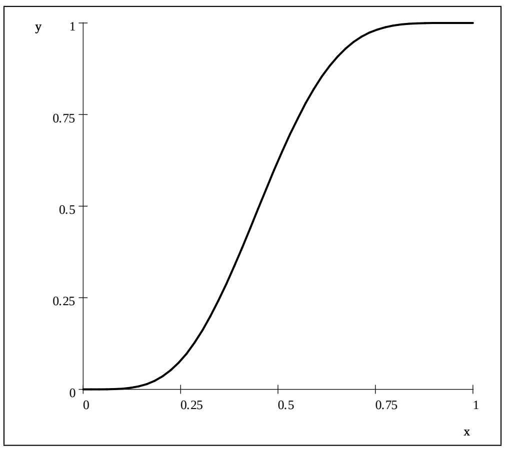


The function $G(\theta \mid y)$ is a perfect continuous distribution for $-\infty < \theta < \infty$ for all y fixed.

Example 4 *The Bernoulli* (θ) *distribution with* $Y = \Sigma_{i=1}^{n} X_i$, *the sum of a random sample. Then* Y *has the binomial* (n, θ) *distribution with* $F(y \mid \theta) = \sum_{x=0}^{y} \binom{n}{x} \theta^x (1-\theta)^{n-x}$, *which is stochastically increasing and*

$$G(\theta \mid y) = \sum_{x=y+1}^{n} \binom{n}{x} \theta^x (1-\theta)^{n-x},$$

for $y = 0, 1, ..., n-1$. *The graph of* $G(\theta \mid 4)$ *for* $n = 10$ *is as follow:*



It is a perfect continuous distribution in θ for $0 \leq \theta \leq 1$ and all $y = 0, 1, ..., n-1$. The case for $y = n$ is explained in detail in Wang (2000).

In both examples above, despite of their small curves, the derivatives g of G do not have nice known forms. In fact, according to our investigation, g does not match any known density function.

Definition 5 *A parameter* $\theta \in \Theta = (-\infty, \infty)$ *is a location parameter of* F *if* $F(y \mid \theta) = F(y - \theta \mid 0)$ *and that* $\theta \in \Theta = (0, \infty)$ *is a scale parameter of* F *if* $F(y \mid \theta) = F(y/\theta \mid 1)$ *for* $y \geq 0$ *and* $= 0$ *for* $y < 0$.

In the above definition, it is assumed that 0 is a location parameter and 1 is a scale parameter. The location and the scale parameters are with respect to the distribution F of the sufficient statistic Y, not with respect to that of the original population. For example, in the normal population, the variance $\theta = \sigma^2$ is a scale parameter of the distribution of the sample variance $Y = S^2$, not that of the normal population. We restrict Y to be $\Pr(Y < 0) = 0$ in case θ is a scale parameter. It is a logical restriction, because a scale parameter is positive and its corresponding statistic should be non-negative.

Lemma 6 *a) If θ is a location or scale parameter of a continuous distribution F, then F is stochastically increasing. b) If $-\theta$ is a location parameter or $1/\theta$ is a scale parameter of a continuous distribution F, then F is stochastically decreasing.*

Proof. a). For fixed y, if θ is a location parameter and $-\infty < \theta_1 < \theta_2 < \infty$, then $F(y \mid \theta_1) = F(y - \theta_1 \mid 0) \geq F(y - \theta_2 \mid 0) = F(y \mid \theta_2)$, because $y - \theta_1 > y - \theta_2$, and if θ is a scale parameter and $0 < \theta_1 < \theta_2 < \infty$, then $F(y \mid \theta_1) = F(y/\theta_1 \mid 1) \geq F(y/\theta_2 \mid 1) = F(y \mid \theta_2)$, because $y/\theta_1 > y/\theta_2$ for $y > 0$.

b). It can be proved in an identical manner. ■

Lemma 6 covers a lot of ground. But there are many other distributions not covered by it. The binomial (n, θ) and the Poisson (θ) distributions are both stochastically increasing in θ. The negative binomial (r, θ) distribution is stochastically decreasing in θ. The parameter θ in these distributions are neither location nor scale parameters.

If a statistic Y and a parameter θ related by a distribution function $F(y \mid \theta) = \Pr(Y \leq y \mid \theta)$, is the family $\{F(y \mid \theta); \theta \in \Theta\}$ of distribution functions automatically either stochastically monotone? That is for $\theta_1 < \theta_2$, $F(y \mid \theta_1) \leq (\geq)\ F(y \mid \theta_2)$ for all y and $F(y \mid \theta_1) < (>)\ F(y \mid \theta_2)$ for some y. The answer is "not necessary". The following counter example which answers the above question was communicated to us by Neil Schwertman (2000).

Example 7 *Let Y be normal $(0, \theta)$ so that $F(y \mid \theta) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\theta}} e^{-x^2/2\theta} dx$. It is easy to check that for $\theta_1 = 2 < 9 = \theta_2$, $F(y \mid \theta_1) < F(y \mid \theta_2)$ for all $-\infty < y < 0$ and $F(y \mid \theta_1) > F(y \mid \theta_2)$ for all $0 < y < \infty$.*

In this example θ is neither a location parameter nor a scale parameter of Y. And Y is not a sufficient statistic of θ. If we make the transformation $Z = Y^2$, then Z is sufficient statistic for θ and θ becomes the scale parameter of Z. By Lemma 6 the distribution $F(z \mid \theta) = \Pr(Z \leq z \mid \theta)$ is stochastically increasing in θ. We believe, but have not been able to prove it, that under the restriction that Y is a sufficient statistic of θ, the distribution function $F(y \mid \theta)$ is either stochastically increasing or decreasing. We also discovered even more that for all fixed y the monotonicity of $F(y \mid \theta)$ is strict in θ. That is if F is stochastically increasing in θ, then for all fixed y, $F(y \mid \theta_1) > F(y \mid \theta_2)$ for all $\theta_1 < \theta_2$. Its proof for θ is a location or a scale parameter is straight-forward, but in general, we have not been able to do it.

Lindley (1958) gives necessary and sufficient conditions that the derivative $-\partial F(y \mid \theta) / \partial\theta$ is a Bayesian posterior distribution for some prior distribution of θ. Lindley assumed that all the distributions are stochastically increasing.

Earlier in this section we stated that an ideal distribution function must take values in $(0, 1)$ inclusively for all fixed y. That is for the parameter space $\Theta = [\theta_1, \theta_2]$, with one or both limits may be open and/or infinite, then, for all $y \in R$, $F(y \mid \theta_1) = \lim_{\theta \to \theta_1^+} F(y \mid \theta) = 1 \ (= 0)$ and $F(y \mid \theta_2) = \lim_{\theta \to \theta_2^-} F(y \mid \theta) = 0 \ (= 1)$, for a stochastically increasing (decreasing) distribution F in θ. We have checked extensively and found that the majority of distributions do take value in $(0, 1)$ inclusively, but there are a few exceptions. Interestingly, the very first example — distribution of the sample correlation coefficient, Fisher (1930) used failed this condition. (He used the 95-percentile of the distribution of the sample correlation coefficient to numerically demonstrate the relation $G = 1 - F$. See Example 11 later in this section.)

For fixed y, we denote

$$\begin{cases} F(y \mid \theta_1) = \lim_{\theta \to \theta_1^+} F(y \mid \theta), \\ F(y \mid \theta_2) = \lim_{\theta \to \theta_2^-} F(y \mid \theta). \end{cases} \tag{16}$$

We are now ready to formally construct the fiducial distributions for the continuous parameter space:

Definition 8 *For the continuous parameter space $\Theta = [\theta_1, \theta_2]$ with sufficient statistic $Y = t(\mathbf{X})$ having the distribution $F(y \mid \theta) = \Pr(Y \leq y \mid \theta)$, conditional on $Y = y$, the fiducial distribution $G(\theta \mid y)$ is*

$$G(\theta \mid y) = \frac{F(y \mid \theta_1) - F(y \mid \theta)}{F(y \mid \theta_1) - F(y \mid \theta_2)}, \tag{17}$$

if $F(y \mid \theta)$ is stochastically increasing in θ, and is

$$G(\theta \mid y) = \frac{F(y \mid \theta) - F(y \mid \theta_1)}{F(y \mid \theta_2) - F(y \mid \theta_1))}, \tag{18}$$

if $F(y \mid \theta)$ is stochastically decreasing in θ. The two limits $F(y \mid \theta_1)$ and $F(y \mid \theta_2)$ are as defined by (16).

If $F(y \mid \theta_1) = 1$, $F(y \mid \theta_2) = 0$ and $F(y \mid \theta)$ is stochastically increasing in θ, (17) reduces to

$$G(\theta \mid y) = 1 - F(y \mid \theta). \tag{19}$$

(19) was taken as the fiducial distribution in Fisher (1930) and in many others' writings. The next Theorem, together with Lemma 6, possibly explain Fisher's and others' reasonings in their fiducial argument.

Theorem 9 *A distribution F satisfies $F(y \mid \theta_1) = 1$, $F(y \mid \theta_2) = 0$, if $\theta \in \Theta = (-\infty, \infty)$ is its location parameter or $\theta \in \Theta = (0, \infty)$ is its scale parameter.*

Proof. Assume θ is a location parameter. For fixed y, $\lim_{\theta\to-\infty} F(y \mid \theta) = \lim_{\theta\to-\infty} F(y-\theta \mid 0) = \lim_{s\to\infty} F(s \mid 0) = 1$, and $\lim_{\theta\to\infty} F(y \mid \theta) = \lim_{\theta\to\infty} F(y-\theta \mid 0)$ $= \lim_{s\to-\infty} F(s \mid 0) = 0$.

Next assume θ is a scale parameter. For fixed y, $\lim_{\theta\to 0^+} F(y \mid \theta) = \lim_{\theta\to 0^+} F(y/\theta \mid 1)$ $= \lim_{s\to\infty} F(s \mid 1) = 1$, and $\lim_{\theta\to\infty} F(y \mid \theta) = \lim_{\theta\to\infty} F(y/\theta \mid 0) = \lim_{s\to 0} F(s \mid 1)$ $= 0$. ■

According to our investigation, absolute majority of the fiducial distributions fall into the type (19). Let us state it as a Corollary to Theorem 9.

Corollary 10 *If θ is a location or a scale parameter of $F(y \mid \theta)$, then the fiducial distribution is* (19).

If $F(y \mid \theta_1) = 0$, $F(y \mid \theta_2) = 1$ and $F(y \mid \theta)$ is stochastically decreasing in θ, (18) becomes

$$G(\theta \mid y) = F(y \mid \theta). \tag{20}$$

This case is rarely mentioned in the literature. A discrete example of (20) is the negative binomial distribution (Example 14) and a continuous example is the exponential distribution $F(y \mid \theta) = 1 - e^{-y\theta}$, $(y > 0)$.

It is easy to check that all the fiducial distributions $G(\theta \mid y)$ possess the properties that:

$$\begin{cases} 1)\ \lim_{\theta\to-\infty} G(\theta \mid y) = 0; \\ 2)\ \lim_{\theta\to\infty} G(\theta \mid y) = 1; \\ 3)\ \theta' < \theta'' \ \text{ implies } \ G(\theta' \mid y) \le G(\theta'' \mid y) \ \text{ for all } \ y. \end{cases}$$

Therefore $G's$ are bona fide distribution functions of θ, and *fiducial probabilities* over the parameter space Θ can be defined accordingly. If G is differentiable with respect θ for all fixed y, the *fiducial density* is

$$g(\theta \mid y) = -\partial F(y \mid \theta)/\partial\theta \left(F(y \mid \theta_1) - F(y \mid \theta_2)\right),$$

for stochastically increasing distribution $F(y \mid \theta)$. And

$$g(\theta \mid y) = \partial F(y \mid \theta) / \partial\theta \left(F(y \mid \theta_2) - F(y \mid \theta_1)\right),$$

for stochastically decreasing distribution $F(y \mid \theta)$. The fiducial densities for special cases (19) and (20) are $g(\theta \mid y) = -\partial F(y \mid \theta)$ and $g(\theta \mid y) = \partial F(y \mid \theta)$, respectively, which can be found in Pedersen (1978) on page 149.

Let us get back to the Lehmann's example (5) in Section 1. In the normal $(\theta, 1)$ distribution case, by Lemma 6 and Corollary 10, the fiducial distribution is (19). Hence

$$\begin{aligned} & \Pr(0.18 < \theta < 4.10 \mid X = 2.14) \\ = \; & G(4.10 \mid 2.14) - G(0.18 \mid 2.14) \\ = \; & F(2.14 \mid 0.18) - F(2.14 \mid 4.10) \\ = \; & \Phi(1.96) - \Phi(-1.96) = 0.95. \end{aligned}$$

Let us have some more examples. The sample correlation coefficient distribution was the very first example Fisher used in his fiducial argument. The parameter θ is neither a location nor a scale parameter, but the distribution F is stochastically increasing in θ.

Example 11 *(Sample correlation coefficient.).* $\Omega = \Theta = [-1, 1]$. *The parameter* $\theta = \rho$ *is the correlation coefficient of the bivariate normal population defined by* (11) *and the statistic is the sample correlation coefficient* $Y = R$ *defined by* (9). *The density of* Y *was derived by Fisher* (1915) *and there are several expressive forms proposed by several authors over the years for the density and distribution. (See Johnson et al* (1995) *Chapter 32). For* $n \geq 3$*, one form of the density is*

$$f_n(y \mid \theta) = \frac{(n-2)\left(1-\theta^2\right)^{(n-1)/2}\left(1-y^2\right)^{(n-4)/2}}{\pi} \int_0^\infty \frac{dw}{(\cosh w - \theta y)^{n-1}}. \qquad (21)$$

For $n = 3$, *the distribution of* Y *can be written as*

$$F_3(y \mid \theta) = \frac{\cos^{-1}(-y) - \pi^{-1}\theta\left(1-y^2\right)\left(1-\theta^2 y^2\right)^{-1/2}\cos^{-1}(-y\theta)}{\pi}. \qquad (22)$$

As we have stated earlier that this is an example of $0 < F_n(y \mid \theta_2) < F_n(y \mid \theta_1) < 1$, *for all* $-1 < y < 1$ *and* $n \geq 3$. *Some values of* $F(y \mid \theta_2)$, $F(y \mid \theta_1)$ *are displayed in Table I below:*

TABLE *I*

y	$F_3(y \mid \theta_1)$	$F_3(y \mid \theta_2)$
.9	*.87635*	*.73761*
.6	*.78*	*.52535*
.3	*.71936*	*.41571*
0	*.65915*	*.34085*
-.3	*.58429*	*.28064*
-.6	*.47465*	*.22*
-.9	*.26239*	*.12365*

The graph of the fiducial correlation coefficient distribution $G_3(\theta \mid y)$, for $n = 3$ and $y = -.6$, 0 and .9 are shown in *Figure I*. For $y = .9$, it is the bottom one, a convex upward curve; for $y = 0$, it is the middle one, a straight line; and for $y = -.6$, it is the top one, a concave downward curve. It can be observed that if $y > 0$ the graph is convex upward from -1 to 1, and it $y < 0$ it is just the opposite, concave downward. For $y = y_0 > 0$ and $y = -y_0$ the graphs are symmetric with respect to the line $y = 0$.

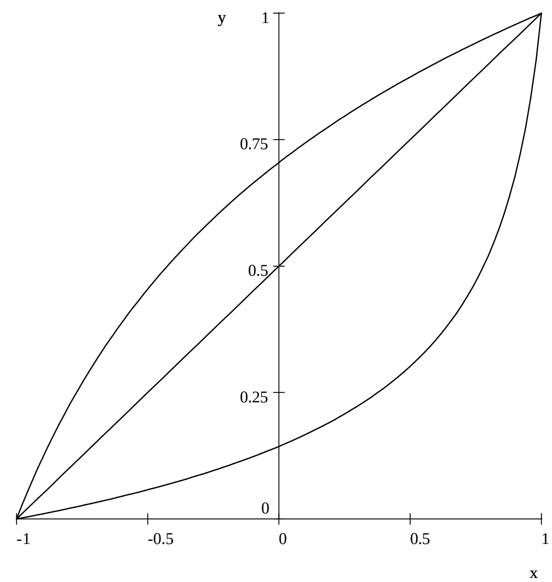


FIGURE I

THE FIDUCIAL CORRELATION COEFFICIENT DISTRIBUTUION

Example 12 *(F-distribution.)* $\Omega = \Theta = (0, \infty)$. *From two independent normal populations, the parameter* $\theta = \sigma_x^2/\sigma_y^2$ *is defined by* (12), *the ratio of two variances and the statistic* $Y = S_x^2/S_y^2$, *by* (10), *the ratio of two sample variances of two independent random samples of sizes* $n+1$ *and* $m+1$. *So that the pivotal quantity* $Z = Y/\theta$ *has the standard F-distribution with* n *and* m *d.f., hence* θ *is the scale parameter of* Y. *The distribution* $F(y \mid \theta)$ *is*

$$F_{n,m}(y \mid \theta) = \int_0^y \frac{1}{\theta} f_{n,m}(t/\theta)\, dt \tag{23}$$

where $f_{n,m}$ *denoting the standard density of the F-distribution with* n *and* m *d.f. By* (19) *the fiducial F-distribution with* n *and* m *d.f.* $G_{n,m}(\theta \mid y) = 1 - F_{n,m}(y \mid \theta)$.

The graph of the fiducial F-distribution $G_{6,10}(\theta \mid y)$ for $n = 6$ and $m = 10$ at $y = 1.5$ is displayed in *Figure II*. It starts from 0 and reaches 1, as the value of θ

increases from 0 to ∞. The shape of the graphs stays unchanged as y increases, but moves to the right-side.

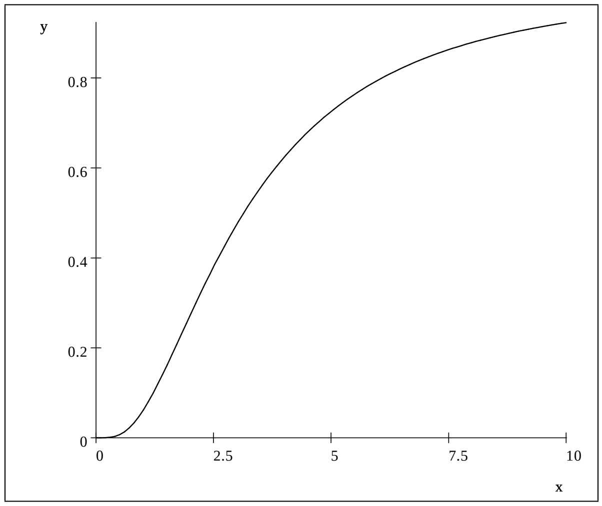


FIGURE II

THE FIDUCIAL F − DISTRIBUTION

Let us consider an example of stochastically increasing distribution with discrete sample space. We have already showed the binomial distribution in Example 4. The Poisson distribution is next.

Example 13 *(Poisson distribution.)* $\Omega = \{0, 1, 2, \ ... \ \}$ *while* $\Theta = [0, \ \infty)$. *The random variable* Y *has the Poisson distribution with intensity* θ. *The distribution* $F(y \mid \theta)$ *is*

$$F(y \mid \theta) = \sum_{i=0}^{y} e^{-\theta} \theta^i / i!, \quad y = 0, 1, 2, ... \ \ \textit{and} \quad \theta \geq 0.$$

F is stochastically increasing, therefore the fiducial Poisson distribution is $G(\theta \mid y) = 1 - F(y \mid \theta)$, *for* $\theta > 0$ *and* $y = 0, 1, 2, ...$.

The graphs of the fiducial Poisson distribution for $y = 0, 3$ and 6 are displayed in *Figure III*. The left-most curve is for $y = 0$, the middle one is for $y = 3$ and the right-most curve is for $y = 6$. They all tend to 1 as θ tends to infinite.

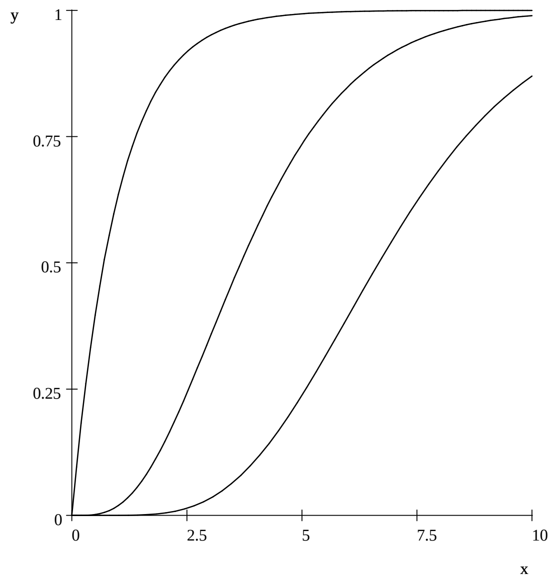


FIGURE III

THE FIDUCIAL POISSON DISTRIBUTUION

Next let us consider an example of stochastically decreasing distribution with discrete sample space.

Example 14 *(Negative binomial.)* $\Omega = \{0, 1, 2, \ ... \ ,\}$ *and* $\Theta = [0, 1]$. *The negative binomial distribution is*

$$F(y \mid \theta) = \sum_{x=0}^{y} \binom{r+x-1}{x} \theta^{r} (1-\theta)^{x}, \quad y = 0, 1, ...; \ r = 1, 2, ...; \ 0 \leq \theta \leq 1.$$

F *is stochastically decreasing with* $F(y \mid 0) = 0$ *and* $F(y \mid 1) = 1$ *for all* $y \geq 0$. *By* (20) *the fiducial negative binomial distribution is* $G(\theta \mid y) = F(y \mid \theta)$.

The graphs of the fiducial negative binomial distribution for the pairs $(r, y) = (2, 2)$, $(2, 5)$ and $(5, 2)$ are displayed in Figure IV. The left-most curve is for the pair $(2, 5)$, the middle curve is for $(2, 2)$ and the right-most curve is for $(5, 2)$. If the value of r increases, the curve tends to sift to the right, while if the value of y increases, the curve tends to sift to the left.

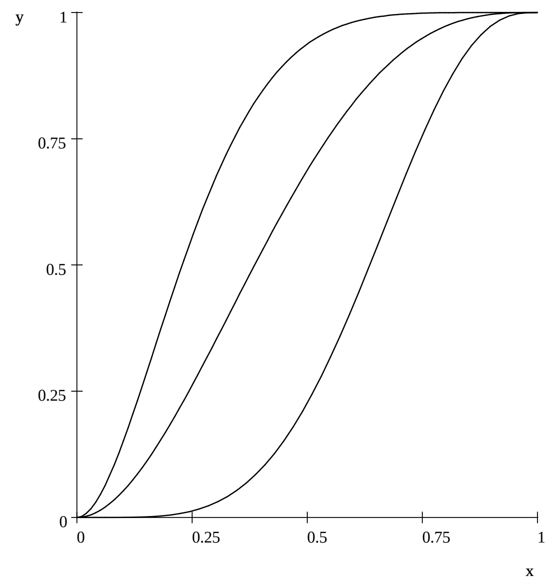


FIGURE IV

THE FIDUCIAL NEGATIVE BINOMIAL DISTRIBUTION

4. SOME APPLICATIONS

4.1. Evaluation of Pre-data Probability and P-value

Let us pose to illustrate some applications of the fiducial distributions (17) and (18). First let us verify the probability $\Pr(3.45 < \theta \mid X = 3.45) = 1/2$ in (2), explain why Student's $\Pr(y_1 \leq \theta \leq y_2) = 1/2$ in (3) is false and how to rectify the incoherence problem of *p-value*. They are relative easy to settle and we shall then go into the problem construction and post-data evaluation of the pre-data confidence intervals in much more details.

For $\Pr(3.45 < \theta \mid X = 3.45) = 1/2$. Without loss of generality, assume F is stochastically increasing in θ so that $G = 1 - F$.

$$\begin{aligned} & \Pr(\theta < 3.45 \mid X = 3.45) \\ = & \ G(3.45 \mid 3.45) \\ = & \ 1 - F(3.45 \mid 3.45) \\ = & \ 1 - 1/2 = 1/2, \end{aligned}$$

because F is continuous and symmetric with respect to θ, $F(t \mid t) = 1/2$ for all $-\infty < t < \infty$. Similarly, $\Pr(3.45 < \theta \mid X = 3.45) = 1/2$.

For Student's $\Pr(y_1 \leq \theta \leq y_2) = 1/2$. Fisher could have misinterpreted it as

$$\Pr(y_1 < \theta < y_2) = 1 - \Pr(y_2 < \theta) - \Pr(\theta < y_1). \quad (24)$$

And further misjudged the right-hand-side of (24) as $\Pr(y_2 < \theta) = \Pr(x_1 \text{ and } x_2 < \theta)$ and $\Pr(\theta < y_1) = \Pr(\theta < x_1 \text{ and } x_2)$, and then concluded that the right-and-side of (24) equals $1 - 1/4 - 1/4 = 1/2$. These arguments are false. The proper way to write (3) is

$$\Pr(y_1 < \theta < y_2 \mid Y_1 = y_1 < Y_2 = y_2). \quad (25)$$

It is a post-data probability conditional on two realized values of Y_1 and Y_2. In our construction of the fiducial distribution, it requires that θ and y be interdependent through the function $F(y \mid \theta)$ which is either stochastically increasing or decreasing in θ for all fixed y. This interdependence of θ and (y_1, y_2) through $F(y_1, y_2 \mid \theta)$ and the monotonicity of $F(y_1, y_2 \mid \theta)$ in θ for fixed y_1 and y_2 must be established before an assertion on (25) can be made. Therefore our arguments in the last section are not applicable here.

On the other hand, had θ have a distribution conditional on (y_1, y_2), then θ would have a mean and a variance which are functions of y_1 and y_2. If not, it can lead to a contradiction, because the variance of θ would have to be independent of the length $y_2 - y_1$. The following is a sharp example: let F be uniform over $(\theta - 1/2, \theta + 1/2)$, then $\Pr(y_1 < \theta < y_2 \mid Y_1 = y_1 < Y_2 = y_2) = 1$, for $y_2 - y_1 > 1/2$, and $\Pr(y_1 < \theta < y_2 \mid Y_1 = y_1 < Y_2 = y_2) < 1$, for $y_2 - y_1 < 1/2$. (See De Groot (1986), p.400). Therefore $\Pr(y_1 < \theta < y_2) \neq 1/2$, not as claimed by Gosset and Fisher. In this example F is symmetry with respect to θ, a stricter condition than requiring θ being the median.

A related problem along the line is in Jeffreys (1932). Jeffreys asserted that "the probability is 1/3 that a third observation X_3 will lie between the first two", i.e. $\Pr(y_1 < x_3 < y_2) = 1/3$. This is trivially true because it is simply a problem of permutation of three numbers x_1, x_2 and x_3, counting the proportion of x_3 being in the middle of the other two in all 3! possible permutations. This problem involves no unknown parameter and requires not symmetry condition.

For Lehmann's post-data probability (5). We have already demonstrated it in section 3 that, with F being normal $(\theta, 1)$, $\Pr(0.18 < \theta < 4.10 \mid X = 2.14) = 0.95$ which is equal to the pre-data probability $\Pr(X - 1.96 < \theta < X + 1.96)$. Thus Lehmann's belief that without θ being random (5) can not be evaluated is rebuffed.

For the incoherence of *p-value*. In section 2, we have an example that according to the current practice of the *p-value* one could simultaneously reject the one-sided null hypothesis $H_0 : \theta \leq 0$ and accept the point-null hypothesis $H_0 : \theta = 0$. Since $\{\theta : \theta = 0\} \subseteq \{\theta : \theta \leq 0\}$, the decision is erroneous. In other words, the current *p-value* is incoherent. What is wrong? Because *p-value* is a post-data probability conditional on a realized value of the test statistic $Y = y$. It is clearly belong to the post-data inference, but being used as a device in the pre-data inference.

In many text books on statistics, *p-value* is often not clearly defined. Here is a good one on page 77 in Garthwaite et al (2002). The *p-value* π with test statistic Y for testing H_0 versus H_1 is

$$\pi(y) = \text{" Pr}\left(Y \text{ is at least as extreme as the observed value } y \mid H_0\right).\text{"}$$

So that in testing $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$, "extreme" means "$\{Y \geq y\}$", and (7) is correct. In testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$, "extreme" means "$\{Y \geq y\} \cup \{Y \leq y\}$", it does not follow that (8) is correct. There is a difference in the logics of the post-data and the pre-data inferences. In testing the point-null hypotheses, according to the frequentist's logic, y can be an extreme value on the left or on the right (when Y is realized), but Y has yet to be observed, and hence the probability (8). But to the fiducialist, it is much simpler, because $Y = y$ has already been observed. An observed value y is either left-extreme or right-extreme, not both. And (8) should be written as

$$\pi_0(\overline{x}) = \begin{cases} \Pr\left(\overline{X} \geq \overline{x}\right), & \text{if } \overline{x} > 0, \\ \Pr\left(\overline{X} \leq \overline{x}\right), & \text{if } \overline{x} < 0. \end{cases}$$

The factor 2 in (8) must be deleted from a fiducial point of view.

Clearly *p-value* belongs to the post-data inference domain and should be treated accordingly.

The proper way to define a *p-value* π is in terms of the fiducial distribution G. For simplicity we assume that Y is a point estimate of θ, as in the normal case above. (In the distributions such as binomial and Poisson a point estimate of θ is simply $\widehat{\theta} = Y/n$, and "extreme" can be equivalently expressed in Y, as in $\widehat{\theta}$.) When testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, if $Y = y$ is observed,

$$\pi(y) = 1 - G(\theta_0 \mid y).$$

When testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, if $Y = y$ is observed,

$$\pi_0(y) = \begin{cases} 1 - G(\theta_0 \mid y), & \text{if } y > \theta_0, \\ G(\theta_0 \mid y), & \text{if } y < \theta_0. \end{cases}$$

This solves the incoherence problem of current definition of the *p-value*.

4.1. Problems with Pre-data Confidence Intervals

Construction of fiducial intervals and post-data evaluation of pre-data random intervals were studied in some details in Wang (2000). For the sake of completeness of this paper, we shall briefly go over some essentials and get into some other topics not covered there.

A random interval $I(Y) = \left(\widehat{\theta}_L(Y), \widehat{\theta}_U(Y)\right)$ obtained by any means becomes a non-random interval $I(y) = \left(\widehat{\theta}_L(y), \widehat{\theta}_U(y)\right)$ once $Y = y$ is observed. By using the fiducial distribution G, its post-data coverage probability is simply

$$\Pr\left(\widehat{\theta}_L(y) < \theta < \widehat{\theta}_U(y) \mid y\right) = G\left(\widehat{\theta}_U(y) \mid y\right) - G\left(\widehat{\theta}_L(y) \mid y\right).$$

For example, with $Y = n\overline{X}$, when $\overline{X} = \overline{x}$ is observed, the Wald interval (6) becomes $I(\overline{x}) = \left(\widehat{\theta}_L(\overline{x}), \widehat{\theta}_U(\overline{x})\right)$ where $\widehat{\theta}_L(\overline{x}) = \overline{x} - z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}$ and $\widehat{\theta}_U(\overline{x}) = \overline{x} + z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}$. The post-data coverage probability is then

$$\begin{aligned} & \Pr\left(\widehat{\theta}_L(\overline{x}) < \theta < \widehat{\theta}_U(\overline{x}) \mid \overline{x}\right) \\ = \; & G\left(\widehat{\theta}_U(\overline{x}) \mid \overline{x}\right) - G\left(\widehat{\theta}_L(\overline{x}) \mid \overline{x}\right), \\ = \; & \Pr\left(Y < y \mid \widehat{\theta}_L(\overline{x})\right) - \Pr\left(Y \le y \mid \widehat{\theta}_U(\overline{x})\right), \end{aligned} \tag{26}$$

where $G(\theta \mid y) = \sum_{i=y+1}^{n} \binom{n}{i} \theta^i (1-\theta)^{n-i}$, for $y = n\overline{x}$. (See Example 4.) For $n = 50$, $1-\alpha = .95$, $y = 8$, $\widehat{\theta}_L(8/50) = .0584$ and $\widehat{\theta}_U(8/50) = .2616$, $\Pr(.0584 < \theta < .2616 \mid 8/50)$ $= .9268$. (All post-data coverage probabilities of $I(\overline{x})$ for $n = 50$, $1-\alpha = .95$ and $Y = 0, 1, \ldots$ are tabulated in Table 1 on page 110 in Wang (2000).)

The post-data coverage probability (26) is void of the parameter θ, thus provides an alternative and effective way to evaluate any pre-data random interval and to compare two or more random intervals, especially when pivotal quantities do not exist.

In general, for a given distribution F, a $(1-\alpha)100\%$ post-data confidence interval $(\theta_1(y), \theta_2(y))$, conditional on $Y = y$, is simply a pair $\theta_1(y)$ and $\theta_2(y)$ satisfying

$$G\left(\widehat{\theta}_1(y) \mid y\right) - G\left(\widehat{\theta}_2(y) \mid y\right) = 1 - \alpha. \tag{27}$$

There are many solutions to (27), among them the optimum one is the one with the shortest width $\left|\widehat{\theta}_U(y) - \widehat{\theta}_L(y)\right|$. Let us formally define a fiducial interval below which is more general than the one given in Definition 3 in Wang (2000).

Definition 15 *Given a statistic Y with distribution F, a $(1-\alpha)\,100\%$ fiducial interval of a parameter θ conditional $Y = y$ is an interval $(\theta_1(y), \theta_2(y))$ satisfying equation* (27).

By Definition 15, in the binomial distribution case, for $1-\alpha = .95$ with $n = 10$ and $y = 3$, the equal-tailed fiducial interval for θ is $(.0667,\ .6525)$ while the optimum fiducial interval is $(.0479,\ .6226)$. The optimum fiducial intervals for θ in the binomial distribution case are compiled in Wang and Chang (2002) for $n = 1, 2, ...50$, $1-\alpha = .90, .95$ and $.99$.

In Wang (2000), there are two examples of continuous sample space, the normal $(\theta, 1)$ and the uniform $(0, \theta)$ distributions. Both are stochastically increasing, because θ is a location parameter in the normal $(\theta, 1)$ case and a scale parameter in the uniform $(0, \theta)$ case. In both examples, it was shown that if $C(Y)$ is the random interval for θ obtained, by inverting tests of point-null hypotheses versus a two-sided alternative, then $C(y)$ is the corresponding fiducial interval obtained by (27). The main reason for this relation is the existence of pivotal quantities in both distributions. We shall formally develop it as a theorem in this section.

Definition 16 *A function $Q(Y, \theta)$ of a statistic Y and a parameter θ is said to be a pivotal quantity if its distribution does not depend upon any unknown parameters.*

The distribution function and the probability density function of the pivotal quantity $Q(Y, \theta)$ shall be denoted by H and h.

If the distribution of Y does not depend on any parameters other than θ, then $Q(Y, \theta) = Y - \theta$, if θ is the location parameter and $Q(Y, \theta) = Y/\theta$, if θ is the scale parameter, are two simple examples of pivotal quantities. Another well-known one is $Q(Y,\theta) = F(Y \mid \theta)$, if the sample space is continuous. There are two problems: One is that it may not be unique. For example, for the normal population with variance σ^2 unknown, in addition to the student's $T = \sqrt{n}(\overline{X} - \theta)/S$, the lesser-known one is $T^* = \sqrt{n}(\overline{X} - \theta)/R$, where $R = Y_n - Y_1$ is the sample range. (Y_i's denote the order statistics of a random sample). The other is it may not exist. The binomial distribution with success probability θ and the Poisson distribution with intensity θ are two well known examples. In fact for all the discrete distributions with continuous parameter space Θ pivotal quantities can not be constructed. We are concern here mainly with distributions having pivotal quantities.

4.2.1. Inverting Tests of Hypotheses Via Pivotal Quantities.—

For the frequentists when testing a point-null hypothesis: $H_0 : \theta = \theta_0$ against the two-sided alternative $H_0 : \theta \neq \theta_0$, in the present of a pivotal quantity $Q(Y,\theta)$, the null hypothesis H_0 is rejected if and only if

$$Q(Y,\theta_0) < \rho(\alpha/2) \text{ or } Q(Y,\theta_0) > \rho(1-\alpha/2),$$

where $\rho(\alpha)$ is the α-percentile of the distribution H, i. e. $H(\rho(\alpha)) = \alpha$, for all $0 < \alpha < 1/2$. (For convenience we take equal-tailed critical regions.) For fixed θ_0 the acceptance region of the test can be written as

$$\mathbf{A}(\theta_0 \mid \alpha) = \{Y : \rho(\alpha/2) \leq Q(Y,\theta_0) \leq \rho(1-\alpha/2)\}.$$

And for each $Y = y$, the set $\mathbf{I}(y \mid \alpha) = \{\theta_0 : \rho(\alpha/2) \leq Q(y,\theta_0) \leq \rho(1-\alpha/2)\}$ is a subset of the parameter space Θ. By letting

$$Q(y,\theta_1(y)) = \rho(1-\alpha/2) \text{ and } Q(y,\theta_2(y)) = \rho(\alpha/2), \tag{28}$$

the inverse of the acceptance region $\mathbf{A}(\theta_0 \mid \alpha)$ is

$$\mathbf{I}(y \mid \alpha) = \{\theta : \theta_1(y) \leq \theta \leq \theta_2(y)\}. \tag{29}$$

Because F is increasing in θ it is necessary that $\theta_1(y) < \theta_2(y)$ and that (29) is an interval. Consequently, a $(1-\alpha)\,100\%$ pre-data confidence interval for θ is the random interval

$$\mathbf{I}(Y \mid \alpha) = \{\theta : \theta_1(Y) \leq \theta \leq \theta_2(Y)\}\,.$$

And the confidence coefficient of the interval $\mathbf{I}(Y \mid \alpha)$ is

$$P_\theta(\theta \in \mathbf{I}(Y \mid \alpha)) = P_\theta(Y \in \mathbf{A}(\theta \mid \alpha)) = 1-\alpha. \tag{30}$$

4.2.2. Fiducial Intervals Via Pivotal Quantities.—

To derive the fiducial interval $[\theta_L(y), \theta_U(y)]$ for θ, in accordance with (27), after $Y = y$ is realized, the upper confidence limit $\theta_U(y)$ is the solution to the equation $F(y \mid \theta_U(y)) = \alpha/2$. Since $F(y \mid \theta) = H(Q(y,\theta))$ for all θ, the upper limit $\theta_U(y)$ is the solution to the equation $Q(y, \theta_U(y)) = \rho(\alpha/2)$, where $\rho\,(\alpha)$ is the α-percentile of the distribution H. Similarly, the lower limit $\theta_L(y)$ is the solution to $Q(y, \theta_L(y)) = \rho(1-\alpha/2)$. Thus, in terms of the pivotal quantity $Q\,(Y,\theta)\,,$ the two quantities $\theta_L\,(y)$ and $\theta_U\,(y)$ are solutions to the equations in (31) below

$$\begin{cases} Q(y, \theta_U(y)) = \rho(\alpha/2), \\ Q(y, \theta_L(y)) = \rho(1-\alpha/2). \end{cases} \tag{31}$$

The above two equations are identical to those in (28) for deriving the two limits $\theta_1(y)$ and $\theta_2(y)$ of the frequentist interval $\mathbf{I}(Y \mid \alpha)$, except that in the frequentist logic y is not really a fixed realization of Y, while in the fiducialist's, $Y = y$ is a realized value of Y. We have just proved the following theorem.

Theorem 17 *If pivotal quantities exist and the distribution F of Y is stochastically increasing in θ, then the random interval $\mathbf{I}(Y \mid \alpha)$ is the frequentist interval for θ if and only if the conditional interval $\mathbf{I}(y \mid \alpha)$ is the fiducial interval.*

In the above theorem, we considered only for F is stochastically increasing, because as we have stated earlier that absolute majority of distributions are stochastically increasing. For the stochastically decreasing case, it can be prove in an identical manner with minor modifications.

According to the above theorem, if a pivotal quantity exists and the distribution F of Y is stochastically increasing in θ, to construct a fiducial interval for θ it is sufficient to construct a frequentist interval and then to conditionalize it on $Y = y$. The converse also holds.

With $Y = y$ a realized value, when a pivotal quantity does not exist, to find a fiducial interval $\left(\widehat{\theta}_1(y), \widehat{\theta}_2(y)\right)$, it is a simple application of the equation (27).

There are examples in which it is a fairly complicated process to derive (the likelihood ratio) tests of hypotheses and then inverting them to form a frequentist confidence interval for a parameter θ. For examples, a) testing the mean θ of a normal population when σ^2 is unknown (Hogg & Craig (1995), p. 413, example 1); b) testing the equality of two normal means $\theta = \mu_1 - \mu_2$, when the common variance σ^2 is unknown (Hogg & Craig (1995), p. 416, Example 2); c) testing the equality two normal variances $\theta = \sigma_1^2/\sigma_2^2$ when both means are unknown.(Hogg & Craig (1995), p. 421, Example 3). If a random interval $I(Y \mid \alpha)$ is desired, it is much easier to construct, by whatever means, a pivotal quantity $Q(y, \theta)$ and using (27) to derive a fiducial interval $I(y \mid \alpha)$ and then converting it into the interval $I(Y \mid \alpha)$.

Let us have five examples below: As usual we denote $\overline{X}$ and S^2 the sample mean and the sample variance and $Y_1 < ... < Y_n$ the order statistics of the random sample $X_1, ..., X_n$, and S_1^2 and S_2^2 the sample variances of two independent random samples $X_1, ..., X_{n+1}$ and $Y_1, ..., Y_{m+1}$.

The Normal (θ, σ^2) distribution with σ^2 unknown: $Y = \overline{X}$ and $Q(Y, \theta) = \frac{\sqrt{n}(Y-\theta)}{S}$, so that H is the t-distribution with $n-1$ d.f. $\rho(\alpha) = t_{n-1;\alpha}$ the α-percentile of the t-distribution with $n-1$ d.f. The fiducial interval for θ is

$$I(y \mid \alpha) = \left[y - t_{n-1;\alpha/2} \frac{s}{\sqrt{n}},\ y + t_{n-1;\alpha/2} \frac{s}{\sqrt{n}} \right].$$

(s denotes the realization of the sample standard deviation S.)

Normal (μ, θ) with μ unknown: $Y = S^2$ and $Q(Y,\theta) = \frac{(n-1)Y}{\theta}$, so that H is the χ^2-distribution with $(n-1)$ d.f. and $\rho(\alpha) = \chi^2_{n-1;a}$, the α-percentile of the $\chi^2(n-1)$ distribution. The fiducial interval for θ is

$$I(y \mid \alpha) = \left[\frac{(n-1)y}{\chi^2_{n-1;1-\alpha/2}},\ \frac{(n-1)y}{\chi^2_{n-1;\alpha/2}} \right].$$

The exponential distribution $E(\mu,\sigma)$ is defined as

$$f(x \mid \mu,\sigma) = \frac{1}{\sigma} e^{-(x-\mu)/\sigma}, \qquad \text{for } x > \mu > -\infty \quad \text{and} \quad \sigma > 0.$$

Exponential $E(\theta,\sigma)$ with σ unknown: $Y = Y_1$ and $Q(Y,\theta) = \frac{n(Y-\theta)}{\widehat{\sigma}}$, where $\widehat{\sigma} = \sum_{i=2}^n (Y_i - Y_1)/(n-1) = \sum_{i=1}^n (X_i - Y_1)/(n-1)$ and $\frac{2(n-1)\widehat{\sigma}}{\sigma}$ has the $\chi^2(2(n-1))$, so that $Q(Y,\theta) = \frac{2n(Y-\theta)/2}{2(n-1)\widehat{\sigma}/2(n-1)}$, has the F-distribution with $(2, 2(n-1))$ d.f. and $\rho(\alpha) = F_{2,2(n-1);\alpha}$, the α-percentile of the $F_{2,2(n-1)}$ distribution. The fiducial interval for θ is

$$I(y \mid \alpha) = \left[y - \frac{\widehat{\sigma}}{n} F_{2,2(n-1);\alpha/2},\ y - \frac{\widehat{\sigma}}{n} F_{2,2(n-1);1-\alpha/2} \right].$$

Exponential $E(\mu,\theta)$ with μ unknown: $Y = \sum_{i=1}^n (X_i - Y_1)$ and $Q(Y,\theta) = \frac{2Y}{\theta}$ so that H is $\chi^2(2(n-1))$ distribution and $\rho(\alpha) = \chi_{2(n-1);\alpha}$. The fiducial interval for θ is

$$I(y \mid \alpha) = \left[\frac{2y}{\chi^2_{2(n-1);1-\alpha/2}},\ \frac{2y}{\chi^2_{2(n-1);\alpha/2}} \right].$$

And finally, (see Example 12)

Two independent normal $N(\mu_1, \sigma_1^2)$ **and** $N(\mu_2, \sigma_2^2)$, **with** μ_1 **and** μ_2 **unknown and** $\theta = \sigma_1^2/\sigma_2^2$. $Y = S_1^2/S_2^2$ and $Q(Y, \theta) = \frac{Y}{\theta}$ so that H is the standard F-distribution with (n, m) d.f. and $\rho(\alpha) = F_{n,m;\alpha}$. The fiducial interval for θ is

$$I(y \mid \theta) = \left[\frac{y}{F_{n,m;1-\alpha/2}}, \frac{y}{F_{n,m;\alpha/2}}\right].$$

REFERENCES

[1] Barnard, G. (1963). Some logical aspects of the fiducial argument. *Journal of the Royal Statistical Society, Series* **B 25**, 111-114.

[2] Brown, L. D., Cai, T. T. and DasGupta, A. (2001). Interval estimation for a binomial proportion (with discussion). *Statistical Science*, **16**, 101-133.

[3] Brown, L. D., Cai, T. T. and DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Annals of Statistics,* **30**, 160-201.

[4] Buehler, R.J. (1983). Fiducial inference. *Encyclopedia of Statistical Sciences.* New York, Wiley, Volume III, 76-81.

[5] De Groot, M.H. (1986). *Probability and Statistics.* 2nd Edition. Addison-Wesley, Reading, Massachusetts.

[6] Dempster, A.P. (1963). On direct probabilities. *Journal of the Royal Statistical Society. Ser.* **B 25**, 100-110.

[7] Dempster, A.P. (1964). On the difficulties inherent in Fisher's fiducial argument. *Journal of the American Statistical Association,* ***59,*** *56-66.*

[8] Dempster, A.P. (1966). New methods for reasoning towards posterior distributions based on sample data. *Annals of Mathematical Statistics,* **37**, 355-374.

[9] Dempster, A.P. (1968). A generalization of Bayesian inference (with discussion). *Journal of the Royal Statistical Society. Series* **B 30**, 205-247.

[10] Edwards, A.W.F. (1976). Fiducial probability. *The Statistician,* **25,** 15-35**.**

[11] Edwards, A.W.F. (1983). Fiducial distributions. *Encyclopedia of Statistical Sciences.* New York. Wiley, Volume III, 70-76.

[12] Fisher, R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika,***10**, 507-521.

[13] Fisher, R.A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society,* **26**, 528-535.

[14] Fisher, R.A. (1939). "Student." *Annals of Eugenics* **9,** 1-9.

[15] Fisher, R.A. (1941). The asymptotic approach to Behrens integral with further tables for the d test of significance. *Annals of Eugenics,* **11**, 141-172.

[16] Fisher, R.A. (1945).The logical inversion of the notion of the random variable. *Sankhyā* **7**, 129-132.

[17] Fisher, R.A. (1956, also 1973). *Statistical Methods and Scientific Inference, 3-rd edition.* Oliver & Boyd, Edinburgh.

[18] Fraser, R.A. (1961). Fiducial method and invariance. *Biometrika,* **48**, 261-280.

[19] Fraser, D.A.S. (1968). *The Structure of Inference.* Wiley, New York.

[20] Hacking, I. (1965). *Logic of Statistical Inference.* Cambridge University Press.

[21] Hogg, R.V. and Craig, A.T. (1995). *Introduction to Mathematical Statistics, 5-th edition.* Prentice Hall, Englewood Cliffs, N.J.

[22] Jeffreys, H. (1932). On the theory of errors and least squares. *Proceedings of the Royal Society of London. Ser.* **A 138**, 38-45.

[23] Jeffreys, H. (1961). *Theory of Probability,* 3rd edition. Oxford University Press.

[24] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distribution, Volume 2, Second Edition. Wiley, New York.*

[25] Kempthorne, O. (1983). A review of "R. A. Fisher: an appreciation" by Fienberg and Hinkley. *Journal of the American Statistical Association.* **78**, 482-490.

[26] Kyburg, H.E. (1963). Logical and fiducial probability. *Bulletin of Institute of International Statistics.* **40**, 884-901.

[27] Kyburg, H.E. (1974). *Logical Foundations of Statistical Inference.* Reidel, Boston.

[28] Lehmann, E.L. (1986). *Testing Statistical Hypotheses.* Wiley, New York.

[29] Lindley, D.V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society, Series* **B 20**, 102-7.

[30] Neyman, J. (1934). On the two different aspects of the representative method (with discussion). *Journal of the Royal Statistical Society,* ***97***, 558-625.

[31] Neyman, J. (1961). Silver Jubilee of my dispute with Fisher. *Journal of Operation Research Society, Japan.* **3**, 145-154.

[32] Pedersen, J.G. (1978). Fiducial inference. *International Statistical Review.* **46,** 147-170.

[33] Sackrowitz, H. and Samuel-Cahn, E. (1999). P values as random variables — expected p values. *The American Statistician.* **53,** 326-331.

[34] Savage, L.J. (1964). Discussion. *Bulletin of International Statistical Institutes.* **40**, 925-927.

[35] Savage, L.J. (1976). On rereading R.A. Fisher. *The Annals of Statistics.* **4,** 441-500.

[36] Schervish, M.J. (1996). P values: What they are and what they are not. *The American Statistician.* **50,** 203-206.

[37] Schwertman, N.C. (2000). Private communication.

[38] Seidenfeld, T. (1979). *Philosophical Problems of Statistical Inference.* Reidel, Dordrecht, Holland.

[39] Stone, M. (1983). Fiducial probability. *Encyclopedia of Statistical Sciences.* New York. Wiley, Volume III, 81-86.

[40] Student. (1908). The probable error of a mean. *Biometrika,* **6,** 1-25.

[41] Wang, Y.H. (2000). Fiducial intervals: what are they? *The American Statistician.* **54**, 105-111.

[42] Wang, Y.H. (2001). Some controversies surrounding the statistical theory. *Proceedings of 14-th Hellenic Conference in Statistics,* **B**, 675-684.

[43] Wang, Y.H. and Chang, H.F. (2002). Optimum confidence intervals for the Bernoulli success probability. *Journal of Applied Statistical Science.* **11**, 235-244.

[44] Zabell, S. (1992). R.A. Fisher and the fiducial argument. *Statistical Science* **7,** 369-387.