Fiducial Inference and Belief Functions

Russell Almond[1]

University of Washington
Department of Statistics

SCREED
April, 1991

This paper was read at an informal seminar at the University of Washington on the work of Fisher in honor of the new volume of the reprinted correspondence of Fisher. The work was facilitated in part by National Science Foundation Grant number DMS-89-02211.

[1] This paper was written while the author was visiting at the University of Washington. Russell Almond is currently at Educational Testing Service, Princeton, NJ.

Fiducial Inference and Belief Functions

Russell Almond

ABSTRACT

One aspect of Fisher's work which has puzzled a great many statisticians is the idea of *Fiducial Inference*. Using "pivotal" variables, Fisher moves from logical statements about restrictions of parameter spaces to probabilistic statements about parameters. The method works, but with a lot of caveats: the pivotal variables must be sufficient statistics, must be continuous, &c.

Dempster's explorations of the fiducial method as applied to the binomial distribution in 1966 through 1968 lead to the formulation of an upper and lower probability model later extended and named "belief functions" by Shafer. Dempster revisits the ideas in a 1989 paper and provides a philosophical synthesis of the ideas of Bayesian, Belief Function and Fiducial Inference. In particular, the theory of belief functions contains representations for logical statements about membership of parameters in sets and probabilistic statments about parameters. It moves between the two representations through subjective statments about the state of knowledge in the form of upper and lower probabilities.

This paper explores the belief function and fiducial arguments using a "state-of-knowledge" argument tie the two methods together. In this light the paper reviews fiducial arguments for the binomial and normal distribution. It also compares the fiducial and belief function arguments for the Poisson process and arrives at an interesting paradox.

Key Concepts: Fiducial Inference, Belief Functions, Normal Inference, Binomial distribution, Poisson process, Pivotal variables

1. Introduction

> *I know only one case in mathematics of a doctrine which has been accepted and developed by the most eminent men of their time, and is now perhaps accepted by men now living, which at the same time has appeared to a succession of sound writes to be fundamentally false and devoid of foundation. Yet that is quite exactly the position in respect of inverse probability.*

Thus Fisher opens his 1930 paper "Inverse Probability" (Fisher[1930]). Fisher is talking about Bayesian statistics here, but it could be argued that Fisher's statement applies even more forcibly to his theory of fiducial probability. As evidence, examine the recently published volume of Fisher's correspondence (Bennet, ed.[1990]); the first 244 pages are devoted to letters between eminent men and Fisher, trying to thrash out the exact nature of the fiducial argument. Names like Barnard, Finney, Fréchet, Jeffreys, Neyman, Savage, Tukey and Yates, eminent men all, yet all had difficulty understanding the fiducial argument and wrote to Fisher for clarification.

Another person who has struggled to understand Fisher's ideas of fiducial probabilities is Arthur Dempster. A particular series of papers from 1963 through 1968 demonstrate the progress of his ideas. His earliest papers start with a discussion of the fiducial argument. A "pivotal" paper, Dempster[1966], applies the fiducial argument to the binomial and multinomial models and arrives at an upper and lower bounds on probability distributions. The later papers then concentrate on exploring these upper and lower distributions, and those papers more or less divorce the theory from its fiducial origins. One of the last papers, entitled *A generalization of Bayesian Inference*, Dempster[1968a], suggests a subjectivist origin of the theory rather than its fiducial origin. Glenn Shafer later (Shafer[1976]) picks up the threads of this theory names it "belief functions," and expands the calculus and defines much of the notation. Shafer[1990] Contains a review of recent developments.

Dempster[1989] later comes back to the fiducial origins of belief functions, and explains how Fisher's most basic fiducial argument, that in the normal case, can be expressed in the belief function terminology. He also explains much of Fisher's attitude can be explained by a synthesis of subjectivism and objectivism which he calls "objectivist–subjectivist":

> *One key is that both Bayes and Fisher were "objectivist" "subjectivists". The latter term refers to the universal (pre-frequentist) semantics that perceives probability as a formal quantitative*

representation of an individual's assessment of uncertainty given current evidence, while the former term conveys the desideratum only knowledge and principles external to any individual scientist, and therefore shared or at least sharable among scientists, be used. For me, belief function models are formal vehicles for reasoning with probabilities, motivated by a similar commitment to objective subjectivism.

If I understand Dempster correctly [2], "Objective-subjectivism" is subjective in that it allows a state of knowledge about an unknown quantity to be expressed as a "subjective" probability distribution. It is objective in that it only allows objective sources for that state of knowledge. "Objective sources" (at least in what I am currently taking for the Fisherian formulation of the problem) here means data, or occasionally a sufficiently compelling *a priori* argument such as a sampling from an urn or the base rate of a genotype in a population of individuals. This may be what Fisher means by the term "fiducial"; a "fiducial probability" is a probability—a subjective state of information about an unknown quantity—which is fiducial—based on empirical evidence or sufficiently well established reasoning to be considered "objective" or "true."

Fisher describes his attitude toward probability and states of knowledge in a series of correspondence with D.J. Finney. In his March 15, 1955 letter to D.J. Finney (Bennet, ed.[1990], p. 96) he writes:

I have recently been thinking a little about the semantics of this word [It is unclear to me from the previous paragraph whether he means "fiducial" or "probability" –R]. *What seems to be implied whenever it is used is a distinction of three levels of knowledge, (a) in which nothing whatever is known about some supposed knowledge, (c) in which the exact value is known, and (b) in which a statement can be made in terms of the concept of mathematical probability, in which the case of a stochastic, or random variable, about which a complete set of probability statements can be made, is typical.*

The problem seems to lie in that Statements (a) and (c) are statements of logic, and Statements (b) and (c) are statements of probability, Statement (c) represents the overlap, or a statement of probability 1. The perceived problem with fiducial reasoning is that slides smoothly from statements of type (a) to statements of type (b). This cannot be done within the subjectivist or objectivist axiomatic frameworks for probability. All three types of statements can be represented by belief functions, and that is one of the greatest strengths of belief functions.

The key to understanding both fiducial and belief function inference, is the iterpretation of probability distributions (or belief functions) as staes of information. It is the combination of a model and an observation which allows the state of information to pass from statements of type (a) to statements of type (b). This paper will follow this idea with several examples.

2. The Fiducial Distribution for the Normal Distribution

To illustrate the *fiducial* argument, let us apply it to the case of a series of normal observation from a population with a known variance and an unknown mean. (Fisher preferred to use the case with a unknown variance in his illustration, it being of greater practical value). In this case we can express the observable random variable as a function of the parameters and a unknown random variable with a known (reference) distribution. In particular, we use the model equation:

$$\bar{X} = \mu + \frac{\sigma}{\sqrt{n}} z \tag{1}$$

where z has a standard normal distribution. Because z has a known distribution, it is possible to find its quantiles, and build confidence intervals, as it is possible to express the quantiles of the distribution of the statistic as functions of the unknown parameter(s) and the quantiles of the reference distribution.

The next step in the fiducial argument is to *pivot* on the random value with the known distribution by solving the model equation (1) for the unknown parameter. In the normal example this yields:

$$\bar{X} - \frac{\sigma}{\sqrt{n}} z = \mu \tag{2}$$

[2] I have just succeeded in generating a hasty series of communications with Art on the subject. Imagine what Fisher's correspondence might have been like had he access to email!

If we replace the equalities in Equations 1 and 2 with inequalities we can make exact probability statements about those inequalities. The variable with a known reference distribution, z, is referred to as a *pivotal variable*. After observing the value of $\bar{X}$, it is possible to describe our state of knowledge about the unknown parameter as a probability distribution: a normal distribution with mean $\bar{X}$ and variance σ^2/n.

I think it is important to make clear the distinction between Equations 1 and 2. The most important difference is that Equation 1 describes the state of knowledge *a priori* while Equation 2 describes the state of knowledge *a posteriori*. Thus in Equation 1 the potential observable value $\bar{X}$ is expressed as a deterministic function of a random variable with known distribution, z, a (completely) unknown quantity, μ and two known quantities σ^2 and n. It describes the "likelihood" of a given observation as a function of the unknown quantity. If we condition on a value of the parameter, it induces a (conditional) probability distribution for the observable quantity.

Once the observation is known, the situation changes to that described in Equation 2. Here the unknown quantity of interest, μ, is expressed as the function of a unknown quantity with known distribution, z, and three known quantities, $\bar{X}$, σ^2 and n. Thus knowledge about μ is equivalent to knowledge about the pivotal variable, and μ is also a random variable. I think it is key to the understanding of Fisher's ideas that this change in the state of knowledge about μ from Equation 1 to 2 comes about not through some action on the part of the pivotal variable, but through the act of observation.

As simple and compelling as the fiducial argument is when applied to this simple case, it soon becomes lost in complexity. The list of caveats and situations in which it can be applied seems to be endless: The observable quantities must be continuous, the statistics involved must be complete and sufficient, the function must be monotonic. If the problem is multivariate the number of caveats explodes including such complex conditions as the Jacobian and all sub-Jacobians must be positive. Fisher writes to Barnard in March of 1962, (Bennet, ed.[1990], pp 44-45).

> *A pivotal quantity is a function of parameters and statistics, the distribution of which is independent of all parameters. To be of any use in deducing probability statements about parameters, let me add*
>
> *(a) it involves only one parameter,*
>
> *(b) the statistics involved are jointly exhaustive for that parameter,*
>
> *(c) it varies monotonically with the parameter.*
>
> [Omitting Fisher's explanation]
>
> *For sets of pivotals then I add*
>
> *(d) the joint distribution is independent of parameters (of as high or higher stratum)*
>
> *(e) all are monotonic, uniformly for variations of parameters of as high or higher stratum.*

This extensive list of caveats has caused many statisticians to shy away from fiducial inference. Almost more damning is the fact that Fisher seems to advocate different fiducial inferences in different situations. The justification for this lies in the fact that Fisher was an applied statistician. For him, each problem deserved a carefully constructed framework which accounted for details of the issues in question; anything less was dishonest. Dempster (1989) claims this as a strength rather than a weakness of fiducial inference, and I concur. For Fisher, Dempster, and myself, applying an old method to a new problem requires first verifying that the old method is appropriate in the new context.

3. Dempster's Fiducial Distribution for the Bernoulli Process

Dempster[1966] applies the pivotal argument to the Bernoulli Process, despite Fisher's injunction against using the fiducial argument with discrete data. The result he gets is nevertheless simple and compelling.

Let $X_1, \ldots, X_n$ be a series of observations from a Bernoulli process such that $\mathrm{P}\{X_i = 1\} = p$ for all i. We now introduce a series of pivotal variables a_i, one for each observation X_i, each having a uniform distribution over the interval $[0, 1]$. We link the pivotal variables with the equation:

$$X_i = \begin{cases} 1 & a_i \leq p \\ 0 & a_i > p \end{cases} \tag{3}$$

This relationship among observables, pivotal variables, and the unknown probability can be expressed by the graphical model shown in Figure 1. Marginalizing out the pivotal variables reduces the problem to the standard likelihood for a Bernoulli process.

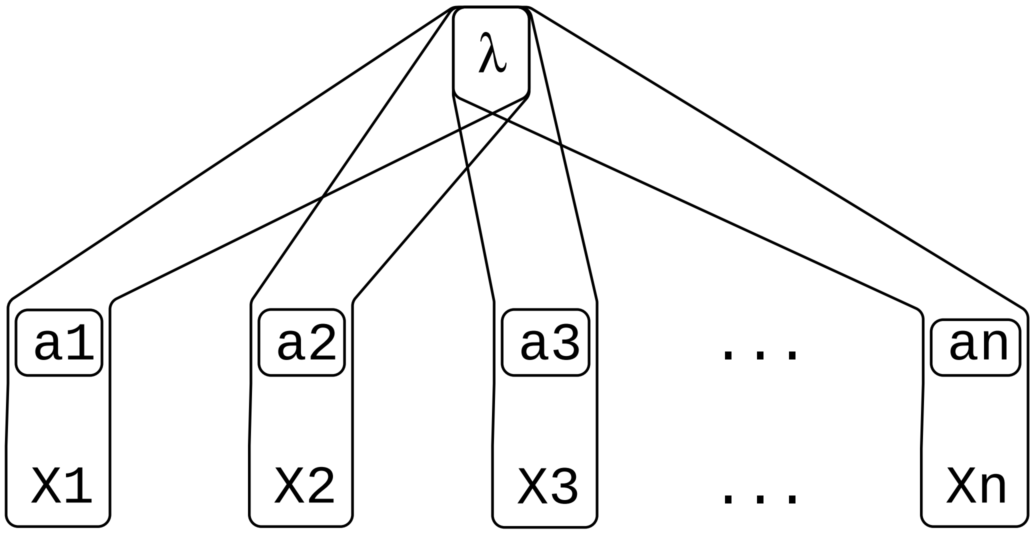

Figure 1. Graphical model linking parameters and variables.

By the assumptions of a Bernoulli process, the order of the observations is unimportant. Therefore, the statistic $X = \sum X_i$ is complete and sufficient. Once we have observed $X = k$, we then have more information about the relationships of the pivotal variables to the parameter p. In particular, exactly k of them must be less than p. Let $\underline{a} = a_{(k)}$ equal the kth order statistic of the pivotal variables and let $\overline{a} = a_{(k+1)}$ equal the $(k+1)$st order statistic. To handle the special cases, $k = 0$ and $k = n$ we define $a_{(0)} \equiv 0$ and $a_{(n+1)} \equiv 1$. Then we know that:

$$\underline{a} \leq p < \overline{a}. \tag{4}$$

As the distribution of $a_1, \ldots, a_n$ is known, the distribution of the order statistics is known as well. In fact the joint p.d.f or *mass function* for $\underline{a}$ and $\overline{a}$ is:

$$\begin{aligned} m(\underline{a}, \overline{a}) &= \frac{n!}{(k-1)!\,(n-k-1)!} \underline{a}^{k-1}(1-\overline{a})^{n-k-1} \\ &= k(n-k)\binom{n}{k}\underline{a}^{k-1}(1-\overline{a})^{n-k-1}, && 0 < k < n; \\ m(0, \overline{a}) &= n(1-\overline{a})^{n-1}, && k = 0; \\ m(\underline{a}, 1) &= n\underline{a}^{n-1}, && k = n; \end{aligned} \tag{5}$$

We can think of Equations 4 and 5 as defining a random interval containing the unknown parameter p. This, in turn, can be used to derive a upper and lower bounds on the probability distribution for p. In particular, for any set B we can get the lower bound on the probability that $p \in B$, or *belief* by:

$$\mathrm{BEL}(B) = \iint_{\substack{[a,b] \subseteq B \\ a \leq b}} m(a, b)\, da\, db\,, \tag{6}$$

and the upper bound on the probability that $p \in B$, or *plausibility* by:

$$\mathrm{PL}(B) = \iint_{[a,b] \cap B \neq \emptyset} m(a, b)\, da\, db\,. \tag{7}$$

.

Dempster[1966] goes on to derive exact formulas for Equations 6 and 7 when the set B under consideration is an interval, but they are not as interesting as the upper and lower expected values of p:

$$E_*p = \iint_{[\underline{a},\overline{a}]\subseteq[0,1]} \underline{a}\, m(\underline{a},\overline{a}) d\underline{a}\, d\overline{a} \;, \tag{8}$$

$$E^*p = \iint_{[\underline{a},\overline{a}]\subseteq[0,1]} \overline{\alpha}\, m(\underline{a},\overline{a}) d\underline{a}\, d\overline{a} \;. \tag{9}$$

In the particular case of the Bernoulli process, this becomes:

$$\frac{k}{n+1} \leq E[p] \leq \frac{k+1}{n+1} \;. \tag{10}$$

Equation 10 tells us a lot about the belief function model. As $n \to \infty$, the upper and lower bounds converge, and the belief function converges on a probability distribution. Even more interesting is that the interval in (10) contains $\frac{k}{n}$, Fisher's maximum likelihood estimate; $\frac{k+1/2}{n+1}$, the Bayes estimate obtained using Jeffreys prior; and $\frac{k+1}{n+2}$, the Bayes estimate using the uniform prior advocated by Bayes and Laplace!

The state of information described by the belief function approach is less precise than the state of information produced by any of the Bayesian models with non-informative priors. This is because the Bayesian models have extra information: the prior. Note that if we did have an informative prior, the belief function methodology would incorporate that extra (independent) information into the final result, producing the same state of information as produced by the Bayesian analysis: the posterior probability distribution.

Fisher does address the fiducial inference problem for the binomial distribution in the third edition of *Statistical Methods and Scientific Inference* (Fisher[1973]). He gives approximate distributions in the case where n is large and hence X is approximately normally distributed. He gives three different approximation methods, one based on his maximum likelihood estimate, one based on Jeffreys' Bayesian method, and one based on the method of Bayes and Laplace. He notices that all three methods converge as n gets very large, but considers his superior. He claims the "exact" solution can be found in his 1957 paper, "The Underworlds of Probability." We will return to that paper later in the discussion.

4. The terminology of belief functions

Not all upper and lower probability schemes result in belief functions. Belief functions assume some additional regularity conditions which result in (or are derived from, depending on your point of view) the random set interpretation. This mass function, which defines a probability distribution over sets of outcomes rather than outcomes, is fundamental to the understanding of belief functions.

Let X be an unknown quantity and let Θ_X be the set of possible values for X, which is called, for reasons which will become obvious in the sequel, a *frame of discernment* or *frame.* For the simplicity of exposition, we will assume that Θ_X is finite. (Wasserman[1988,1990] rigorously develops the measure-theoretic details of the continuous case). We next define a *mass function* $m : 2^{\Theta_X} \to [0,1]$ which is a probability function over the power set of the frame. The empty set is defined to have zero mass. The support of $m(\cdot)$ is called the set of *focal elements* of the belief function.

It is here worth mentioning some special cases. (I) In the case were there is only one focal element, A, then the belief function amounts to a logical assertion that $X \in A$. There are two further specializations of this case: (a) If the only focal element is the entire frame, then the belief function amounts to the trivial proposition that $X \in \Theta_X$ and we call the belief function *vacuous*. (c) If the only focal element is a single outcome, it amounts to the logical proposition that the specified outcome is the true value of X, we call this belief function *deterministic*. (II) If all of the focal elements each contain a single outcome, then the mass function m is essentially a probability mass function over those outcomes. In this case we call the belief function *Bayesian*. We note that deterministic belief functions are also Bayesian. Returning to Fisher's letter to Finney quoted earlier, Fisher's statement (a) is represented by a vacuous belief function, (b) is represented by a Bayesian belief function and (c) is represented by a deterministic belief function.

Dempster[1967] introduces a combination rule for two independent belief functions:

$$m_1 \oplus m_2(C) = \frac{\sum_{\substack{A,B\subseteq\Theta \\ A\cap B=C}} m_1(A)\cdot m_2(B)}{1-\sum_{\substack{A,B\subseteq\Theta \\ A\cap B=\emptyset}} m_1(A)\cdot m_2(B)} \qquad \forall C\subseteq\Theta \tag{11}$$

In two special cases Dempster's rule reduces to a more familiar form. In the case that both belief functions are logical, it becomes set intersection, providing the logical conjunction of the two propositions. In the case where both belief functions are Bayesian (and one can be interpreted as a conditional distribution given the other) the rule reduces to Bayes Theorem. The exact nature of the *independence* required for the rule to work has never been completely defined. Dempster rather circularly defines *independence* as cases where the combination rule works. Both he and Shafer supply plenty of anecdotal evidence on which to base reasoning about independence.

Shafer[1976] describes rather elegant operator (or pair of operators) for changing the frame of discernment of a belief function. As these are easiest to understand in a multivariate framework, let us consider three variables, X, Y and Z. Corresponding to each variable is a frame of discernment Θ_X, Θ_Y and Θ_Z. Corresponding to each pair of variables is a larger frame of discernment which is the cross product of the smaller frames, thus: (X,Y) corresponds to $\Theta_X\times\Theta_Y$, (Y,Z) to $\Theta_Y\times\Theta_Z$ and (X,Z) to $\Theta_X\times\Theta_Z$. All three variables taken jointly have the frame $\Theta_X\times\Theta_Y\times\Theta_Z$. As it is clear than any collection of variables corresponds to a frame, I shall be deliberately careless and use the term *frame* for both a collection of variables, and the set of their possible outcomes.

Shafer argues that it should be possible to minimally extend a belief function over X to the frame (X,Y) without adding any additional information about Y. He does this by taking every focal element of the belief function over X and extending it by taking its cross product with the frame of discernment for Y, Θ_Y. Thus if A was a focal element of the old belief function, $A\times\Theta_Y$ would be a focal element of the new belief function.

Similarly, it is possible to marginalize a belief function over a multivariate frame to a smaller frame. If A is a focal element in the belief function over the frame (X,Y), and the marginal beliefs about X are required, then A is projected onto the X margin. As a series of such projections could result in more than one focal element in the larger frame being projected onto the same focal element in the smaller frame, the mass of all such multiple projections are summed. Mariginalization throws away information about the eliminated variables. A generalized projection operation is formed by combining minimal extension and marginalization. Thus to project a belief function from the frame (X,Y) to the frame (Y,Z), you would first extend it to the frame (X,Y,Z) and then marginalize to the frame (Y,Z).

Marginalization is a familiar operation from the probability theory, but minimal extension is not. Lauritzen and Speigelhalter[1988] implicitly define a (uniform) extension operator for "potential" representations of probability, although there method requires that fully specified probability models be built. For an explanation of the Lauritzen and Spiegelhalter model in a belief function framework, see Shenoy and Shafer[1988] or Almond[1989].

5. The Normal Problem Revisited

As a further illustration of the theory of belief functions, we examine the fiducial argument for the normal sample in the belief function framework (Dempster[1989]). First we define the variables of the problem; there are three: μ—the unknown "true" mean, $\bar{X}$—the observable sample mean, and z—the unobservable "pivotal" variable with known distribution. All have outcome spaces corresponding to the entire real line, and thus the joint outcome space is $\Re^3$.

We have *a priori* two pieces of information: the model (or likelihood) equation, Equation 1 (reproduced below) and the known distribution for z.

$$\bar{X} = \mu + \frac{\sigma}{\sqrt{n}}z \tag{12}$$

Equation 1 (or 12) defines a logical restriction in $\Re^3$ that constrains $(\bar{X}, \mu, z)$ to lie in a plane. The known, standard normal distribution for z produces a Bayesian belief function over z. We minimally extend that belief function to the frame $(\bar{X}, \mu, z)$ where it becomes a series of random planes in $\Re^3$, each defined by a possible value of z. We combine it with the belief function defined by Equation 1. The result is a series of random lines in $\Re^3$, representing the intersection of the random lines with the fixed plane.

We now receive an addition piece of information, namely the value of $\bar{X}$. This defines a deterministic belief function over $\bar{X}$ which is minimally extended to a plane in $\Re^3$. Intersecting this "data plane" with the random line which represents the sum of our *a priori* beliefs produces the *a posteriori* belief function over the joint space which is a random point. Projecting (marginalizing) that point onto the frame μ produces the normal fiducial distribution for μ.

Why does the pivotal method produce upper and lower bounds for the binomial parameter and an exact distribution for the normal parameter? The answer lies in the amount of information provided by the equal sign in the piovotal equation (Equation 1/12). The equal sign provides two pices of information: (1) the likelihood of the observed values of $\bar{X}$ for each value of μ and (2) the fact taht the likelihood contains complete information (hence the equality). It is this second piece of information that allows the fiducial argument to turn the likelihood into something which behaves like a posterior distribution. This ideas is illustrated even more clearly in the example of the following section.

6. The fiducial distribution for the Poisson process

Following Dempster[1966], Almond[1989] develops a similar fiducial distribution for the Poisson process. Assume that X is the number of event generated by a Poisson process with unknown rate λ within a time period of known length t. We introduce a series of pivotal random variables v_i which are the independent exponentials. The variables $w_i = \frac{1}{\lambda} v_i$—the unit exponentials scaled by the failure rate—are the waiting times between the $(i-1)$th and the ith event of the Poisson process. From the data we know that λ must lie in the set Σ as defined below:

$$\begin{aligned} \Sigma &= \{\lambda : \sum_{i=1}^{X} w_i \le t < \sum_{i=1}^{X+1} w_i\} \\ &= \{\lambda : \sum_{i=1}^{X} v_i \le \lambda t < \sum_{i=1}^{X+1} v_i\} \\ &= \{\lambda : \frac{1}{t}\sum_{i=1}^{X} v_i \le \lambda < \frac{1}{t}\sum_{i=1}^{X+1} v_i\} \ . \end{aligned} \tag{13}$$

Let $V_j = \sum_{i=1}^{j} v_i$. Note that $\frac{1}{t} V_j$ has a $\Gamma(j, t)$ distribution for $j > 0$ and is identically equal to zero otherwise. Also, the random variable $\frac{1}{t}(V_k - V_j)$ is independent of $\frac{1}{t} V_j$ and has a $\Gamma(k-j, t)$ distribution for $k > j$. Let $\underline{a} = \frac{1}{t} V_X$ and $\overline{a} = \frac{1}{t} V_{X+1}$ be two dependent random variables. Then $[\underline{a}, \overline{a}]$ Forms a random interval with mass function:

$$m\left([\underline{a}, \overline{a}]\right) = \frac{t^{X+1}}{(X-1)!} \underline{a}^{X-1} e^{-\overline{a}t} \ , \tag{14}$$

This random interval is in fact the random set Σ containing λ and therefore it forms the (continuous) mass function for a belief function expressing the information contained in the data about λ. The upper and lower expectations for λ are $\frac{X}{t}$ and $\frac{X+1}{t}$ respectively. Note that the lower expected value is the maximum likelihood estimate for the Poisson process, and the upper expected value is the Bayes estimate obtained via the uniform prior. The Bayesian analysis using Jeffrey's rule as a prior yields $\frac{X+1/2}{s}$ as the expected failure rate which lies between the two extremes.

Fisher does talk about the Poisson process in his *Statistical Methods* book. He assumes that the arrival times are observed, as these form the complete statistic. If these arrival times are observed, an exact fiducial distribution, instead of an upper and lower bound can be found. Fisher would no doubt accuse me (or whoever collected only the number of events) of designing an inefficient estimation scheme. Fisher's distribution leads to a curious paradox, and so is worth examining more closely.

Assume that we have observed the process continuously from time 0 to time t and that the last event occured at time t_1. Let $t_2 = t - t_1$ be the time remaining since the last ovservation. The waiting time t_1 should follow a Gamma(X, λ) distribution; that is:

$$t_1 = \frac{1}{\lambda} V_X \tag{15}$$

Where V_X is a Gamma$(X, 1)$ random variable. This is essentially Equation 26 from Fisher[1973], page 57; the only difference being that Fisher uses a χ^2 with $2n$ degrees of freedom instead of the Gamma. Observing t_1 allows us to pivot the equations yielding a fiducial distrabution for λ; t_1 follows a Gamma(X, t_1) distribution. This can be represented by a Bayesian belief function, call its mass function—an ordinary gamma density function—m_1

Now look at the remaining time, t_2. Observing no events a the time period of length t_2 yeilds the following belief function:

$$m_2([0, \overline{a}]) = te^{-\overline{a}t} \tag{16}$$

Note that this belief function is define only over intervals containing 0; but is otherwise very similar to the random interval distribution shown in Equation 14. By the usual properties of Poisson processes, the belief functions m_1 and m_2 are independent and can be combined by Dempster's product-intersection rule.

Because m_1 and m_2 are continuous distribution, the sum in the numerator of Equation 11 becomes an integral. As the belief function m_1 is a Bayesian belief function, all of its focal elements are single element sets; therefore the mass function $m_1 \oplus m_2$ will also have only single element sets as focal elements. Let $[a, a]$ be a focal element of m_1; its intersection will be non-empty with any focal element $[0, \overline{a}]$ of m_2 such that $a < \overline{a}$: the intersection of the two intervals is $[a, a]$. The denominator of Dempster's rule is mearly a normalization constant, therefore:

$$m_1 \oplus m_2([a, a]) \propto \int_a^\infty a^{X-1} e^{-at_1} e^{-\overline{a}t_2}\, d\overline{a} \propto a^{X-1} e^{-a(t_1+t_2} \tag{17}$$

Normalizing the expression on the right hand side yields a Gamma(X, t) distribution. Therefore, using Fisher's method the total information about λ is that it follows a Gamma(X, t) distribution.

This result is somewhat surprizing. First it is the same distribution, for all values of $0 < t_1/t \leq 1$. This, however, is mearly an artifact of the Poisson process assumption of independent increments. By that assumption, all time periods are exchangeable, and it should make no difference in our state of knowledge about the parameter if some of the time t_2 had been observed before the last failure and hence was in t_1. The fact that the same fiducial distribution holds even for small values of t_1/t is an artifact of the assumptions of the Poisson process, and it seems unrealistic because often the assumptions of the Poisson process are unrealistic for many situations.

What is somewhat more puzzling is the difference in the results of Equations 14 and 17. Why should one set of information result in upper and lower bounds, while one additional piece of information—no matter what it is—result in an exact distribution. In this case knowing the value of t_1, no matter what it is, changes the state of information about the process rate from a random interval distribution to a more familiar probability distribution. This is essentially the same paradox that Diaconis [1978] discusses in his review of a Shafer's 1976 book (see also, Diaconis and Zabell[1978]).

The key to understanding this dilema, lies in recognizing that the equal sign in Equation 15 is once again doing double duty. It is not only telling us that the distribution of the waiting time is one particular gamma, but also it carries with it the assumption that this is complete information about the problem. The inequalities of Equation 13 carry incomplete information, and remain compatable with several "non-informative" assumptions of prior information. The question still remains, are the assumptions of Equation 15 too strong or are the assumptions of Equation 13 too weak?

7. "The Underworld of Probability"

In 1957, Fisher wrote a paper called the "Underworld of Probability" in which he describes uncertainty models of various ranks.

> *Since in statistical discussion it is common to speak of a* probability *as an unknown quantity, appropriate ranks should be assigned to such statments as follows:*
> Cenrtainty. *The Event will occur.*
> Uncertainty of Rank A. *The Event has a known probability* π
> Uncertainty of Rank B. *For all* P *a function* π_P *is known such that the probability that* π *is less than* π_P *is exactly* P
> Uncertainty of Rank C. π_P *is given as a random variable, such that a known function of* P *and* P'*, namely* $\pi_{PP'}$ *shall exceed* π_P *with frequency* P'*.*

Fisher identifies Uncertainty of Rank A with a distribution with known parameters and Uncertainty of Rank B with a distribution whose parameters are known up to a probability distribution—the standard Bayesian model. Good[1979] proposes a similar hierarchy of models using uncertainty about parameters in place of the uncertainty about order statistics used by Fisher. Uncertainties of Rank C are random quantities with a distribution whose unknown parameters have a distribution whose unknown hyperparameters have a completely specified distribution. Both Fisher and Good note that the hierarchy can be extended, *ad absurdum*.

How do belief functions fit into this hierarchy? The "Underworld" paper was not an influence on Dempster's development of belief functions, but has Fisher anticipated the theory of belief functions the way he anticipated so many other important ideas in statistics and genetics? Perhaps the answer lies in Fisher's exposition of the "exact" fiducial distribution for the binomial.

A belief function model has uncertainty of infinite rank. To see this, note that for any uncertainty model of finite rank, we can find the expected value of the probability of the event by integrating out the other variables. A belief function model, in contrast, has generally only an upper and lower expectation.

Imagine n observations from a binomial process where the probability of a success for each trial is π. Let

$$C_r = \binom{n}{r} \pi^r (1-\pi)^{n-r} \tag{14}$$

be the rth term in the binomial probability function. Let λ be an independent random variable with the uniform distribution over range $(0, 1)$. We then define the quantity

$$P_\pi = \sum_0^{a-1} C_r + \lambda C_a \tag{15}$$

which is a function of π and λ. Let π_0 be a value between zero and 1. Then $P_{\pi_0} = P\{\pi \in [0, \pi_0]\}$. P_{π_0} is a random variable with a known distribution. Fiducial confidence limits could be found by setting $P_{\pi_0} = \alpha$ and solving for π_0.

Dempster[1966] solves for the belief and plausibility of an interval $[\alpha, \beta]$. Specializing to the interval $[0, \pi_0]$ we find that

$$\mathrm{BEL}([0, \pi_0]) = \sum_{r=0}^{a-1} C_r \qquad \mathrm{PL}([0, \pi_0]) = \sum_{r=0}^{a-1} C_r + C_a \tag{16}$$

Thus the belief agrees with Fisher's Rank C model when $\lambda = 0$ and the plausibility agrees with Fisher's model when $\lambda = 1$. The difference between the Rank C model and the belief function model is that Fisher's model puts a distribution over a space of probability distributions compatible with the belief function model.

Even though Dempster's argument and resulting upper and lower bounds seem much more compelling than Fisher's uncertainty model of Rank c, they are just two possible models on an infinite collection of conceivable models for the Bernoulli process. Each would have circumstances under which it would be more applicable. Similarly, there are contexts in which my modl for the Poisson process will be more appropriate than Fisher's and *vise versa.* The fact that when the sample size is large, both models produce similar inferences (which are in turn similar to those produced from Bayesian models with common choices of non-informative priors) is comforting.

Nor, do I believe, did Fisher intend for his Rank C model to be definitive. In the abstract to his "Underworld" paper, Fisher writes: *The following paper is much less than an exploration of the varieties of logical uncertainty in normal experience. It takes one short step towards a recognition of these varieties and introduces the beginnings of a classification, more detailed than that of my book. The illustrations given will, I believe, help others to engage in a deeper penetration.*

It is in the same spirit of exploration that Dempster and others have developed the theory of belief functions.

Acknowledgments.

I am grateful to Art Dempster for taking the time to help me understand his thinking on the relationships between belief functions and fiducial inference. I am also grateful to Peter Guttorp and Elizabeth Thompson for organizing the seminar at which this paper was originally presented and encouraging me to delve into the fascinating history of the ideas at the foundations of statistical inference.

Bibliography

Rather than a standard list of reference, I have prepared a more extensive bibliography of Fisher's papers on fiducial inference and Dempster's papers on fiducial inference and belief functions. For other authors I only list references cited in the paper.

Fisher

There are two collections of Fisher's writing which are of note:

Bennet, J.H., ed. [1990] *Statistical Inference and Analysis: Selected Correspondence of R.A. Fisher.* Clarendon Press, Oxford, 1990.

Bennet, J.H. and Cornish, E.A., ed. *Collected Papers of R.A. Fisher*, University of Adelaide. Published in five volumes.

As my sources for Fisher's published papers has primarily been the latter volumes, I include in brackets after each Fisher citation, its location in the collected papers. The form of this citation is [CP #n, **v**, p] where n is the number of the paper in the collected papers, v is the volume in which it is found and p is the page number.

Fisher, R.A. [1930] "Inverse Probability," *Proceedings of the Cambridge Philosophical Society*, **26**, 528–535. [CP # 84, **2**, pp428–436.]

— [1932] "Inverse probability and the use of Likelihood," *Proceedings of the Cambridge Philosophical Society*, **28**, 257–261. [CP # 95 **3**, 13–17.]

— [1933] "The concepts of Inverse Probability and Fiducial Probability Referring to Unkonwn Parameters," *Proceedings of the Royal Society of London, A* **139**, 343–348. [CP # 102 **3**, 78–83.]

— [1935] "The Fiducial Argument in Statistical Inference," *Annals of Eugenics,* **6**, 391–398. [CP # 125 316–324.]

— [1935] "Statistical Tests," *Nature*, **136**, 474. [CP # 127, **3** 328–329.]

— [1941] "The Asymptotic Approach to Behrens's Integral, With Further Tables for the d Test of Significance." *Annals of Eugenics,* **11**, 141–172. [CP #180, **4**, 323–354.]

— [1945] "The Logical Inversion of the Notion of the Random Variable," *Sankhyā,* **7**, 129–132. [CP #203, **4**, 506–509.

— [1948] "Conclusions fiduciaires," *Annales de l'Institut Henri Poincaré,* **10**, 191–213. [CP # 222, **5**, 13–35.]

— [1955] "Statistical Methods and Scientific Induction," *Journal of the Royal Statistical Society, B*, **17**, 69–78. [CP # 261, **5**, 339–348.]

— [1957] "The Underworld of Probability," *Sankhyā*, **18**, 201–210 . [CP #267, **5** 366–375 .]

— [1959] "Mathematical Probability in the Natural Science," *Technometrics,* **1**, 21–29; reprinted in *Metrika,* **2**, 1–10; translated into Italian in *La Scuola in Azione,* **20**, 5–19. [CP # 273, **5**, 398–406.]

— [1960] "On Some Extensions of Bayesian Inference Proposed by Mr. Lindley," *Journal of the Royal Statistical Society, B*, **22**, 299–301. [CP #283, **5**, 485–487.] (See also, Lindley, D. V. [1958], "Fiducial distributions and Bayes's theorem," *Journal of the Royal Statistical Society, B*, **20** 102–107.)

— [1962] "Some Examples of Bayes' Method of the Experimental Determination of Probabilities *A Priori*," *Journal of the Royal Statistical Society, B*, **24**, 118–124. [CP # 289, **5** 521–527.]

— [1973] *Statistical Methods and Scientific Inference*. Hafner Press. (This edition was revised posthumously to include some notations Fisher had made in his copy for later revision). Earlier editions: 1956, 1959.

Dempster

These are the papers tracing the early history of Dempsters exploration of fiducial inference and belief functions.

Dempster, A.P. [1963] "On direct probabilities," *Journal of the Royal Statistical Society, B* **20**, 102–107.

— [1963a] "Further examples of inconsistencies in the fiducial argument," *Annals of Mathematical Statistics,* **34**, 884–891.

— [1963b] "On a paradox concerning inference about a covariance matrix." *Annals of Mathematical Statistics*, **34**, 1414–1418.

— [1964] "On the difficulties inherent in Fisher's fiducial argument," *Journal of the Americal Statistical Association,* **59**, 56–66.

— [1964a] "Some Revisionist Approaches to Fiducial Inference from Samples," *Bulletin of the International Statistical Institute,* **40**, 870. (Abstract only), also discussion in same issue.

— [1966] "New Methods for Reasoning Towards Posterior Distributions Based on Sample Data." *Annals of Mathematical Statistics,* **37**, 355-374.

— [1967a] "Upper and Lower Probabilities Induced by a Multivalued Mapping." *Annals of Mathematical Statistics,* **38**, 325-339.

— [1967b] "Upper and lower probabilities inferences based on a sample from a finite univariate population. *Biometrika,* **54**, 515–528.

— [1968a] "A Generalization of Bayesian Inference (with discussion)." *Journal of the Royal Statistical Society, Series B,* **30**, 205-247.

— [1968b] "Upper and Lower Probabilities Generated by a Random Closed Interval." *Annals of Mathematical Statistics,* **38**, 325-339.

— [1989] "Bayes, Fisher, and Belief Functions." in Geisser, Hodges, Press and Zellner, eds. *Bayesian Likelihood Methods in Statistics and Econometrics.* North-Holland.

Other Reference

Almond, R. G. [1989] *Fuision and Propagation of graphical belief models: an implementation and an example,* Ph.D. dissertation and Technical Report, S-130, Harvard University, Department of Statistics.

Diaconis, P. [1978] "Review of *A Mathematical Theory of Evidence,*" *Journal of the American Statistical Society,* **78**, 677–678.

— and Zabell, S. L. [1986]. Some alternatives to Bayes Rule. In Grofman, B. and Guillemo, O. *Information pooling and group decision making,* JAI Press, Greenwich, Conn.

Good, I. J. [1979] "Some history of the hierarchical Bayesian methodology," in Bernardo, J. M., *et al,* eds. *Bayesian Statistics,* University of Valencia Press, 489–519, with discussion. Also in Good, I.J.[1983] *Good Thinking,* pp 95–105, without discussion.

Lauritzen, D. J. and Spiegelhalter, S. L. [1988]. "Fast manipulation of probabilities with local representations—with applications to expert systems (with discussion)," *Journal of the Royal Statistical Society, B,* **50**, 205–247.

Shafer, G. [1976], *A Mathematical Theory of Evidence,* Princeton University Press.

Wasserman, L. A. [1988] *Some applications of belief functions to statistical inference.* PhD.Dissertation, Department of Preventive Medice and Biostatistics, University of Toronoto.

— [1990] "Belief functions and statistical inference," *The Canadian Journal of Statistics*, **18**, 183–196.