Reprinted from JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS
All Rights Reserved by Academic Press, New York and London

Vol. 175, No. 2, May 15, 1993
Printed in Belgium

Bernoulli Trials: From a Fuzzy Measure Point of View

YUAN YAN CHEN

U.S. Army Concept Analysis Agency,
Bethesda, Maryland 20814

Submitted by Ulrich Höhle

Received February 19, 1991

Bernoulli ("Ars Conjectandi," Basle, 1713) proved the first limit theorem of law of large numbers which provided the foundation of probability and statistical theory. However, the problem of Bernoulli trials is still unsettled (e.g., see Hacking, "The Emergence of Probability," Cambridge Univ. Press, Cambridge, 1975). It is from different interpretations of the relationship between the Bernoulli trials and relative frequency that we have different schools of probability theories (e.g., see Cox (*Amer. J. Phys.* **14**, No. 1 (1946), 1–13) and Fine (*IEEE Trans. Inform. Theory* **IT-16**, No. 3 (1970), 251–257)). In this paper we give a new treatment of the Bernoulli trials based on fuzzy measure, and we interpret the Bernoulli trials through the interaction of probability and possibility measures.

1. INTRODUCTION

In the Bernoulli trials, if we are given a probability of success, p, then we can predict the probability of success of any trial. The question is where the priori knowledge of p came from. Bernoulli showed that p will tend to s_n, which is the observed relative frequency of success as n increase. Thus one school of probability (e.g., Von Mises) considered the limit of the relative frequency as the definition of probability. But the limit of relative frequency is simply a conceptual value, and we cannot afford an infinite number of trials. Another school of probability (e.g., Neyman and Pearson) proposed to determine p by hypotheses testing of H_0: $p = p_0$ vs H_1: $p \neq p_0$. However, failing to reject the null hypothesis does not imply we are certain that $p = p_0$, since we can set up a slightly different null hypothesis and still fail to reject it. Thus by assigning a single value p for Bernoulli trials is just a priori, which cannot be confirmed precisely through sample trials. Nevertheless, do we need a single value of p in order to make inference in the Bernoulli trials? Intuitively even without priori when we observed several number of trials we still can state that the probability of success for

0022-247X/93 $5.00

the next trial is approximately s_n. So to determine which priori p for the Bernoulli trials seems to be less critical.

An important characteristic of the Bernoulli trials is the assumption that the propensity of trials is unchanged throughout the trials. Under this assumption we can claim that the observed trials are the sample realization of the population of all possible trials. Since each trial is either a success or a failure so the trials can be characterized by a parameter θ, which is the proportion of successes of the population. If we know θ, then the prediction of each trial is determined. So under such a model it is not the priori p we intend to estimate; it is the population parameter θ we are estimating. Based on the sample trials we can only obtain partial knowledge about θ. If we represent this knowledge by probability measure, then we have the Bayesian inference. However, if we represent the partial knowledge of θ by possibility measure, then we have a new inference proposed in this paper.

Although Bernoulli's theorem provided the foundation for the school of objective probability, but I believe Bernoulli viewed probability as a subjective entity. This can be seen from the title of his book, from his considering probability as a degree of certainty, and from his proposal of nonadditive probability. This is the philosophy we adopt in this paper. We consider probability theory as a collection of intrinsic rules of human thinking. It is a theory of belief logic, and an extension of Boolean logic. By extending the Boolean logic to the belief logic there are two kinds of extensions. One is the logic of expectation and the other is the logic of likelihood, which are corresponding to aleatory (chance) and epistemic (belief) concepts of probability. If a belief is for predicting a sample event, then we can split the belief into several pieces, since there can be several possible truths. But if a belief is for estimating a hypothesis, then the belief cannot be subdivided, since there can be only one possible true. Thus we need different calculus to represent these two types of belief logic.

2. Plausibility and Belief Measure

First we define a plausibility measure which is a subclass of the fuzzy measure.

Definition. Let $\mathscr{B}$ be a Borel-σ-Algebra on Ω. A fuzzy measure $\mathrm{Pl}: \mathscr{B} \to [0, 1]$ is a plausibility measure if

$$
\begin{aligned}
&\text{(i)} \quad \mathrm{Pl}(\varnothing) = 0, \qquad \mathrm{Pl}(\Omega) = 1,\\
&\text{(ii)} \quad A \subset B \Rightarrow \mathrm{Pl}(A) \leqslant \mathrm{Pl}(B) \text{ (Isotonicity)},\\
&\text{(iii)} \quad \mathrm{Pl}(A \cup B) \leqslant \mathrm{Pl}(A) + \mathrm{Pl}(B) \text{ (Subadditive)},\\
&\text{(iv)} \quad A_n \uparrow A \Rightarrow \mathrm{Pl}(A_n) \uparrow \mathrm{Pl}(A).
\end{aligned}
\tag{2.1}
$$

A conjugate of plausibility measure is a belief measure which is defined as $\mathrm{Bel}(A) = 1 - \mathrm{Pl}(\bar{A})$. So a belief measure satisfies

$$\begin{aligned} &(\mathrm{i})' \quad \mathrm{Bel}(\varnothing) = 0, \qquad \mathrm{Bel}(\Omega) = 1, \\ &(\mathrm{ii})' \quad A \subset B \Rightarrow \mathrm{Bel}(A) \leqslant \mathrm{Bel}(B) \text{ (Isotonicity)}, \\ &(\mathrm{iii})' \quad \mathrm{Bel}(A) + \mathrm{Bel}(B) - 1 \leqslant \mathrm{Bel}(A \cap B), \\ &(\mathrm{iv})' \quad A_n \downarrow A \Rightarrow \mathrm{Bel}(A_n) \downarrow \mathrm{Bel}(A). \end{aligned} \tag{2.2}$$

From (iii)′ we have $\mathrm{Bel}(A) + \mathrm{Bel}(\bar{A}) \leqslant 1$, therefore $\mathrm{Bel}(A) \leqslant \mathrm{Pl}(A)$, $\forall A$. If (iii) and (iii)′ are replaced by

$$(\mathrm{v}) \quad \mathrm{Pl}(A \cup B) = \sup\{\mathrm{Pl}(A), \mathrm{Pl}(B)\}, \ \forall A, B, \tag{2.3}$$

and

$$(\mathrm{v})' \quad \mathrm{Bel}(A \cap B) = \inf\{\mathrm{Bel}(A), \mathrm{Bel}(B)\} \ \forall A, B, \tag{2.4}$$

then we have the possibility and necessity measure proposed by Zadeh [12].

Obviously the plausibility measure contains the possibility and the probability measures. Without ambiguity we use Pl, Bel to denote the plausibility and belief measure, as well as the possibility and necessity measures. And we use P to denote the probability measure.

Remark. Conditions (iii) and (iii)′ are weaker than the belief function defined by Shafer [9], thus this class of fuzzy measure also contains the belief function measure.

Similar to the probability measure, the possibility measure can be characterized by a distribution function. Let $l(\theta) = \mathrm{Pl}(\{\theta\})$, then

$$\mathrm{Pl}(A) = \sup_{\theta \in A} l(\theta), \qquad \text{and} \qquad \mathrm{Bel}(A) = 1 - \sup_{\theta \notin A} l(\theta). \tag{2.5}$$

We refer to the function $l(\theta)$ as the likelihood function, since it is related to the likelihood function in statistical inference. This function is called the possibility distribution function by Zadeh, except we restrict the sup norm of likelihood function to be 1.

Next we define a rule of conditioning.

Definition. If $\mathrm{Pl}(B) > 0$, then the conditional plausibility of A given B is

$$\mathrm{Pl}(A \mid B) = \mathrm{Pl}(A \cap B)/\mathrm{Pl}(B) \tag{2.6}$$

and the conditional belief of A given B is $\mathrm{Bel}(A \mid B) = 1 - \mathrm{Pl}(\bar{A} \mid B)$.

If a plausibility measure is a belief function measure then (2.6) reduces to Dempster's rule of conditioning. If it is a probability measure, then (2.6) reduces to the Bayes's rule of conditioning.

Remark. Under this rule of conditioning Shafer's belief function can be represented by $\mathrm{Pl}(A) = \sum_B \mathrm{Pl}(A \mid B)\, P(B)$, and $\mathrm{Bel}(A) = \sum_B \mathrm{Bel}(A \mid B)\, P(B)$, where $\mathrm{Pl}(A \mid B)$ and $\mathrm{Bel}(A \mid B)$ are conditional possibility and necessity measures, and P is the basic probability assignment. Thus Shafer's belief function theory can be considered as a random set theory, where B's are the sampling units.

3. Random Variable and Stationary Variable

In this section we develop a concept of stationary variable which is associated with the likelihood judgment. A variable whose true value is unique but unknown to us is called a stationary variable. The reason for using the term of stationary variable is because its true value is unchanged with respect to time. In general a stationary variable can be considered as an unknown parameter, an unknown hypothesis, or an unknown past event. In statistical inference, when we are sampling from a fixed unknown population, we can consider the perspective sample statistics as random variables and consider the population statistics as stationary variables. We propose to measure stationary variable by possibility measure. Note that stationary variable is also referred to as fuzzy variable by some authors.

There are two basic axioms for the belief constructions.

Axiom 1. If X is a random variable on the sample space $\mathbf{X}$ and each value in $\mathbf{X}$ is equally likely to be true, then we have $P(X = x) = 1/|\mathbf{X}|$, $\forall x \in X$, where $|\mathbf{X}|$ is the cardinal of $\mathbf{X}$.

Axiom 2. If θ is a stationary variable in the hypothesis space Θ and each value in Θ are equally likely to be true, then we have $l(\theta) = 1$, $\forall \theta \in \Theta$, which implies $\mathrm{Bel}(A) = 0$, $\mathrm{Pl}(A) = 1\ \forall A \subsetneq \Theta$.

Both axioms are principle of insufficient reasoning, but depending on sample space or hypothesis space, we have different belief representations.

Next we show how to combine random variable and stationary variable.

Definition. If $X \mid \theta$ is a random variable on $\mathbf{X}$ with probability function $p(x \mid \theta)$ and θ is a stationary variable on Θ with likelihood function $l(\theta)$, then the joint plausibility measure is

$$\mathrm{Pl}(X \in A, \theta \in B) = \sup_{\theta \in B} \int_A \mathrm{pl}(x, \theta)\, dx, \tag{3.3}$$

where $\mathrm{pl}(x, \theta) = p(x \mid \theta) \mid (\theta)$ is a joint plausibility function.

From (3.3) we can obtain two marginal meausures. The marginal measure for θ is still a possibility measure, since $\mathrm{Pl}(\theta \in B) = \sup_{\theta \in B} \int_X p(x \mid \theta)\, l(\theta)\, dx = \sup_{\theta \in B} l(\theta)$; but the marginal measure for X is not a probability measure.

THEOREM 3.1. *If $X \mid \theta \sim p(x \mid \theta)$ and $l(\theta)$ is a likelihood function, then*

$$\begin{aligned} \mathrm{Pl}(X \in A) &= \sup_{\theta \in \Theta} \int_A p(x \mid \theta)\, l(\theta)\, dx \\ \mathrm{Bel}(X \in A) &= 1 - \sup_{\theta \in \Theta} \int_{\bar{A}} p(x \mid \theta)\, l(\theta)\, dx \end{aligned} \tag{3.4}$$

are plausibility and belief measures.

Proof.

$$\begin{aligned} \mathrm{Pl}(X \in A \cup B) &= \sup_{\theta \in \Theta} \int_{A \cup B} p(x \mid \theta)\, l(\theta)\, dx \\ &\leqslant \sup_{\theta \in \Theta} \left\{ \int_A p(x \mid \theta)\, l(\theta)\, dx + \int_B p(x \mid \theta)\, l(\theta)\, dx \right\} \\ &\leqslant \sup_{\theta \in \Theta} \int_A p(x \mid \theta)\, l(\theta)\, dx + \sup_{\theta \in \Theta} \int_B p(x \mid \theta)\, l(\theta)\, dx. \end{aligned}$$

Remark. Equation (3.4) is a special case of fuzzy integral in [4], where $\otimes$ is the usual multiplication and $\oplus = \vee$ is a pseudo-addition.

If $l(\theta)$ is degenerate then we have $\mathrm{Bel}(X \in A) = \mathrm{Pl}(X \in A)$, and it reduces to the probability model. If $\Theta = \mathbf{P}$ is a class of probability measures and $l(P) = 1\ \forall P \in \mathbf{P}$ then we have $\mathrm{Pl}(X \in A) = \sup\{P(X \in A) \mid P \in \mathbf{P}\}$ and $\mathrm{Bel}(X \in A) = \inf\{P(X \in A) \mid P \in \mathbf{P}\}$, and it reduces to an upper and lower probabilities model of Huber and Strassen [7].

EXAMPLE 1. If an urn contains ten balls, and we have only partial knowledge that either three or four of them are red. What is the probability that a ball selected at random will be red?

The prior information indicates that

$$l(\theta) = \begin{cases} 1, & \text{if } \theta = 3/10, 4/10; \\ 0, & \text{otherwise.} \end{cases}$$

So we have

$$\mathrm{Pl}(X = 1) = \sup_{0 \leqslant \theta \leqslant 1} \theta\, l(\theta) = 4/10,$$

$$\mathrm{Bel}(X = 1) = 1 - \sup_{0 \leqslant \theta \leqslant 1} (1 - \theta)\, l(\theta) = 3/10.$$

THEOREM 3.2. *If $X_1, ..., X_n \mid \theta \sim$ i.i.d. $p(x \mid \theta)$ is a prior belief for θ. Then the posterior belief for $\theta \mid x_1, ..., x_n$ is*

$$l(\theta \mid x_1, ..., x_n) = \frac{l(\theta)\, p(x_1, ..., x_n \mid \theta)}{\sup\{l(\theta)\, p(x_1, ..., x_n \mid \theta) \mid \theta \in \Theta\}} \tag{3.5}$$

Proof.

$$\begin{aligned} l(\theta \mid x_1, ..., x_n) &= \text{pl}(\theta, x_1, ..., x_n)\, \text{pl}(x_1, ..., x_n) \\ &= \text{pl}(\theta, x_1, ..., x_n)/\sup_{\theta \in \Theta} \text{pl}(\theta, x_1, ..., x_n) \\ &= p(x_1, ..., x_n \mid \theta)\, l(\theta)/\sup_{\theta \in \Theta} p(x_1, ..., x_n \mid \theta)\, l(\theta). \end{aligned}$$

If $l(\theta)$ and $l(\theta \mid x_1, ..., x_n)$ represent the relative odds of the prior and posterior beliefs, then in terms of odds ratio, the belief update by (3.5) is equivalent to the belief update of the Bayesian Inference. However, if the prior belief is vacuous we can always let $l(\theta) = 1$, $\forall \theta \in \Theta$, thus the improper prior problem of the Bayesian inference can be avoided.

If the prior belief is vacuous (3.5) reduces to $l(\theta \mid x_1, ..., x_n) = kp(x_1, ..., x_n \mid \theta)$, which also satisfies a classical statistical principle, $L(\theta \mid x_1, ..., x_n) \propto p(x_1, ..., x_n \mid \theta)$. Under this inference a point estimate is less important, since we can use $\text{Bel}(\theta \in A \mid x_1, ..., x_n)$ as our confidence level that θ is in A. However, if $(\hat{\theta} \mid x_1, ..., x_n) = 1$ then $\hat{\theta}$ is an MLE, which is the same as MLE in classical statistical inference.

THEOREM 3.3. *If $X_1 \mid \theta$ and $X_2 \mid \theta$ are independent random variables with probability function $p_1(x_1 \mid \theta)$ and $p_2(x_2 \mid \theta)$, respectively, and $l(\theta) = 1$, $\forall \theta \in \Theta$. If $\sup_{\theta \in \Theta} l_1(\theta \mid x_1)\, l_2(\theta \mid x_2) > 0$, then*

$$l(\theta \mid x_1, x_2) = \frac{l_1(\theta \mid x_1)\, l_2(\theta \mid x_2)}{\sup\{l_1(\theta \mid x_1)\, l_2(\theta \mid x_2) \mid \theta \in \Theta\}} \tag{3.6}$$

Proof.

$$\begin{aligned} l(\theta \mid x_1, x_2) &= \text{pl}(\theta, x_1, x_2)/\text{pl}(x_1, x_2) = p(x_1, x_2 \mid \theta)/\sup_{\theta \in \Theta} p(x_1, x_2 \mid \theta) \\ &= p_1(x_1 \mid \theta)\, p_2(x_2 \mid \theta)/\sup_{\theta \in \Theta} p_1(x_1 \mid \theta)\, p_2(x_2 \mid \theta) \\ &= \text{pl}_1(\theta, x_1)\, \text{pl}_2(\theta, x_2)/\sup_{\theta \in \Theta} \text{pl}_1(\theta, x_1)\, \text{pl}_2(\theta, x_2) \\ &= l_1(\theta \mid x_1)\, l_2(\theta \mid x_2)\, \text{pl}_1(x_1)\, \text{pl}_2(x_2) \\ &\quad /\sup_{\theta \in j} l_1(\theta \mid x_1)\, l_2(\theta \mid x_2)\, \text{pl}_1(x_1)\, \text{pl}_2(x_2) \\ &= l_1(\theta \mid x_1)\, l_2(\theta \mid x_2)/\sup_{\theta \in \Theta} l_1(\theta \mid x_1)\, l_2(\theta \mid x_2). \end{aligned}$$

Equation (3.6) is known as the likelihood rule. It has been discussed in Smets [11] and Shafer [10]. Similar to the Dempster's rule of combination this rule reduces to the Bernoulli's rule, when $l(\theta \mid x_1)$ and $l(\theta \mid x_2)$ are likelihood functions of simple belief functions focus on the same element A.

THEOREM 3.4. *If* $X_1, ..., X_n \mid \theta$ *i.i.d.* $\sim p(x \mid \theta)$ *and* $l(\theta) = 1$, $\forall \theta \in \Theta$, *then*

$$\mathrm{Pl}(X_{n+1} \in A \mid x_1, ..., x_n) = \sup_{\theta \in \Theta} \int_A p(x_{n+1} \mid \theta)\, l(\theta \mid x_1, ..., x_n)\, dx_{n+1} \tag{3.7}$$

$$\mathrm{Bel}(X_{n+1} \in A \mid x_1, ..., x_n) = 1 - \sup_{\theta \in \Theta} \int_{\bar{A}} p(x_{n+1} \mid \theta)\, l(\theta \mid x_1, ..., x_n)\, dx_{n+1}.$$

Proof.

$$\begin{aligned}
\mathrm{pl}(x_{n+1} \mid x_1, ..., x_n) &= \mathrm{pl}(x_1, ..., x_n, x_{n+1})/\mathrm{pl}(x_1, ..., x_n) \\
&= \sup_{\theta \in \Theta} p(x_1, ..., x_n, x_{n+1}, \theta)/\sup_{\theta \in \Theta} p(x_1, ..., x_n, \theta) \\
&= \sup_{\theta \in \Theta} p(x_1, ..., x_n, x_{n+1} \mid \theta)/\sup_{\theta \in \Theta} p(x_1, ..., x_n \mid \theta) \\
&= \sup_{\theta \in \Theta} p(x_{n+1} \mid \theta)\, p(x_1, ..., x_n \mid \theta)/\sup_{\theta \in \Theta} p(x_1, ..., x_n \mid \theta) \\
&= \sup_{\theta \in \Theta} p(x_{n+1} \mid \theta)\, l(\theta \mid x_1, ..., x_n).
\end{aligned}$$

The previous theorem indicates that the posterior likelihood $l(\theta \mid x_1, ..., x_n)$ can be considered as the prior belief for predicting X_{n+1}.

THEOREM 3.5 (Rule of Succession). *If* $X_1, ..., X_n \mid \theta$ *i.i.d.* $\sim$ *Bernoulli*(θ) *and* $l(\theta) = 1$, $0 \leqslant \theta \leqslant 1$, *then*

$$l(\theta \mid x_1, ..., x_n) = \begin{cases} \dfrac{\theta^x (1-\theta)^{n-x}}{(x/n)^x\, ((n-x)/n)^{n-x}}, & \text{if } 0 < x < n \\ \theta^n, & \text{if } x = n \\ (1-\theta)^n, & \text{if } x = 0. \end{cases}$$

$$\mathrm{Bel}(X_{n+1} = 1 \mid x_1, ..., x_n) = \begin{cases} 1 - \left(\dfrac{n-x+1}{n+1}\right)\left(\dfrac{n-x+1}{n-x}\right)^{n-x} \times \left(\dfrac{n}{n+1}\right)^n, & \text{if } x < n \\ 1 - \left(\dfrac{1}{n+1}\right)\left(\dfrac{n}{n+1}\right)^n, & \text{if } x = n \end{cases}$$

$$\mathrm{Pl}(X_{n+1}=1 \mid x_1, ..., x_n) = \begin{cases} \left(\frac{x+1}{n+1}\right)\left(\frac{x+1}{x}\right)^x \left(\frac{n}{n+1}\right)^n, & \text{if } x>0, \\ \left(\frac{1}{n+1}\right)\left(\frac{n}{n+1}\right)^n, & \text{if } x=0 \end{cases} \tag{3.8}$$

where $x = \sum_{i=1}^n x_i$.

Proof. By directly applying (3.5) and (3.7) we have the results.

THEOREM 3.6. *If* $X_1, ..., X_n \mid \theta$ *i.i.d.* $\sim$ *Bernoulli*(θ) *and* $l(\theta)=1$, $0 \leqslant \theta \leqslant 1$, *then we have*

$$\mathrm{Bel}(X_{n+1}=1 \mid x_1, ..., x_n) \leqslant s_n \leqslant \mathrm{Pl}(X_{n+1}=1 \mid x_1, ..., x_n), \tag{3.9}$$

where $s_n = x/n$.

Proof. From (3.4) we have

$$\mathrm{Bel}(X \in A) \leqslant P(X \in A \mid \hat{\theta}) \leqslant \mathrm{Pl}(X \in A), \qquad \text{if } \hat{\theta} \text{ is an MLE.}$$

From (3.8) $l(s_n \mid x_1, ..., x_n) = 1$, so we have the result.

The model of the Bernoulli trials can be considered as extension of the urn model in Example 1. In the Bernoulli trials we have infinite number of balls in the urn. Based on finite number of trials we have only partial knowledge about θ; therefore the prediction of the next trial becomes imprecise. Table I gives a few values of Bel and Pl for different n and x. From this table we can see that Bel and Pl converge rather fast to relative frequency (The rate of convergence is $O(1/n)$). Thus a single value s_n can provide a reasonable approximation if n is sufficiently large.

Theorem 3.5 also indicates that if we accept the two principles of insufficient reasoning; then the consistency of relative frequency and probability becomes the consequence of the two principles. Therefore there is no need to define probability as the limit of the relative frequency. Although Bayesian argument, which does not distinguish the two insufficient reasoning principles, also produces a similar result, e.g., Laplace's rule of succession: $P(X_{n+1}=1 \mid x_1, ..., x_n) = (x+1)/(n+2)$. But the bias of the Laplace's rule against the relative frequency seems to be unreasonable. Moreover, in the case if we observe $x_1 = x_2 = \cdots x_n = 1$, the need of upper and lower probability becomes inevitable. Since if we assign

$$\text{Probability } (X_{n+1}=1 \mid x_1 = x_2 = \cdots x_n = 1) = p,$$

TABLE I

Predictive Intervals for the Next Success, Given x Successes out of n Trials

$n=10$			$n=100$			$n=1{,}000$			$n=10{,}000$			$n=100{,}000$		
x	Bel	Pl	x	Bel	Pl	x	Bel	Pl	x	Bel	Pl	x	Bel	Pl
1	0.0953	0.1402	10	0.0995	0.1044	100	0.09995	0.10045	1,000	0.099995	0.100045	10,000	0.0999995	0.1000045
3	0.2860	0.3323	30	0.2985	0.3035	300	0.29985	0.30035	3,000	0.299985	0.300035	30,000	0.2999985	0.3000035
5	0.4767	0.5233	50	0.4975	0.5025	500	0.49975	0.50025	5,000	0.499975	0.500025	50,000	0.4999975	0.5000025

(i) if $p=1$, this implies the next success is a certainty. However, we cannot make such a strong statement based on the sample evidence.

(ii) if $p<1$, then no matter how close p is to 1 this implies there is a small chance the next instance will fail; also in the long run we expect to observe some failure with certainty. Again such a statement cannot be supported by the sample evidence.

The only way to circumvent this logical inconsistency is by assigning

$$\text{Probability } (X_{n+1}=1 \mid x_1=x_2=\cdots x_n=1) \geqslant p.$$

Theorem 3.7 (law of large number) If $X_1, ..., X_n \mid \theta$ i.i.d. $\sim$ Bernoulli (θ) and $l(\theta)=1$, $0 \leqslant \theta \leqslant 1$, then $\forall \varepsilon > 0$, we have

$$\lim_{n \to \infty} \text{Bel}(|\theta - s_n| < \varepsilon \mid x_1, ..., x_n) \to 1. \tag{3.10}$$

To prove the theorem we need the following well known lemma.

LEMMA. *Let t be a nonnegative real numbers, and suppose $0<\lambda<1$. Then $t^{\lambda} \leqslant (1-\lambda)+\lambda t$.*

Proof of the Theorem. It is trivial for the case $s_n=0$, or $s_n=1$, so we only show for the case $0<s_n<1$. Since $l(\theta \mid x_1, ..., x_n)$ is increasing for $\theta < s_n$ and decreasing for $\theta > s_n$, we have $\text{Pl}(|\theta - s_n| \geqslant \varepsilon \mid x_1, ..., x_n) \leqslant \sup(l(s_n-\varepsilon \mid x_1, ..., x_n), \ l(s_n+\varepsilon \mid x_1, ..., x_n) = \sup((1-\varepsilon/s_n)^x (1+\varepsilon/(1-s_n))^{n-x}, (1+\varepsilon/s_n)^x (1-\varepsilon/(1-s_n))^{n-x})$. By the previous lemma we have $(1-\varepsilon/s_n)^{s_n} (1+\varepsilon/(1-s_n))^{1-s_n} \leqslant (1-\varepsilon)(1+\varepsilon) = 1-\varepsilon^2$, thus we obtain $l(s_n - \varepsilon \mid x_1, ..., x_n) \leqslant (1-\varepsilon^2)^n$. Similarly $l(s_n+\varepsilon \mid x_1, ..., x_n) \leqslant (1-\varepsilon^2)^n$; this proves the theorem.

Theorem 3.7 is different from the classical law of large number. Since our conditional belief is based on the observed empirical evidence $x_1, ..., x_n$, thus s_n is simply a constant and not a random variable. However, the theorem indicates that as the number of trials increases the knowledge of θ becomes crisp. Therefore, in a loose sense we can use a single value probability for the Bernoulli trials.

4. CONSISTENCY OF BOOLEAN LOGIC AND BELIEF LOGIC

The logical view of probability has been advocated by many authors (e.g., – Keynes [8]). In this section we show that the belief logic is consistent with the Boolean logic under the rule of conditioning. Under the belief logic a proposition A is true if $\text{Bel}(A)=1$ or $\text{Pl}(\bar{A})=0$, and A is false if $\text{Bel}(\bar{A})=1$ or $\text{Pl}(A)=0$. Note that $\text{Bel}(A)=0$ does not implies A is false since we might have $\mathrm{Pl}(A) > 0$. However, $\mathrm{Pl}(A) = 0$ does imply A is false, since it implies $\mathrm{Bel}(A) = \mathrm{Pl}(A) = 0$. The conditional belief $\mathrm{Bel}(B \mid A)$ is interpreted as the degree of certainty that the proposition "If A then B" is true. A degree of certainty can be a chance or a belief depending on the proposition.

THEOREM 4.1 (Modus Ponens $p \to q,\ p \Rightarrow q$).

$$\mathrm{Bel}(B \mid A) = 1, \qquad \mathrm{Bel}(A) = 1 \Rightarrow \mathrm{Bel}(B) = 1.$$

Proof.

$$\begin{aligned} \mathrm{Pl}(\bar{B} \cap A) &= \mathrm{Pl}(\bar{B} \mid A)\,\mathrm{Pl}(A) = 0, \\ \mathrm{Pl}(\bar{B} \cap \bar{A}) &= \mathrm{Pl}(\bar{B} \mid A)\,\mathrm{Pl}(\bar{A}) = 0 \\ &\Rightarrow \mathrm{Pl}(\bar{B}) \leqslant \mathrm{Pl}(\bar{B} \cap A) + \mathrm{Pl}(B \cap \bar{A}) = 0. \end{aligned}$$

THEOREM 4.2 (Modus Tollens $p \to q,\ \sim q \Rightarrow \sim p$).

$$\mathrm{Bel}(B \mid A) = 1, \qquad \mathrm{Pl}(B) = 0 \Rightarrow \mathrm{Pl}(A) = 0.$$

Proof.

$$\begin{aligned} \mathrm{Pl}(A \cap B) &= \mathrm{Pl}(A \mid B)\,\mathrm{Pl}(B) = 0, \\ \mathrm{Pl}(A \cap \bar{B}) &= \mathrm{Pl}(\bar{B} \mid A)\,\mathrm{Pl}(A) = 0 \\ &\Rightarrow \mathrm{Pl}(\bar{A}) \leqslant \mathrm{Pl}(A \cap B) + \mathrm{Pl}(A \cap \bar{B}) = 0. \end{aligned}$$

THEOREM 4.3 (Hypothetical Syllogism $p \to q,\ q \to r \Rightarrow p \to r$).

$$\mathrm{Bel}(B \mid A) = 1, \qquad \mathrm{Bel}(C \mid B) = 1 \Rightarrow \mathrm{Bel}(C \mid A) = 1.$$

Proof.

$$\begin{aligned} \mathrm{Pl}(\bar{C} \cap A) &\leqslant \mathrm{Pl}((\bar{C} \cap B) \cup (\bar{B} \cap A)) \leqslant \mathrm{Pl}(\bar{C} \cap B) + \mathrm{Pl}(\bar{B} \cap A) \\ &= \mathrm{Pl}(\bar{C} \mid B)\,\mathrm{Pl}(B) + \mathrm{Pl}(\bar{B} \mid A)\,\mathrm{Pl}(A) = 0 \\ &\Rightarrow \mathrm{Pl}(\bar{C} \mid A) = 0. \end{aligned}$$

THEOREM 4.4 (Disjunctive Syllogism $p \vee q,\ \sim p \Rightarrow q$).

$$\mathrm{Bel}(A \cup B) = 1, \qquad \mathrm{Pl}(A) = 0 \Rightarrow \mathrm{Bel}(B) = 1.$$

Proof.

$$\mathrm{Pl}(\bar{B}) = \mathrm{Pl}((\bar{B} \cap A) \cup (\bar{B} \cap \bar{A})) \leqslant \mathrm{Pl}(\bar{B} \cap A) + \mathrm{Pl}(\bar{B} \cap \bar{A}) = 0.$$

THEOREM 4.5 (Constructive Dilemma $p \to q$, $r \to s$, $p \vee r \Rightarrow q \vee s$).

$$\text{Bel}(B \mid A) = 1, \qquad \text{Bel}(D \mid C) = 1, \qquad \text{Bel}(A \cup C) = 1 \Rightarrow \text{Bel}(B \cup D) = 1.$$

Proof.

$$\begin{aligned} &\text{Pl}((A \cap \bar{B}) \cup (C \cap \bar{D}) \cup (\bar{A} \cap \bar{C})) \\ &\quad \leqslant \text{Pl}(\bar{B} \mid A)\, \text{Pl}(A) + \text{Pl}(\bar{D} \mid C)\, \text{Pl}(C) + \text{Pl}(\bar{A} \cap \bar{C}) = 0, \\ &\quad \text{and} \quad \bar{B} \cap \bar{D} \subset (A \cap \bar{B}) \cup (C \cap \bar{D}) \cup (\bar{A} \cap \bar{C}) \\ &\quad \Rightarrow \text{Pl}(\bar{B} \cap \bar{D}) = 0. \end{aligned}$$

THEOREM 4.6 (Absorption $p \to q \Rightarrow p \to p \wedge q$).

$$\text{Bel}(B \mid A) = 1 \Rightarrow \text{Bel}(A \cap B \mid A) = 1.$$

Proof.

$$\text{Pl}(\bar{B} \mid A) = 0, \qquad \text{Pl}(\bar{A} \mid A) = 0$$
$$\Rightarrow \text{Pl}(\bar{A} \cup \bar{B} \mid A) = 0.$$

THEOREM 4.7 (Simplification $p \wedge q \Rightarrow p$).

$$\text{Bel}(A \cap B) = 1 \Rightarrow \text{Bel}(A) = 1.$$

Proof.

$$\text{Bel}(A) \geqslant \text{Bel}(A \cap B) = 1.$$

THEOREM 4.8 (Conjunction $p, q \to p \wedge q$).

$$\text{Bel}(A) = 1, \qquad \text{Bel}(B) = 1 \Rightarrow \text{Bel}(A \cap B) = 1.$$

Proof.

$$\text{Pl}(\bar{A} \cup \bar{B}) \leqslant \text{Pl}(\bar{A}) + \text{Pl}(\bar{B}) = 0.$$

THEOREM 4.9 (Addition. $p \Rightarrow p \vee q$).

$$\text{Bel}(A) = 1 \Rightarrow \text{Bel}(A \cup B) = 1.$$

Proof.

$$\text{Bel}(A \cup B) \geqslant \text{Bel}(A) = 1.$$

The above theorems provide some justification of the proposed belief model. They also indicate that the probabilistic logic and the possibilistic logic are consistent with the Boolean logic.

5. Concluding Remark

In this paper we show that the Bernoulli trials can be more appropriately explained by the combination of probability and possibility measures than by the probability measure alone. Probability and possibility are the two basic belief measures; from their combination we have a variety of fuzzy measures. However, probability and possibility measures play different roles in belief logic inference; one is suitable for the inference of prediction and the other is suitable for the inference of estimation. In statistics the possibility measure can provide an inference method for hypothesis evaluation and interval estimate. This inference is a theory of support as opposed to a theory of decision. A discussion of these difference can be seen in Hacking [5].

Belief logic and fuzzy logic are both multi-valued logic. However, there is a definite unknown truth in the belief logic, but there is no definite truth in the fuzzy logic. Therefore, the inference rules of the two logic are different.

The rule of succession derived in this paper is a rule of inductive inference; and the law of large number is also a law of inference and not a law of nature. They are based on the analogy of past and future events in the Bernoulli trials. Therefore, if the future trials has different propensity from the past trials then we have a departure of probability and relative frequency.

References

1. J. Bernoulli, "Ars Conjectandi," Basle, 1713.
2. J. T. Cox, Probability, Frequency and reasonable expectation, *Amer. J. Phys.* **14**, No. 1 (1946), 1–13.
3. T. L. Fine, On the apparent convergence of relative frequency and its implications, *IEEE Trans. Inform. Theory* **IT-16**, No. 3 (1970), 251–257.
4. H. Ichihashi, H. Tanaka, and K. Asai, Fuzzy integral based on pseudo-additions and multiplication, *J. Math. Anal. Appl.* **130** (1988), 354–364.
5. I. Hacking, "Logic of Statistical Inference," Cambridge Univ. Press, Cambridge, 1965.
6. I. Hacking, "The Emergence of Probability," Cambridge Univ. Press, Cambridge, 1975.
7. J. Huber and V. Strassen, Minimax tests and the Neyman–Pearson lemma for capacities, *Ann. Statist.* **1** (1973), 251–263.
8. J. M. Keynes, "A Treatise on Probability," 3rd ed., Macmillian Press, London, 1973.
9. G. Shafer, "A Mathematical Theory of Evidence," Princeton Univ. Press, Princeton, NJ, 1976.
10. G. Shafer, Belief functions and possibility measures, *in* "The Analysis of Fuzzy Information" (J. C. Bizdek, Ed.), Vol. 2, CRC Press, Boca Raton, FL, 1987.
11. P. Smets, Discussion of 'Belief functions and parametric models', *J. Roy. Statist. Soc. B* **33** (1982), 343.
12. L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems* **1** (1978), 3–28.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium