LA-UR-03-8840

Approved for public release;
distribution is unlimited.

Title: PROBABILITY, CHANCE, AND THE PROBABILITY OF CHANCE

Author(s): NOZER SINGPURWALLA, GEORGE WASHINGTON UNIVERSITY

ALYSON WILSON, LANL

Submitted to: The American Statistician

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Form **836 (10/96)**

Probability, Chance, and The Probability of Chance*

Nozer D. Singpurwalla
The George Washington University, Washington, D.C. 20052

Alyson G. Wilson
Los Alamos National Laboratory, Los Alamos, NM 87545

October 7, 2004

Abstract

In our day-to-day discourse on uncertainty, words like belief, chance, likelihood, and probability are commonly encountered. Often, these words are used interchangeably, because they are intended to encapsulate some vaguely articulated notion about the unknowns. The purpose of this paper is to propose framework that is able to show how each of these terms can be made to reflect a distinct meaning. To construct this framework, we use a basic scenario upon which caveats are introduced. Each caveat motivates us to bring in one or more of the above labels. The scenario considered here is very basic; it arises in both the biomedical and the industrial contexts, and easily generalizes.

This paper is expository and much of what is said here is not new. We require of the reader an appreciation of the calculus of probability. However, in order to make our distinctions transparent, probability has to be interpreted subjectively, not as a relative frequency.

***Key Words*:** Belief Functions, Biometry, Likelihood, Plausibility, Quality Assurance, Reliability, Survival Analysis, Uncertainty, Vagueness.

*Los Alamos Unclassifed Report LA-UR-03-8840

1 Probability and Chance

1.1 Introduction: Statement of the Problem

Consider the following archetypal problem that commonly arises in the contexts of biomedicine, engineering, and the physical sciences.

Suppose that at some reference time τ, the "now time", YOU are asked to predict the time to failure T of some physical or biological unit. The capitalized YOU is to emphasize the fact that it is a particular individual, namely yourself, that has been asked to make the prediction. To facilitate prediction, you examine the unit carefully and learn all that you can about its genesis: how, when, and where it was made. You denote this information by $\mathcal{H}(\tau)$, for history at time τ. In the case of biological units, $\mathcal{H}(\tau)$ would pertain to genetic and/or medical information. Suppose, as is generally true, that based on $\mathcal{H}(\tau)$ you conclude that prediction with certainty is not possible. Consequently, you are now faced with two choices: walk away from the problem, or make an informed guess about T.

Suppose that you choose the second option and are prepared to make guesses about the event $(T \geq t)$, for some $t > 0$. In reliability, $t > 0$ is known as the "mission time". There are several additional caveats to this basic problem that go into forming our overall framework; these will be presented in Sections 2 and 3.

To keep the mathematics simple, you introduce a counter, say X, and adopt the convention that $X = 1$ (a "success") whenever $T \geq t$, and $X = 0$ (a "failure"), otherwise. Thus the events $(T \geq t)$ and $(X = 1)$ are isomorphic; however, there is a loss of granularity in going from T to X. This is because X continues to equal one, even when $T \geq t + a$, for any and all $a > 0$. With the introduction of X, informed guesses about $(T \geq t)$ boil down to informed guesses about $(X = 1)$. But what do we mean by an informed guess, and how shall we make this operational?

1.2 Personal Probability: Making Guesses Operational

By informed guess, we mean a quantified measure of your uncertainty about the event $(X = 1)$ in the light of $\mathcal{H}(\tau)$, and subsequent to a thoughtful evaluation of its consequences. Now, it is generally well acknowledged that probability is a satisfactory way to quantify uncertainty, and to some, like Lindley (1982), the only satisfactory way. There are several interpretations of probability [cf. Good(1965)]. The one we shall adopt is *personal probability*, also known as *subjective probability*. Here, you quantify your uncertainty about the event $(X = 1)$, based on $\mathcal{H}(\tau)$, by your personal probability denoted

$$P_{\mathcal{Y}}(X = 1; \mathcal{H}(\tau)). \tag{1}$$

The subscript indexing P emphasizes the fact that the specified probability is that of a particular individual, namely, you. For convenience, we set $\tau = 0$ and denote $\mathcal{H}(0)$ by simply $\mathcal{H}$. Henceforth, we also omit the subscript associated with P, so that Equation 1 is written

$$P(X = 1; \mathcal{H}) = p, \tag{2}$$

where $0 < p < 1$. The p so specified is a personal probability because it is not unique to all persons; more important, it can change with time for the same individual. This is because the background history for this person also changes, and it is the history that plays a key role in specifying a personal probability. Thus an informed guess is tantamount to specifying a p, where p is a personal probability.

To make an informed guess operational, that is, to make a pragmatic use of it, we need to interpret p. For this we appeal to de Finetti (1974) who proposed that p represent the amount you—the specifier of p—is willing to stake in a *two-sided bet* (or gamble) about the event $(X = 1)$. That is, should X turn out to be one, you receive as a reward one monetary unit against the p staked out by you. Should X turn out to be zero, then the amount staked, namely p, is lost. By a two-sided bet, we mean the willingness to stake p for the event $(X = 1)$, or an amount $(1 - p)$ for the event $(X = 0)$. That is, you are

indifferent between the two gambles: one monetary unit in exchange for p if $(X = 1)$, or one monetary unit in exchange for $(1 - p)$ if $(X = 0)$. It is useful to bear in mind that in keeping with the spirit of the individual nature of personal probability, the amount p represents your stake. For the same event $(X = 1)$, your colleague may choose to stake a different amount $\widetilde{p}$, with $\widetilde{p} \neq p$. It is also important to note that with p interpreted as a gamble, the bet will only be settled when X reveals itself. Thus bets can only be made operational for events that are ultimately observed. We do not consider here the disposition of the second party in the bet; we assume that the second party is willing to accept any bet put forth by you.

Thus to summarize, in the context of this paper, the word "probability" is used to denote the amount an individual is prepared to stake in a two-sided bet about an uncertain event. This probability can be specified based on $\mathcal{H}$ alone, and it is not essential that $\mathcal{H}$ contain data on items judged to be similar to the item in question. That is, personal probabilities can be specified without the benefit of having observed data.

1.3 Chance or Propensity: A Useful Abstraction

Whereas specifying a personal probability can be done solely by introspection considering $\mathcal{H}$, a more systematic approach, which involves breaking the problem into smaller, easier problems, begins with invoking the law of total probability on the event $(X = 1; \mathcal{H})$. Specifically, for some unknown quantity θ, $0 < \theta < 1$, and an entity $\pi(\theta; \mathcal{H})$, whose interpretation is given later in Section 1.4,

$$P(X = 1; \mathcal{H}) = \int_0^1 P(X = 1|\theta; \mathcal{H})\, \pi(\theta; \mathcal{H}) d\theta, \tag{3}$$

$$= \int_0^1 P(X = 1|\theta)\, \pi(\theta; \mathcal{H}) d\theta, \tag{4}$$

if you assume that X is independent of $\mathcal{H}$ given θ. That is, were you to know θ, then knowledge of $\mathcal{H}$ is unnecessary. The meaning of θ, known as a *parameter*, remains to be

discussed, but for now we state that in the language of personal probability, Equation 3 implies an *extension of the conversation* from $P(X = 1; \mathcal{H})$ to $P(X = 1|\theta; \mathcal{H})$. The idea here is that after invoking the assumption of independence, you may find it easier to quantify your uncertainty about $(X = 1)$ were you to know θ, than quantifying the uncertainty based on $\mathcal{H}$. Whereas the dimension of $\mathcal{H}$ can be very large, the dimension of θ is one. Thus the role of the parameter θ is to simplify the process of uncertainty quantification by imparting to X independence from $\mathcal{H}$.

In Equation 4, the quantity $P(X = 1|\theta)$ is known as a *probability model* for the binary X. Following Bernoulli, you let $P(X = 1|\theta) = \theta$, where $P(X = 1|\theta)$ represents your bet (personal probability) about the event $(X = 1)$ were you to know θ. This brings us to the question of what does θ mean? That is, how should we interpret θ?

The meaning of θ was made transparent by de Finetti [cf. Lindley and Phillips (1976)] in his now famous theorem on binary exchangeable sequences. Loosely speaking, this theorem says that if a large number of units judged similar to each other (the technical term is *exchangeable*) and to the unit in question were to be observed for their survival or failure until t, and if $X_i = 1$ if the ith item survived until t ($X_i = 0$ otherwise), then

$$\theta = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i; \tag{5}$$

that is θ is the average of the X_i's, when the number of X_i's is infinite. De Finetti refers to this θ as a *chance* or *propensity*. Note that there is no personal element involved in defining θ, other than the fact that θ derives from the behavior of exchangeable sequences, and exchangeability is a judgment. What you judge to be exchangeable may not sit well with your colleagues. Because θ connotes the limit of an exchangeable binary sequence, θ can be seen as an objective entity. More important, since θ cannot be actually observed (n in the Equation 5 is infinite), we claim that chance is an abstract construct. It is a useful abstraction all the same, because in writing $P(X = 1|\theta) = \theta$, you are saying that your stake on the uncertain event $(X = 1)$ is θ, were you to know θ. But no one can

possibly tell you what θ is, and this is what leads us to the next section. But before we do so, it may be of interest to mention a few words about two other interpretations of θ.

One is due to Laplace, who in keeping with the scientific climate of his time, and being influenced by Newton, was concerned with cause and effect relationships. Accordingly, to Laplace, θ was the cause of an effect, namely, the event $(X = 1)$. The second interpretation of θ stems from the relative frequency interpretation of probability. Indeed, here θ is taken to be the probability that $X = 1$.

Finally, even though the notion of chance introduced here has been in the context of binary variables, a parallel notion also exists for other kinds of variables.

1.4 Probability of Chance: Taking Chances with Chance

Since θ is unknown, and in principle can never be known, you are uncertain about θ. In keeping with the dictum that all uncertainty be described by probability, you let $P_{\mathcal{Y}}(\Theta \leq \theta; \mathcal{H})$ encapsulate your bet on the event $(\Theta \leq \theta)$. Here, in keeping with standard convention, all unknown quantities are denoted by capital letters and their realized values by the corresponding small letters. Since Θ can take all values in the continuum $(0, 1)$, we shall assume that $P_{\mathcal{Y}}(\Theta \leq \theta; \mathcal{H})$ is "absolutely continuous," so that its density at θ exists, for $0 < \theta < 1$. We denote this density by $\pi_{\mathcal{Y}}(\theta; \mathcal{H})$ and interpret it as

$$\pi(\theta; \mathcal{H}) d\theta \approx P\left(\theta \leq \Theta \leq \theta + d\theta; \mathcal{H}\right).$$

For convenience, the subscript has been dropped.

Thus $\pi(\theta; \mathcal{H}) d\theta$ is approximately your personal probability that the unknown chance Θ is in the interval $[\theta, \theta + d\theta]$. Since θ will never be known, the bet on Θ cannot be settled. However, since $\pi(\theta; \mathcal{H})$ goes into determining $P\left(X = 1; \mathcal{H}\right)$—see Equation 6 below—and since bets on $(X = 1; \mathcal{H})$ can be settled, $\pi(\theta; \mathcal{H})$ can also be interpreted as a technical device that helps you specify your bet on an observable.

With the above in place, plus the fact that in our case $P(X = 1|\theta) = \theta$, Equation 4

becomes

$$P\left(X=1;\mathcal{H}\right)=p=\int_0^1 \theta\cdot\pi(\theta;\mathcal{H})d\theta. \tag{6}$$

Equation 6 above is noteworthy. It embodies (i) a personal probability about the event $(X=1)$—the left hand side; (ii) a chance Θ taking the value θ; and (iii) a personal probability about the chance Θ belonging to the interval $[\theta,\theta+d\theta]$—the entity $\pi(\theta;\mathcal{H})d\theta$. This equation helps us make transparent the difference between probability, chance and the probability of chance.

There is another angle from which Equation 6 can be viewed. This comes from the fact that the right-hand side of Equation 6 is your *expected value* of Θ, the expected value being determined by your $\pi(\theta;\mathcal{H})$. Denoting this expected value by $E_{\mathcal{Y}}(\Theta)$, we have

$$P\left(X=1;\mathcal{H}\right)=p=E_{\mathcal{Y}}(\Theta),$$

implying that your personal probability for the event $(X=1)$ is your expected value of the chance Θ with respect to $\pi(\theta;\mathcal{H})$, your personal probability about chance.

2 The Likelihood of Chance

2.1 Introducing the Caveat of Data

We supplement the framework of the basic problem of Section 1.1 by introducing our first caveat. Suppose that in addition to $\mathcal{H}(\tau)$, you also have at hand the binary $x_1,\ldots,x_n$, where $x_i=1$ if the life-length of the i-th item has <u>actually</u> been observed to exceed t, and $x_i=0$, otherwise. The n items that go into constituting the data $\mathbf{x}=(x_1,\ldots,x_n)$ are judged by you, prior to observing the $\mathbf{x}$, to be similar (or exchangeable) to the item in question. What can you now say about the unobserved X? In other words what is your prediction for the event $(X=1)$ in the light of $\mathcal{H}(\tau)$ as well as $\mathbf{x}$? Certainly, the observed $\mathbf{x}$ should help you sharpen your prediction. Consequently, you are now called upon to assess $P(X=1;\mathbf{x},\mathcal{H})$.

One possibility would be to think hard about all that you have at hand, namely, $\mathbf{x}$ and $\mathcal{H}$, and then simply specify $P(X = 1; \mathbf{x}, \mathcal{H})$ as p^*, where $p^* \in (0, 1)$. Here p^* encapsulates your bet on the event $(X = 1)$ in the light of $\mathbf{x}$ and $\mathcal{H}$. If p^* happens to be identical to the p of Equation 2, then you are declaring the opinion that the data $\mathbf{x}$ has not had a sufficient impact on your beliefs for you to change your bet from your original p. From a philosophical point of view, there is nothing in the theory of subjective probability that stops you from specifying a p^* by introspection alone. However, from a computational point of view, it is efficient to proceed formally along the lines given below, because introspection to specify p^* subsequent to having specified p may lead to an inconsistency (technically *incoherence*). By incoherence, we mean a scenario involving a gamble in which "heads I win, tails you lose."

2.2 Bayes' Law: The Mathematics of Changing Your Mind

To address the scenario presented in Section 2.1, you start by pondering the matter of assessing your uncertainty about $(X = 1)$, in the light of $\mathcal{H}$, <u>were</u> you to know (but do not know) the disposition of $X_1, \ldots, X_n$; here $X_i = 1$, if the i-th item judged to be similar to the item in question has a life-length that exceeds t ($X_i = 0$, otherwise). That is, what would your $P(X = 1|X_1, \ldots, X_n, \mathcal{H})$ be? To address this question, you follow the same line of reasoning used to arrive upon Equation 4; that is, extend the conversation to θ, and obtain

$$\begin{aligned} P\left(X = 1|X_1, \ldots, X_n; \mathcal{H}\right) &= \int_0^1 P(X = 1|\theta, X_1, \ldots, X_n) \cdot \pi(\theta|X_1, \ldots, X_n; \mathcal{H}) d\theta, \\ &= \int_0^1 P(X = 1|\theta) \cdot \pi(\theta|X_1, \ldots, X_n; \mathcal{H}) d\theta, \\ &= \int_0^1 \theta \cdot \pi(\theta|X_1, \ldots, X_n; \mathcal{H}) d\theta. \end{aligned} \tag{7}$$

The second equality is a consequence of your judgment that X is independent of

$X_1, \ldots, X_n$, were you to know θ, and the third a consequence of choosing $P(X = 1|\theta) = \theta$ as a probability model for X. The quantity $\pi(\theta|X_1, \ldots, X_n, \mathcal{H})$ is the probability density at θ of your $P(\Theta \leq \theta|X_1, \ldots, X_n, \mathcal{H})$.

To obtain $\pi(\theta|X_1, \ldots, X_n, \mathcal{H})$ you invoke Bayes' Law; thus

$$\begin{aligned} \pi(\theta|X_1, \ldots, X_n; \mathcal{H}) &\propto P(X_1, \ldots, X_n|\theta; \mathcal{H}) \cdot \pi(\theta; \mathcal{H}) \\ &= \prod_{i=1}^{n} P(X_i = x_i|\theta) \cdot \pi(\theta; \mathcal{H}), \end{aligned} \tag{8}$$

by the multiplication rule, and by the independence of the X_i's from each other, were you to know θ, and with $x_i = 1$ or 0. For $P(X_i = x_i|\theta)$, you once again choose Bernoulli's model, so that $P(X_i = x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$.

With the above in place, you now have

$$\pi(\theta|X_1, \ldots, X_n; \mathcal{H}) \propto \prod_{i=1}^{n} \left\{\theta^{x_i}(1-\theta)^{1-x_i}\right\} \pi(\theta; \mathcal{H}). \tag{9}$$

Since $\pi(\theta; \mathcal{H})$ encapsulates your uncertainty about Θ in the light of $\mathcal{H}$ alone, and $\pi(\theta|X_1, \ldots, X_n; \mathcal{H})$ your uncertainty about it were you to be provided additional information via the $X_1, \ldots, X_n$, we say that Bayes' Law provides a mathematical prescription for changing your mind about the unobservable Θ. Once Equation 9 is at hand we may incorporate it in Equation 7 to write

$$P(X = 1|X_1, \ldots, X_n; \mathcal{H}) \propto \int_0^1 \theta \prod_{i=1}^{n} \left\{\theta^{x_i}(1-\theta)^{1-x_i}\right\} \pi(\theta; \mathcal{H}) d\theta, \tag{10}$$

as a prescription of how to change your mind about the event $(X = 1)$ itself.

2.3 Likelihood Function: The Weight of Evidence

There are two aspects of Equations 8 through 10 that need to be emphasized. The first is that the left-hand sides of these equations pertain to conditional events, namely the proposition that "were you to know the disposition of the X_i's, $i = 1, \ldots, n$" ; that is,

supposing you were provided with the realizations of each X_i. The second feature is that they inform the reader as to how you express your uncertainties (or bets) about Θ and X respectively, once the X_i's reveal themselves as x_i. Implicit to this bet is your particular choice of probability models $P(X = x|\theta)$ and $P(X_i = x_i|\theta),\ i = 1, \ldots, n$.

In actuality, however, the X_i's have indeed revealed themselves in the form of data, as $\mathbf{x} = (x_1, \ldots, x_n)$, where each x_i is known to you as being one or zero. In view of this, the left-hand sides of Equations 8 through 10 should be re-written as $\pi(\theta; x_1, \ldots, x_n, \mathcal{H})$ and $P(X = 1; x_1, \ldots, x_n, \mathcal{H})$ respectively. But more significant is the fact that the quantity $P(X_i = x_i|\theta)$ of Equation 8 can no longer be interpreted as a probability. This is because the notion of probability is germane only for events that have yet to occur, or for events that have occurred but whose disposition is not known to you. In our case, X_i is known to you as $x_i = 1$ or $x_i = 0$, thus $P(X_i = x_i|\theta)$ is <u>not</u> a probability. So what does the quantity $P(X_i = x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, with x_i fixed as 1 or 0, and θ unknown, mean? Similarly, in the context of Equation 9 with $r = \sum_{i=1}^{n} x_i$, what does the quantity

$$\prod_{i=1}^{n} \left\{\theta^{x_i}(1-\theta)^{1-x_i}\right\} = \theta^r(1-\theta)^{n-r}, \tag{11}$$

with n and r known, but θ unknown, mean? Note that r is the total number of successes.

As a function of θ, with n and r fixed, the quantity $\theta^r(1-\theta)^{n-r}$ is called the *likelihood function* of θ; it is denoted, $\mathcal{L}_{\mathcal{Y}}(\theta; n, r)$, the subscript, which will henceforth be dropped, signaling the fact that like probability, the likelihood function is also personal. Since $\mathcal{L}(\theta; n, r)$ is not a probability, the likelihood function, even though it is derived from a probability model, is not a probability. It can be viewed as a function that assigns weights to the different values θ that Θ can take, in the light of the known n and r; these latter quantities can be viewed as *evidence*. Thus the likelihood function can be interpreted as a function that prescribes the weight of evidence provided by the data for the different values that chance Θ can take. For example, with $n = r = 1$, $\mathcal{L}(\theta; n = r = 1) = \theta$; this suggests—see Figure 1—that with $n = r = 1$, more weight is given by the likelihood

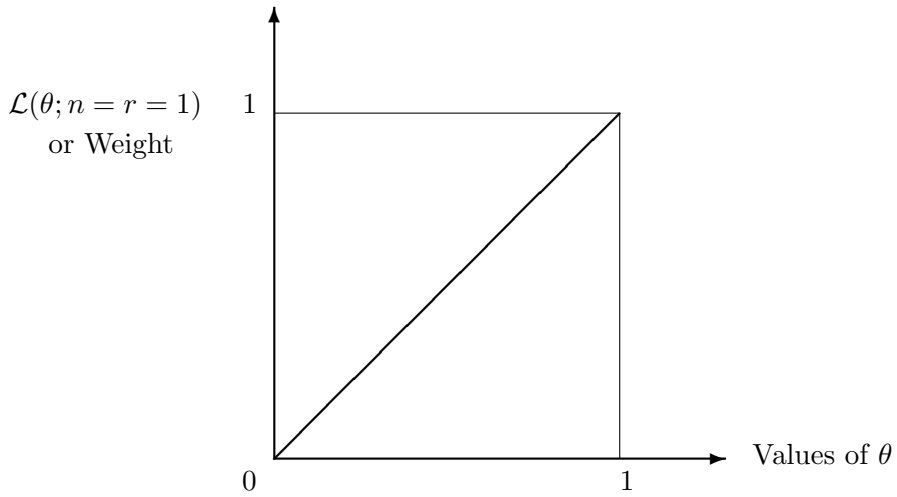

Figure 1: The likelihood function with $n = r = 1$

function to the large values of θ than to the smaller values.

To summarize, the expression $P(X_i = x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, specifies a probability of the event $(X_i = x_i)$ when X_i is unknown, and θ is assumed known; whereas with X_i known as x_i, it specifies a likelihood for the unknown θ. With $\mathbf{x}$ known, Equation 10 when correctly written becomes

$$P(X = 1; \mathbf{x}, \mathcal{H}) \propto \int_0^1 \theta(\theta^r(1-\theta)^{n-r}) \cdot \pi(\theta; \mathcal{H}) d\theta. \tag{12}$$

Equation 12 is interesting. It encapsulates, as we read from left to right, the four notions we have introduced thus far: personal probability (the left-hand side); chance (the parameter θ); the likelihood of chance (the quantity $\theta^r(1-\theta)^{n-r}$); and the probability of chance (the quantity $\pi(\theta; \mathcal{H})$).

Note also that the right-hand side of Equation 12 is the expected value of a function of Θ, namely, the function $\Theta^{r+1}(1-\Theta)^{n-r}$. Thus we may say that the effect of the data $\mathbf{x}$ is to change your bet on the event $(X = 1)$ from $E_{\mathcal{Y}}(\Theta)$ to $E_{\mathcal{Y}}(\Theta^{r+1}(1-\Theta)^{n-r})$.

3 Imprecise Surrogates: Motivation for Vagueness & Belief

In Section 1 we outlined a problem that is the focus of our discussion, and in Section 2 we added a feature to it by bringing in the role of data. The notions used in Sections 1 and 2 are probability, chance, and likelihood. Are these the only ones needed to address all problems pertaining to uncertainty? Are there circumstances that pose a challenge to us in terms of being able to lean on these notions alone? If so, what are these, and under what scenarios do we need to go beyond what has been introduced and discussed? The purpose of this section is to address the above and related questions. But first we bring into play our second caveat and explore the circumstances under which the notions of probability, chance, and likelihood will suffice to address this caveat. The caveat in question pertains to the presence or not of detectable anomalies during inspection, quality control, and other diagnostic testing functions.

3.1 Anomalies: A Surrogate of Failure

To keep our discussion simple, suppose that in order to assess your uncertainty about the event $(X = 1)$, you have at your disposal $\mathcal{H}$ and also a knowledge of the presence or the absence of a detectable anomaly. An anomaly could be a visible defect, or noticeable damage, or some other suitable indicator of imperfection. Anomalies could be present and yet not be detected. We denote the presence of a detected anomaly by letting a binary variable Y take the value 1; the absence of a detectable anomaly by letting $Y = 0$. The presence of an anomaly does not necessarily imply that X will be zero; similarly, its absence is no assurance (to you) that X will be one; see Figure 2. Rather, like the $X_1, \ldots, X_n$ of Section 2, the presence or absence of a detectable anomaly helps you sharpen your assessment of the uncertainty about $(X = 1)$.

Suppose then, that $Y = y$ has been observed, with $y = 1$ or 0, and that you are required to assess $P(X = 1; y, \mathcal{H})$. A simple way to proceed would be to treat y as a part of $\mathcal{H}$,

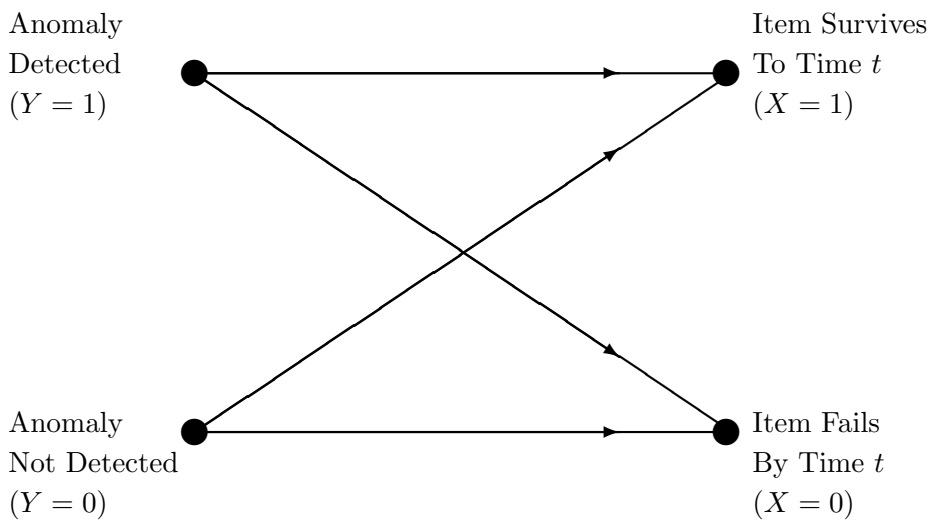

Figure 2: Effect of anomalies on survival

and upon careful introspection specify

$$P(X = 1; y, \mathcal{H}) = \widehat{p}, \quad 0 < \widehat{p} < 1,$$

as your bet on the event $(X = 1)$. The $\widehat{p}$ above is like the p of Section 1, in the sense that if $\widehat{p} = p$, then y has had no effect on your disposition about $(X = 1)$. There is, of course a more systematic way to incorporate the effect of y into your analysis, and this involves a use of the likelihood. To see how, start by pondering the matter of assessing your uncertainty about the event $(X = 1)$, in the light of $\mathcal{H}$, <u>were</u> you to know (but do not know) the disposition of Y. This is what was also done in Section 2.2. That is, you ask yourself what $P(X = 1|Y; \mathcal{H})$ should be? By Bayes' Law

$$P(X = 1|Y; \mathcal{H}) \propto P(Y = y|X = 1; \mathcal{H}) \cdot P(X = 1; \mathcal{H}),$$

$y = 1$ and 0. For $P(X = 1; \mathcal{H})$ you may use your p of Equation 2. To proceed further, you need to specify a probability model for Y, conditional on $(X = 1)$. That is, you need to specify $P(Y = 1|X = 1; \mathcal{H})$ and $P(Y = 0|X = 1; \mathcal{H})$; this is tantamount to specifying

a joint distribution for X and Y. Once this can be done, you have

$$P(X = 1|Y; \mathcal{H}) \propto P(Y = y|X = 1; \mathcal{H}) \cdot p. \tag{13}$$

However, in actuality, Y has been observed as $y = 1$ or $y = 0$. Consequently, Equation 13 becomes

$$P(X = 1; y, \mathcal{H}) \propto \mathcal{L}(X = 1; y, \mathcal{H}) \cdot p, \tag{14}$$

where $\mathcal{L}(X = 1; y, \mathcal{H})$ is your likelihood function for the unknown event $(X = 1)$ in the light of the evidence y and $\mathcal{H}$. The probability model $P(Y = y|X = 1; \mathcal{H})$ helps you specify the likelihood. Equation 14 says that your bet on the event $(X = 1)$ in the light of y and $\mathcal{H}$, is proportional to your bet on $(X = 1)$ based on $\mathcal{H}$ alone, multiplied by your likelihood. The approach prescribed above is more systematic than the one involving the specification of $\widehat{p}$ based on introspection alone, because it incorporates the p of Equation 2. A key point to note is that $\mathcal{L}(X = 1; y, \mathcal{H})$ is the likelihood of an observable event; it is not the likelihood of chance Θ discussed in Section 2.3. Should you prefer to work with the likelihood of chance, then you must introduce chance into your pondering. To do so, you may proceed as follows:

$$P(X = 1|Y; \mathcal{H}) = \int_0^1 P(X = 1|\theta, Y; \mathcal{H}) \cdot \pi(\theta|Y; \mathcal{H}) d\theta,$$

which extends the conversation to θ, as was done to arrive at Equation 3. If you now assume that $(X = 1)$ is independent of both Y and $\mathcal{H}$, were you to know θ, and assume Bernoulli's model, then

$$P(X = 1|Y; \mathcal{H}) = \int_0^1 \theta \cdot \pi(\theta|Y; \mathcal{H}) d\theta. \tag{15}$$

But by Bayes' law

$$\pi(\theta|Y; \mathcal{H}) \propto P(Y = y|\theta; \mathcal{H}) \cdot \pi(\theta; \mathcal{H}). \tag{16}$$

Consequently, to proceed further, you need to specify a probability model for the anomaly Y, were you to know θ, and also $\pi(\theta; \mathcal{H})$, an entity that has already appeared in Sections 1

and 2. Since Y has in actuality been observed (as $y = 1$ or $y = 0$), Equation 16 becomes

$$\pi(\theta; y, \mathcal{H}) \propto \mathcal{L}(\theta; y, \mathcal{H}) \cdot \pi(\theta; \mathcal{H}),$$

where $\mathcal{L}(\theta; y, \mathcal{H})$ is the likelihood function of the chance Θ, in the light of $\mathcal{H}$ and evidence about the anomaly y. With the above in place Equation 15 becomes

$$P(X = 1; y, \mathcal{H}) \propto \int_0^1 \theta \cdot \mathcal{L}(\theta; y, \mathcal{H}) \cdot \pi(\theta; \mathcal{H}) d\theta.$$

To compare the above equation with Equation 14 (their left hand sides are the same), we note that since $p = E(\Theta)$, Equation 14 may also be written as

$$P(X = 1; y, \mathcal{H}) \propto \int_0^1 \theta \cdot \mathcal{L}(X = 1; y, \mathcal{H}) \cdot \pi(\theta; \mathcal{H}) d\theta.$$

The last two equations signal the fact that in order to incorporate the effect of the detected anomalies into the assessment of your uncertainty about $(X = 1)$, you should be prepared to either specify the likelihood of $(X = 1)$ in the light of y (and $\mathcal{H}$), or the likelihood of θ in the light of y (and $\mathcal{H}$), whichever is more convenient. To specify these likelihoods, you may want to specify $P(Y = y|X = 1; \mathcal{H})$ or $P(Y = y|\theta; \mathcal{H})$, probability models for Y, were you to know X or θ, respectively. Of these, the former may be easier to assess than the latter, since it is based only on observables. We shall therefore focus on the case $P(Y = y|X; \mathcal{H})$, and refer to it as a *postmortem probability model.*

3.2 Eliciting Postmortem Probabilities: Potential Obstacles

The material of Sections 1 and 2 required of you the specification of $P(X = x|\theta)$ and $\pi(\theta; \mathcal{H})$, for $x = 1$ or 0. For the former, Bernoulli's model is a natural choice; for the latter, a beta density with parameters α and β is a choice with much flexibility. Thus, for $0 < \theta < 1$

$$P(X = x|\theta) = \theta^x (1 - \theta)^{1-x},$$

and

$$\pi(\theta; \mathcal{H}) = \pi(\theta; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}.$$

Coming to the scenario of Section 3, you are required to specify the above, and also a model for the postmortem probability $P(Y = y|X = x; \mathcal{H})$, for x, $y = 1$ or 0. The latter could pose two difficulties. The first is that you should be able to probabilistically relate detectable anomalies and failure; Figure 2 with the direction of the arrows reversed could provide guidance. The second—a bigger problem—can arise because of the fact that the absence or the presence of any trait which qualifies as an anomaly has may not be easily determined. For example, both a surface scratch and a dent qualify as defects, but the former could be less deleterious to an item's survival than the latter. Also, at what point does a rough scratch get labelled as a dent? The classification of an anomaly is therefore not crisp, so that the event "anomaly" is not well defined. It is this lack of crispness that motivates a consideration of "vagueness" as another aspect of uncertainty quantification; more on this will be said in Section 4.

One manifestation of this absence of crispness is that responses to questions for eliciting postmortem probabilities tend to be unhelpful. The following two responses from an actual scenario are illustrative:

i) "If the unit works, there is a less than 20% chance that we would have detected an anomaly. If it does not, we would be seeing something 20-40% of the time."

ii) "If it works, that means that it was well manufactured. If it does not, then it means that it was handled poorly when it was shipped."

Clearly, pinning down postmortem probabilities from statements like *i*) and *ii*) above is not possible. At best *i*) can provide bounds on the postmortem probabilities, and *ii*) has no probabilistic content whatsoever. Yet *i*) and *ii*) provide information, albeit not in the form required by the calculus of probability.

To summarize, as long as the event "anomaly" is well defined so that one is able to precisely specify the postmortem probabilities, the development of Section 3.1 can be used, and to do so all that one needs are the notions of probability, chance, and likelihood. Once difficulties of the type discussed above come into play, postmortem probabilities cannot be elicited. When such is the case, the notions of "vagueness" and "belief" enter the arena of uncertainty quantification. We emphasize that we do not see these notions as a prelude to supplanting probability; rather, they enhance probability by making its use more encompassing. However, to some, like Zadeh (1978), the notion of vagueness invites alternatives to probability, a matter upon which we disagree.

4 Harnessing Vagueness: Uncertainty Quantification Under Imprecision

What do we mean by the term "vagueness"? Is it synonymous with the term "imprecision"? How do vagueness and imprecision enter the arena of uncertainty quantification? These are some of the questions that we aim to address in this section. We shall use the scenario of anomalies discussed in Section 3 as a point of discussion.

4.1 Fuzzy Sets and the Uncertainty of Classification

As a preamble, recall that in Section 3.1, Y was a binary variable taking values $y = 0$ or $y = 1$, with $Y = 0(1)$ denoting the absence (presence) of a detectable anomaly. Declaring that $Y = 0$ or 1 is often a judgment call, which does not encapsulate the degree of the anomaly. In this section we refine the above process by introducing some granularity to the values y that Y can take. To do so, we let Y denote some undesirable characteristic of the item in question that can be quantified—for instance the depth of a scratch—and allow Y to take a continuum of values y in some well defined range, say $\mathcal{R} = [0, M]$, where M is specified. Let $\widetilde{A}$, a subset of $\mathcal{R}$, be the set of all y's that lead to the assessment that the item in question has an anomaly. Now if there exists a value y^* such that for any

$y \geq y^*$ an anomaly is declared, then $\widetilde{A}$ is called a *crisp* (or a sharp) set; crisp to reflect the fact that $\widetilde{A}$ has well defined boundaries. Consequently, any y can be placed with precision in the set $\widetilde{A}$, or its complement. Crisp sets are said to adhere to the *law of the excluded middle*, in the sense that any y either does belong or does not belong to $\widetilde{A}$. However, if it is not possible to identify a y^* of the kind described above, then a boundary of $\widetilde{A}$ is not well defined. Consequently, we are unable to classify the membership of certain y's in $\widetilde{A}$ with definitiveness (or precision). Such y's can simultaneously belong and not belong to $\widetilde{A}$. Sets which exhibit the property of having boundaries that are not sharp are said to be *fuzzy*. Fuzzy sets do not adhere the law of the excluded middle. In the context of the scenario considered here, one may not be able to classify, with definiteness, certain defects as being anomalies. That is, there could arise, in practice, scenarios in which there is an uncertainty (in a subject matter specialist's mind) about classifying a defect as being an anomaly or not, and also an unwillingness (of the specialist) to assign probabilities to the uncertainty of classification.

To summarize, fuzzy sets are those whose boundaries are not well defined, and imprecision pertains to an inability to place with certainty every element of a set, such as $\mathcal{R}$, into its fuzzy subset such as $\widetilde{A}$. That is, imprecision is a consequence of vagueness.

The Kolmogorov axiomatization of probability is developed on the premise that probability measures be defined on sharp sets [cf. Billingsley (1985), p. 20]. Thus the appearance of fuzzy sets requires of us ways to develop approaches whereby probabilities can be endowed to fuzzy sets as well. A strategy for doing so is via the introduction of "membership functions" which, though not probabilistic, can be seen as a subject matter specialist's classification "probabilities." Membership functions are discussed in Section 4.2 and their use for inducing probabilities on fuzzy sets discussed in Section 4.3. As a final reminder, it is important to keep in mind that the material of Sections 4.2 and 4.3 will not come into play if the event "anomaly" can be well defined.

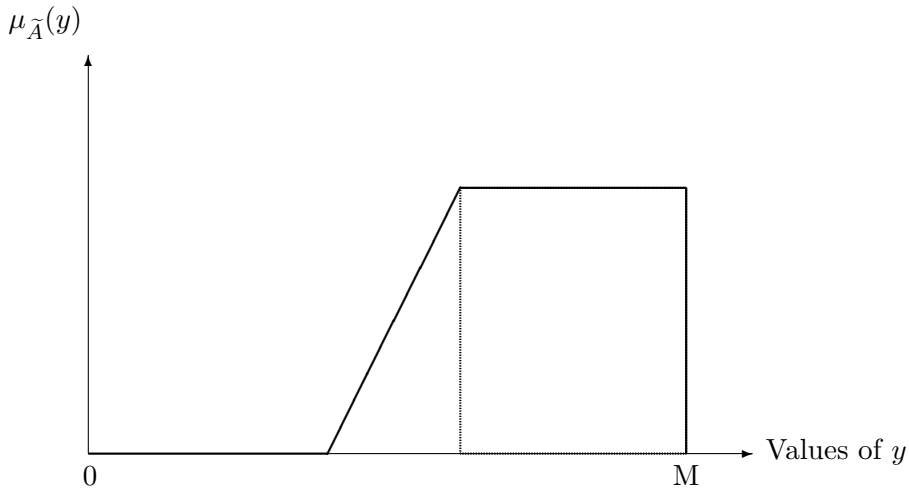

Figure 3: Membership function of a fuzzy set $\widetilde{A}$

4.2 The Membership Function of a Fuzzy Set

The *membership function* of a fuzzy set $\widetilde{A}$ encapsulates the degree to which any $y \in \mathcal{R}$ belongs to $\widetilde{A}$. It is denoted by $\mu_{\widetilde{A}}(y)$, for every y. It is important to note that $\mu_{\widetilde{A}}(y)$ is not a probability, because $\sum_y \mu_{\widetilde{A}}(y)$ need not be one; however, it is often the case that $0 \leq \mu_{\widetilde{A}}(y) \leq 1$, for all y. Operations with fuzzy sets, such as unions, intersections and complements are facilitated by the membership function. Like probability, the membership function is subjectively specified, and may change from person to person. The membership function of a crisp set is an identity function; i.e. if $\widetilde{A}$ is a crisp set, then $\mu_{\widetilde{A}}(y) = 0$ for $y < y^*$ and $\mu_{\widetilde{A}}(y) = 1$, otherwise. For the scenario of anomalies considered here, with y encapsulating the magnitude of a defect, $\mu_{\widetilde{A}}(y)$ would be of the form illustrated in Figure 3. Small values of y would certainly not be viewed as an anomaly and large values certainly would. For the intermediate values of y, $\mu_{\widetilde{A}}(y)$ shows the extent to which y would be judged (by one particular individual) to be an anomaly.

4.3 Endowing Probabilities to Fuzzy Sets

By endowing probabilities to fuzzy sets we mean assessing our personal probability that Y belongs to $\widetilde{A}$ in the light of the membership function $\mu_{\widetilde{A}}(y)$. For this we first need to assess our personal probability that Y reveals itself as y—that is our probability that the outcome of Y is y—and our personal probability that the revealed y belongs to $\widetilde{A}$. Supposing Y to take discrete values, we denote the above personal probabilities by $P_{\mathcal{Y}}(Y = y)$ and $P_{\mathcal{Y}}(y \in \widetilde{A})$ respectively. The need for this latter probability entails a philosophical argument with roots that can be traced to Laplace. By interpreting $\mu_{\widetilde{A}}(y)$ as a likelihood function and invoking Bayes' law, Singpurwalla and Booker (2003) argue that:

$$P_{\mathcal{Y}}(Y \in \widetilde{A}|\mu_{\widetilde{A}}(y)) = \sum_y \left[1 + \frac{1 - \mu_{\widetilde{A}}(y)}{\mu_{\widetilde{A}}(y)} \cdot \frac{P(y \notin \widetilde{A})}{P(y \in \widetilde{A})}\right]^{-1} P_{\mathcal{Y}}(Y = y). \tag{17}$$

4.4 Assessing Failure Probability with Imprecisely Specified Anomalies

With Equation 17 in place, it is a relatively straightforward matter to obtain an analogue of the postmortem probability when the classification of anomalies is imprecise, as

$$P(Y \in \widetilde{A}|X; \mu_{\widetilde{A}}(y)) = \sum_y \left[1 + \frac{1 - \mu_{\widetilde{A}}(y)}{\mu_{\widetilde{A}}(y)} \cdot \frac{P(y \notin \widetilde{A})}{P(y \in \widetilde{A})}\right]^{-1} P(Y = y|X), \tag{18}$$

where for convenience the subscripts associated with all the P's have been omitted. The key difference between Equations 17 and 18 is in the last term. The former entails an unconditional probability for Y; the latter, a conditional probability that Y reveals itself as y, given X, the disposition of an item's status—surviving or failed. Note that $P(Y = y|X)$ is like the postmortem probability of Section 3.1, save for the fact that Y can now take a range of values y, instead of it being 0 or 1.

To assess an item's survival probability were an imprecisely specified anomaly be declared as $Y \in \widetilde{A}$, we consider the analogue of Equation 13. Specifically, we have

$$P(X = 1|Y \in \widetilde{A}; \mathcal{H}) \propto P(Y \in \widetilde{A}|X = 1; \mu_{\widetilde{A}}(y)) \cdot p, \tag{19}$$

where the middle term is given by Equation 18, and as before, p is our prior probability that $(X = 1)$.

Equation 19 forms the basis of assessing the item's survival probability when the presence of an anomaly is actually declared, but not the extent of the defect that is believed to result in an anomaly. That is, we are not given the value of y. In this case $P(Y \in \widetilde{A}|X = 1; \mu_{\widetilde{A}}(y))$ is viewed as the likelihood and the left-hand side of Equation 19 becomes $P(X = 1; Y \in \widetilde{A}, \mathcal{H})$, the required probability. Consequently, Equation 19 leads us to

$$P(X = 1; Y \in \widetilde{A}, \mathcal{H}) \propto \mathcal{L}(X = 1; Y \in \widetilde{A}, \mu_{\widetilde{A}}(y)) \cdot p, \tag{20}$$

which is our personal probability that $(X = 1)$, given the presence of an anomaly that is vaguely specified.

5 A Reason to Believe

Sections 3 and 4 required of us the specification of a conditional probability $P(Y = y|X = x; \mathcal{H})$ and the membership function $\mu_{\widetilde{A}}(y),\ y \in [0, M]$, as a way of dealing with vagueness and anomalies. What if vagueness and other reasons create an unwillingness to specify the conditional probability but a willingness to specify a marginal probability $P(Y = y; \mathcal{H})$?

The notion of "belief" was introduced by Dempster (1967) as a way of dealing with such partial specifications. Dempster's development is articulated via a key feature of axiomatic probability theory, namely, that in order to induce probability measures from a probability measure space to another measure space it is necessary that the mapping from the former to the latter be a many-to-one map. As an example, a random variable is a many-to-one map. Consequently, its probability distribution function can be induced from the probability measure space on which the random variable is defined. When the mapping is a one to many map—as is the case with our anomaly (see Figure 2)—the induced measure will no more be a probability measure. For a more detailed appreciation

of this argument, we refer the reader to Appendix A. The induced measure not being a probability measure, alternate labels for it become germane. Dempster's choice of a label is *basic probability assignment*, abbreviated BPA.

With respect to the problem at hand, suppose that we are able to elicit personal probabilities of the type $P(Y = y; \mathcal{H})$, $y = 1$ or 0, as p_a and $(1 - p_a)$ respectively. Given p_a, and the mapping of Figure 2, how may we describe our uncertainty about the survival (or failure) of the item to time t? That is, how may we express our uncertainty about the event $(X = x)$ for $x = 1$ or 0?

The "belief function" approach of Dempster starts by noting that the mapping from $Y = y$ to $X = x$ is a one to many map. In particular, if Γ denotes the mapping from the Y-space to the X-space, then $\Gamma(Y = 1) = \{X = 1, X = 0\}$. That is, the singleton $(Y = 1)$ maps into the set $\{X = 1, X = 0\}$ via the map Γ; in other words, Γ is a *set-valued* map, similarly with $\Gamma(Y = 0)$. However, in order to make the essence of our development more transparent, we follow Wilson and Sentz (2003) and suppose that $\Gamma(Y = 0) = (X = 1)$. This means that the absence of an anomaly tantamount to the item's success. In other words, the mapping from $Y = 0$ to the X-space is a one-to-one map. Consequently, in Figure 2, the arc joining the nodes $(Y = 0)$ and $(X = 0)$ needs to be removed.

With the above in place, the next step in the development of the belief function approach is to induce measures of uncertainty from the Y-space to the X-space. Recall, that it is only the Y-space that has been endowed with probability as the measure of uncertainty. Since the X-space has only two elements, $(X = 1)$ and $(X = 0)$, $\mathcal{F}(X)$, the measure space (i.e., the set of all sets) generated by X, has four elements, namely,

$$\mathcal{F}(X) = \{\{\phi\}, \{X = 1\}, \{X = 0\}, \{X = 1, X = 0\}\}.$$

With $\Gamma(Y = 1) = \{X = 1, X = 0\}$ and $\Gamma(Y = 0) = (X = 1)$, the induced measure, say m, on $\mathcal{F}(X)$ will be of the form: $m(\phi) = 0,\ m(X = 1) = P(Y = 0) = 1 - p_a,\ m(X = 0) = 0$ and $m\{X = 1, X = 0\} = P(Y = 1) = p_a$. Recall that in Dempster's terminology, the

$m(\bullet)$'s constitute a BPA. It is easy to verify that m possesses the following two properties: $m(\phi) = 0$, and for $F \in \mathcal{F}(X)$, $\sum_{F \in \mathcal{F}(X)} m(F) = 1$. However, m is not countably additive and thus is not a probability measure. To make m a probability measure we should be prepared to apportion p_a between the events $(X = 1)$ and $(X = 0)$.

Once the BPA's are in place, the *belief function* induced by the map Γ on $\mathcal{F}(X)$ is defined , for any $F, G \in \mathcal{F}(X)$ as:

$$Bel(F) = \sum_{G \subseteq F} m(G),$$

and $Bel(F)$ is then considered as a quantified measure of uncertainty about F. Thus for our problem at hand $Bel(X = 1) = 1 - p_a$, whereas $Bel(X = 0) = 0$; also, $Bel\{X = 1, X = 0\} = 1 - p_a$.

Dempster has also introduced the dual of the belief function, called the *plausibility function*, where for any $F \in \mathcal{F}(X)$

$$Pl(F) = 1 - Bel(F^c);$$

F^c is the complement of F. For our problem at hand $Pl(X = 1) = 1$, whereas $Pl(X = 0) = p_a$.

To make these ideas operational, that is, to make a pragmatic use of them, we need to interpret $Bel(\bullet)$ and $Pl(\bullet)$. Using bets, $Bel(X = 1)$ is the most you are willing to pay for a bet on $(X = 1)$: if $Bel(X = 1) = 1 - p_a$, you are willing to pay at most $1 - p_a$ to receive one monetary unit if $(X = 1)$. $Pl(X = 1)$ is (1 - the most you are willing to pay for a bet on $(X = 1)^c$): if $Pl(X = 1) = 1$, you are not willing to pay anything to bet on $(X = 1)^c = (X = 0)$.

5.1 Summarizing "Beliefs"

By way of a closure, we claim that the notion of belief, or its dual plausibility, comes into play when joint probabilities of the type $P(y = 1, X = 1; \mathcal{H})$ cannot be elicited, and when

the marginal probabilities of the type $P(Y = 1; \mathcal{H}) = p_a$ cannot be apportioned in a one to many map. Intuitively, the uncertainty measure $Bel(\bullet)$ seems reasonable; it can be seen as a lower bound on probability. When the mapping under discussion is a one-to-one or a many-to-one, belief and probability agree, and thus the belief function will obey the rules of probability. We may conclude by saying that there is a price to be paid for not being able to elicit the required conditional probabilities, and the price is to forsake the notion of probability and its accompanying virtues. Dempster has also proposed rules for combining uncertainties, the details about which can be found in Shafer (1976) or in Wasserman (1990).

References

[1] Billingsley, P. (1985), *Probability and Measure*, 2nd Edition, John Wiley & Sons, Inc., New York.

[2] de Finetti, B. (1974), *Theory of Probability*, John Wiley, New York.

[3] Dempster, A. P. (1967), "Upper and Lower Probabilities Induced for a Multivariate Mapping", *Annals of Mathematical Statistics* **38**, 325-339.

[4] Good, I. J. (1965), *The Estimation of Probabilities*, The MIT Press, Cambridge, MA.

[5] Lindley, D. V. (1982), "Scoring Rules and the Inevitability of Probability", *International Statistical Review* **50**, 1-26.

[6] Lindley, D. V. and Phillips, L. D. (1976), "Inference for a Bernoulli Process (a Bayesian View)", *The American Statistician* **30**, 112-119.

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

[8] Singpurwalla, N. D. and Booker, J. M. (2003), "Membership Functions and Probability Measures of Fuzzy Sets", To appear in *Journal of the American Statistical Association*

[9] Wasserman, L. A. (1990), "Belief Functions and Statistical Inference", *The Canadian Journal of Statistics* **18**(3), 183-196.

[10] Wilson, A. and Sentz, K. (2003), "Information Integration Using Belief Functions", Submitted to *American Statistician.*

[11] Zadeh, L. A. (1978), "Fuzzy Sets as a Basis for a Theory of Possibility", *Fuzzy Sets and Systems* **1**, 3-28.

Appendix A

Belief and Plausibility

In order to gain an appreciation of the notion of "belief" and its dual "plausibility," it is best that we start off with a look at the essentials of measure theoretic probability. This we do below via the following seven steps, each of which serves as a prelude to the next step. We assume of the reader some familiarity with these steps. From Step 8 and onwards, our discussion highlights arguments necessary to motivate the notions of belief and plausibility.

1. Let $(\Omega, \mathcal{F}(\Omega), \mu)$ be a probability measure space, with ω as an element of Ω, and μ assessed for all members A of $\mathcal{F}(\Omega)$.

2. Let $(X, \mathcal{F}(X))$ be some measure space with x as an element of X. This is our space of interest.

3. Let $B \subset X$; since $\mathcal{F}(X)$ is a σ-field generated by X, $B \in \mathcal{F}(X)$.

4. Our aim is to endow the space $(X, \mathcal{F}(X))$ with a measure that encapsulates our uncertainty about any B, where $B \subset X$, or about a singleton x, where $x \in X$, should X have countable elements. Ideally, our measure of uncertainty should be probability.

5. The measure that we endeavor to endow $(X, \mathcal{F}(X))$ with, should bear some relationship to the measure μ. This is because we have been able to assess probabilities on the space $(\Omega, \mathcal{F}(\Omega))$; i.e. we are prepared to place bets only on members of $\mathcal{F}(\Omega)$.

6. In order to be able to do the above, we should connect the spaces $(\Omega, \mathcal{F}(\Omega), \mu)$ and $(X, \mathcal{F}(X))$. This connection can be made in several ways, <u>two</u> of which are indicated below:

i) A mapping from Ω as the domain, to X as the range, or

ii) A mapping from Ω as the domain, to $\mathcal{F}(X)$ as the range.

7. The standard approach is 6 *i*) above; this is what leads us to the notion of a *real valued random variable*, say Z.

   Specifically, we take X to be the real line $\mathbb{R}$, or a countably infinite set of integers $I = \{0, \pm 1, \pm 2, \ldots\}$, or a countably finite set of integers $I_N = \{0, \pm 1, \ldots, \pm N\}$. When $X = \mathbb{R}$, $\mathcal{F}(X) = \mathcal{B}(X)$—the Borel sets of $\mathbb{R}$. When $X = I_N$, then $\mathcal{F}(X)$ is the power set of I_N.

   Suppose that $X = \mathbb{R}$. Then Z is a mapping with domain Ω and range $\mathbb{R}$. Furthermore, Z is a *many-to-one* map from Ω to $\mathbb{R}$. Specifically, for every $\omega \in \Omega$, there is one and only one $Z(\omega)$, and $Z(\omega) \in \mathbb{R}$. However, we do allow for the possibility that for any two (or more) $\omega_1, \omega_2 \in \Omega$, $Z(\omega_1) = Z(\omega_2)$.

   Now, a (fortunate) consequence of the many to one map Z is that such a map is able to induce a probability measure, say μ^*, on $(X, \mathcal{F}(X))$ [or to put it more correctly on $(\mathbb{R} < \mathcal{F}(\mathbb{R}))$]. Specifically, for any $a \in \mathbb{R}$, the set $(Z(\omega) \leq a) \in \mathcal{F}(X)$, and

   $$\mu^*(Z(\omega) \leq a) = \mu\{\omega \in \Omega : Z(\omega) \leq a\}$$

   is a probability measure of the set $(Z(\omega) \leq a)$. Consequently, we now have a probability measure space $(X, \mathcal{F}(X), \mu^*)$ in addition to our original probability measure space $(\Omega, \mathcal{F}(\Omega), \mu)$.

   Thus with a many-to-one map, we are able to describe our uncertainties about events of interest in $\mathcal{F}(X)$ via a probability μ^*, with μ^* being based on μ.

8. Suppose now that the connection between the spaces $(\Omega, \mathcal{F}(\Omega), \mu)$ and $(X, \mathcal{F}(X))$ is established via a mapping Γ whose domain is Ω (as before) but whose range is $\mathcal{F}(X)$ instead of X. That is, $\Omega \xrightarrow{\Gamma} \mathcal{F}(X)$. More specifically, for every $\omega \in \Omega$, $\Gamma(\omega) = B$, where $B \in \mathcal{F}(X)$.

If we assume that the above mapping is many-to-one, in the sense that every $\omega \in \Omega$ gets mapped to one and only one set B (where B may or may not be a singleton), then this mapping is known as a many-to-one *set valued* map. When such is the case Γ is also able to induce a probability measure, say μ^{**}, on the space $(\mathcal{F}(X), \mathcal{F}(\mathcal{F}(X)), \mu^{**})$, where $\mathcal{F}(\mathcal{F}(X))$ is a σ-field of sets generated by $\mathcal{F}(X)$. Consequently, for any set $C \in \mathcal{F}(\mathcal{F}(X))$,

$$\mu^{**}(C) = \mu\{\omega \in \Omega : \Gamma(\omega) = C\}.$$

Thus to summarize, a many-to-one set valued map is also able to induce a probability measure μ^{**} on the space $(\mathcal{F}(X), \mathcal{F}(\mathcal{F}(X)))$, assuming that the latter space is of interest to us. But what about the space $(X, \mathcal{F}(X))$? This after all, is our space of interest.

9. The fact that Γ is a many-to-one set valued map on $\mathcal{F}(X)$ is tantamount to the fact that Γ is a *many-to-many* point valued map on X. In particular, if $X = \mathbb{R}$ and $\mathcal{F}(X) = \mathcal{B}(\mathbb{R})$, then Γ is a many-to-many real valued map on $\mathbb{R}$. Consequently, for every $\omega \in \Omega$, $\Gamma(\omega)$ can take any and <u>all</u> values in an interval, say $\mathcal{I}$, where $\mathcal{I} \in \mathcal{B}(\mathbb{R})$. Inducing a probability measure on $\mathcal{I}$ or any subset of $\mathcal{I}$ boils down to <u>smearing</u> $\mu(\omega)$, the probability measure on ω, over $\mathcal{I}$. How should this measure be smeared? What if one is unwilling to specify a strategy for smearing (or distributing) $\mu(\omega)$ over $\mathcal{I}$? When such is the case we are unable to induce a probability measure form the space $(\Omega, \mathcal{F}(\Omega), \mu)$ to $(X, \mathcal{F}(X))$. As a consequence, an alternative measure called *plausibility*, abbreviated $Pl(\bullet)$, has been proposed on $\mathcal{F}(X)$. But before examining $Pl(\bullet)$, it may be useful to better articulate this matter of smearing $\mu(\omega)$ by looking at a special case of $\mathcal{I}$, namely an $\mathcal{I}$ consisting of a countable number of elements, say two; denote these by $\{x_1, x_2\}$. Suppose that $\Gamma^{-1}\{x_1, x_2\} = \omega$; then $\mu(\omega)$ is the induced probability measure of $\{x_1, x_2\}$. However, to induce a probability measure on x_1 or x_2, we need to split (apportion) $\mu(\omega)$ in some logical and meaningful manner.

To summarize, whenever the map connecting two measure spaces is a many-to-one set valued, or a many-to-one point valued map, a probability measure can always be induced from the domain space to the range space. Probability measures cannot be induced when the mapping is a one-to-many, or a many-to-many, point valued map, unless additional assumptions are made. When such assumptions cannot be made, a compromise has to be struck and upper and lower probabilities enter the foray of uncertainty assessment. These are discussed below.

10. Consider the subset B of X. Suppose that there does <u>not</u> exist an induced probability measure from $(\Omega, \mathcal{F}(\Omega), \mu)$ to B. That is, $\nexists$ an $\omega \in \Omega$, such that $\Gamma(\omega) = B$.

    Now consider a set $C \in \mathcal{F}(\mathcal{F}(X))$ with the feature that $C \cap B \neq \phi$; suppose that C is the only set in $\mathcal{F}(\mathcal{F}(X))$ that intersects with B. Since $C \in \mathcal{F}(\mathcal{F}(X))$, $\mu^{**}(C)$ is known. Let $\omega_1, \omega_2, \ldots, \omega_n$ be such that $\Gamma(\omega_i) = C,\ i = 1, \ldots, n$. Then, the plausibility if B, denoted $Pl(B)$ is the (probability) measure $Pl(B) = \mu\{\omega_1, \ldots, \omega_n\}$. Alternatively put,

    $$Pl(B) = \mu\{\omega \in \Omega; \Gamma(\omega) = C \text{ and } B \cap C \neq \phi\}.$$

    The above expression generalizes when more than one set intersects B. For example, suppose that $B \cap C_i \neq \phi$, for $i = 1, \ldots, k$, with $C_i \in \mathcal{F}(\mathcal{F}(X))$. Then

    $$Pl(B) = \mu\{\omega \in \Omega; \Gamma(\omega) = C_i \text{ and } B \cap C_i \neq \phi,\ i = 1, \ldots, k\}.$$

    Since there are several sets C_i that intersect with B, there are overlapping ω's in the definition of $Pl(B)$. Consequently, it is also called an "upper probability".

11. A notion dual to $Pl(\bullet)$—in a sense to be explained later—is $Bel(\bullet)$; here

    $$Bel(B) = \mu\{\omega \in \Omega; \Gamma(\omega) = C_i, C_i \subset B,\ i = 1, \ldots, k\}.$$

    $Bel(B)$ is a lower probability, with $0 \leq Bel(B) \leq Pl(B) \leq 1$. Also, $Bel(B) = 1 - Pl(B^c)$.

The measures $Pl(\bullet)$ and $Bel(\bullet)$ are not probability measures in the sense that

$$Bel(A \cup B) \geq Bel(A) + Bel(B);$$

i.e. because of an overlap of ω's, $Bel(\bullet)$ is super-additive.