

Chapter 1

Introduction

Statistics is often defined as the science of collecting, summarizing, and analyzing data. Our focus in this body of work will be on the summarization aspect, the reduction of large amounts of data into a single number or set of numbers. These numbers are also called statistics. How can we choose an appropriate statistic? How do we know how well the statistic summarizes the data set?

A statistic that summarizes an entire data set without losing any information about the family of distributions or the model is often called a sufficient statistic. Generally we like to use the statistic that provides the most information about the model. Sometimes there are several sufficient statistics for one model. At other times the only sufficient statistic is the entire data set. A very large data set can be rather unwieldy to work with. In this case, can we use a statistic that, though not sufficient, does summarize the data set somewhat? How much information will we lose? How can we compare two statistics that aren't sufficient in terms of the amount of information they provide?

A statistical experiment consists of taking a sample and forming a statistic based on the sample data. Two different experiments can yield two different statistics. Under certain conditions, we can compare these statistics by comparing the experiments. Though there is literature on the theory of comparison of experiments, the theory is set in a quite rigid format. There is also a great deal of literature involving information theory. Much of this theory requires that certain assumptions be made about distributions.

In this paper we will present a new method for comparing experiments. No assumptions need to be made and no conditions need to be met in order for this new method to measure the

amount of information contained in almost any statistic. We only need to know the distribution of the statistic under study. Indeed, this new theory can be applied where many of the existing methods often fail to be useful. It will allow us to open a new door in the fields of information theory and comparison of experiments.

We will begin our study with a literature review of existing methods of measuring information in statistics and comparing experiments. The shortcomings of each of these methods is discussed. Next we will present the new approach to comparing experiments, with preliminary definitions, theoretical results, and comparisons to old methods. Examples will be given where existing methods either produce undesirable results, or produce no results at all. Then we will introduce a method to self-calibrate our measure of information. This will give us a meaningful scale against which to compare other statistics. Last, we will apply our new method to Casino Blackjack, where we will be trying to discriminate between a full fifty-two card deck, and a fifty-two card deck with one card missing. This is an area where we will gain new insight about information in card-counting systems.

Chapter 2

Literature Review

Literature addressing “information” abounds in every discipline. In the statistical literature there are numerous approaches to defining and quantifying “amount of information”. We will review several of these approaches as well as the shortcomings of each.

2.1 Fisher's Information

R. A. Fisher (1925, pp. 700-725) introduced a measure that summarizes the amount of information an observed sample contains about the parameter space. Suppose an independent, identically distributed sample is observed, $X_n = (x_1, x_2, \dots, x_n)$, and $f(x_i, \theta)$ is the common density function of the X_n 's, then one way of expressing this information measure for one dimensional space is:

$$I(f, \theta) = E_{\theta} \left\{ \left(\frac{\partial}{\partial \theta} \sum_{i=1}^n \log f(x_i; \theta) \right)^2 \right\}$$

We must assume that certain regularity conditions are met. For r -dimensional parameter space the measure can be expressed as the Fisher information matrix:

$$I(f, \theta) = \left\| E_{\theta} \left\{ \frac{\partial}{\partial \theta_k} \sum_{i=1}^n \log f(x_i; \theta) \cdot \frac{\partial}{\partial \theta_l} \sum_{i=1}^n \log f(x_i; \theta) \right\} \right\|_{k, l = 1, \dots, r}$$

If two experiments, E and F , have associated density functions f and g , respectively, one should be able to compare the amount of information in each experiment by comparing $I(f, \theta)$ to $I(g, \theta)$. However, in order to compare two experiments or statistics, we must assume that certain regularity conditions are satisfied. One of these conditions (Condition A) is of particular

concern because it will not be met in the blackjack problem. These regularity conditions follow (Rohatgi (1976, p. 361)). Let q be an open interval of the real line, and $\{f_\theta : \theta \in \Theta\}$ be a family of probability density functions (pdf's) or probability mass functions (pmf's). Assume that the set $\{f_q(x) = 0 \text{ for every } q \in \Theta\}$ is independent of q . Suppose that:

CONDITION A: $\frac{\partial}{\partial q} f_q(x)$ exists and is finite for every q .

CONDITION B:
$$\begin{cases} \frac{\partial}{\partial \theta} \int f_\theta(x) dx = \int \frac{\partial}{\partial \theta} f_\theta(x) dx = 0 \text{ if } f_\theta \text{ is a pdf, for every } \theta \in \Theta \\ \frac{\partial}{\partial \theta} \sum_x f_\theta(x) = \sum_x \frac{\partial}{\partial \theta} f_\theta(x) = 0 \text{ if } f_\theta \text{ is a pmf, for every } \theta \in \Theta. \end{cases}$$

Let y be defined on Θ and be differentiable there, and let T be an unbiased estimate of y such that $E_\theta T^2 < \infty$ for all q . Assume that

CONDITION C:
$$\begin{cases} \frac{\partial}{\partial \theta} \int T(x) f_\theta(x) dx = \int T(x) \frac{\partial}{\partial \theta} f_\theta(x) dx = 0 \text{ if } f_\theta \text{ is a pdf, for every } \theta \in \Theta \\ \frac{\partial}{\partial \theta} \sum_x T(x) f_\theta(x) = \sum_x T(x) \frac{\partial}{\partial \theta} f_\theta(x) = 0 \text{ if } f_\theta \text{ is a pmf, for every } \theta \in \Theta. \end{cases}$$

These conditions can be adapted to the r -dimensional case. Of course, these regularity conditions can be restrictive. There will often be times when they are not met. Other drawbacks are discussed in section 3.4.

2.2 Weight of Evidence and Shannon's Entropy

Suppose we wish to discriminate between the pair (q_1, q_2) in Θ . Perhaps P_{θ_1} is the distribution under hypothesis H and P_{θ_2} is the distribution under hypothesis H'. Let $f(X; q_1)$ be the density function for the random variable X under H and $f(X; q_2)$ be the density function for the random variable X under H'. Then Good (1950, p. 63) describes the weight of evidence for H given X as the log of the ratio of the likelihoods of H and H'. More formally,

$$W(H:X) = \log \frac{f(X; \theta_1)}{f(X; \theta_2)}.$$

Shannon (1948, p. 19) introduced a related measure of the amount of information without regard to any hypotheses. Shannon proposed a measure, G, of the expected uncertainty before an observation is taken or the expected gain in information from a single observation:

$$G = - \sum_{i=1}^{\infty} p_i \log p_i \quad \text{where } p_i = P(X = x_i) \quad i=1,2,\dots$$

This is often referred to as entropy. Good (1950, p. 75) notes that the difference of the entropies assuming H' and H, respectively, is equal to the expected weight of evidence. We will continue a discussion of the expected weight of evidence in section 2.4.

In reference to entropy there is some difficulty in defining whose uncertainty or information is being measured (Denbigh and Denbigh (1985, p. 104)). For continuous random variables entropy is either infinite or undefined (Good and Osteyee, 1974).

2.3 Comparison of Risk Functions

Blackwell (1951, pp. 93-102) summarized several ideas introduced in unpublished notes by Bohnenblust, Shapley and Sherman in which he compared two experiments by comparing their risk functions. He defined one experiment, E , as more informative than another, F , if for every risk function, any risk attainable with F is also attainable with E . Blackwell also develops the idea of one experiment being sufficient for another experiment. Sufficiency in this sense is a special case of being “more informative than”. Notice that this only partially orders the experiments in terms of the amount of information each provides.

2.4 Kullback Information Function

A measure of information more closely related to the proposed problem is the expected weight of evidence or the Kullback information function (Good (1950, pp. 72-73), Kullback and Leibler (1951, p. 80)) for discriminating between the pair (q_1, q_2) in Θ . It is given by:

$$I(q_1, q_2) = E_{\theta_1} \left(\log \frac{f(X; \theta_1)}{f(X; \theta_2)} \right)$$

where $f(X; q_i)$ is the density function for the random variable X . This could be evaluated for each statistic and for each q_i and the results compared to determine which statistic discriminates the best between distributions. Kullback (1968) continues the study of the relationship between information and sufficiency.

One assumption that must be made when using the Kullback information function is that the probability measures P_{q_1} and P_{q_2} are absolutely continuous with respect to one another. When this assumption is not met the Kullback information function may be equal to $\pm\infty$.

DEFINITION 2.4.1: Probability measures P_{q_1} and P_{q_2} are absolutely continuous with respect to one another if there exists no set $A \in \mathcal{A}$ (see definition 2.6.1) such that $P_{q_1}(A) = 0$ and $P_{q_2}(A) \neq 0$ or $P_{q_1}(A) \neq 0$ and $P_{q_2}(A) = 0$.

If P_{q_1} and P_{q_2} are absolutely continuous with respect to one another we will say that they have the same supports.

DEFINITION 2.4.2: A support for P_{q_i} consists of any set A such that $P_{q_i}(A) = 1$.

However, in problems where these assumptions are not met, the Kullback information function is not a useful measure of information. It is, in fact, equal to $\pm\infty$. For instance, if $f(X; q_1) = .5$ and $f(X; q_2) = 0$ then $I(q_1, q_2) = \infty$ and $I(q_2, q_1) = -\infty$. Results involving singular distributions are discussed by Good and Osteyee (1974). Some statisticians feel that a problem like this is not worth considering. Kullback (1968, p. 3) states, "Since there is no essential problem in the rejection of statistical hypotheses that may have been possible prior to the observations but are impossible after the observations, our mathematical assumption is to exclude this contingency." Savage (1954, p. 127) says, "definitive observations do not play an important part in statistical theory, precisely because statistics is mainly concerned with uncertainty, and there is no uncertainty once an observation definitive for the context at hand has been made." Certainly in our blackjack problem, (if we are trying to distinguish between a full fifty-two card deck and a deck with one ace missing), if we observe four aces dealt then we know definitely that the deck was composed of fifty-two cards, but if we observe three or fewer aces, we are still left with the problem of discriminating between the two possible deck compositions. If we followed the opinions of Kullback (1968, p. 3) and Savage (1954, p. 127), we would never even try to measure information in card-counting statistics when we are trying to distinguish between different deck compositions.

To meet the problem concerning Kullback information. Dr. I. J. Good suggests that weight of evidence, $W(H:X)$, could be replaced by $\tanh(\lambda W(H:X))$ before expectations are taken, where λ depends on the application. This expression lies between -1 and 1 and thus avoids the problem of infinities. For small $W(H:X)$ this expression is proportional to $W(H:X)$.

2.5 Comparison of Power Functions

Lehmann (1986, p. 86) also ordered two experiments by comparing their power functions. Suppose two experiments, E and F , are available for testing a simple hypothesis, H , against a simple alternative, H' . If $\beta(\alpha)$ and $\beta'(\alpha)$ represent the power of the most-powerful α -level test based on E and F respectively, then E is more informative than F if $\beta(\alpha) \geq \beta'(\alpha)$ for all α . This technique only partially orders the experiments and Lehmann (1986, p. 89) points out that “...pairs of experiments permitting such a strong ordering are rare.”

2.6 Deficiency and Insufficiency

Le Cam (1964, pp. 1419-1455) presented a complicated formula for comparing two experiments based on the comparison of their risk functions. It is called the deficiency distance. In order to define it we must present several definitions. Let us review some properties of σ -fields (Billingsley (1986, pp. 17-18)):

DEFINITION 2.6.1: Let A be a collection of subsets of a set X . Then A is a field in X if and only if:

- (a) $X \in A$.
- (b) If $A \in A$ then $A^c \in A$.
- (c) If $A_1, A_2, \dots, A_n \in A$ then $\bigcup_{i=1}^n A_i \in A$.

If in addition:

$$(d) \text{ If } A_1, A_2, \dots \in A \text{ then } \bigcup_{i=1}^{\infty} A_i \in A,$$

then A is called a σ -field in X .

DEFINITION 2.6.2: Let A be a σ -field in X . A set $x \subseteq X$ is A -measurable if $x \in A$.

We also need to review the concept of a subfield. Bahadur (1954, p. 427-432) introduced the notion of a sufficient subfield which Le Cam used extensively in his work.

DEFINITION 2.6.3: A_1 is a subfield of A , written $A_1 \subseteq A$, if A_1 is a field and if every set that is A_1 -measurable is also A -measurable.

DEFINITION 2.6.4: Given X , A , and Θ , the subfield A_1 of A is sufficient for Θ if for each A -measurable set A , there exists an A_1 -measurable function f_A such that:

$$(a) \int_X f_A(x) dP_{q_i} \text{ exists and is finite.}$$

$$(b) \int_{A_1 \cap A} dP_{q_i} = \int_{A_1} f_A(x) dP_{q_i} \text{ for } A_1 \in A_1 \text{ and for all } q_i \in \Theta$$

Now let's define a statistical experiment. We will use this definition of an experiment throughout this dissertation (Torgersen (1991, p. 4)).

DEFINITION 2.6.5: The pair (X, A) is a measurable space if A is a σ -field in X .

DEFINITION 2.6.6: Define $E=(X, \mathcal{A} ; P_{q_i} : q_i \in \Theta)$ as a statistical experiment, where (X, \mathcal{A}) is a measurable space, Θ is a set and P_{q_i} is a probability measure for each $q_i \in \Theta$.

Le Cam's deficiency involves the use of some decision theory. We will state some basic definitions from Torgersen (1991, p. 62) and Lee (1989 pp. 218-219). Let's first define a decision space and a loss function:

DEFINITION 2.6.7: The decision space (T, \mathcal{L}) is a measurable space where T is a non-empty set of decisions available to the statistician, \mathcal{L} is a σ -field in T and the loss functions are extended real valued \mathcal{L} measurable functions on T .

DEFINITION 2.6.8: Let $(L_{q_i} : q_i \in \Theta)$ be a family of loss functions on the decision space. Then define loss function $L_{q_i}(t)$ as the loss that a statistician suffers if he takes decision $t \in T$ when $q_i \in \Theta$ is the true state of nature.

Decision rules and risk functions are also involved in measuring deficiency.

DEFINITION 2.6.9: A decision rule r is the randomization from the measurable space (X, \mathcal{A}) of experiment E to the decision space (T, \mathcal{L}) . This rule tells us, for each $x \in X$, which decision $t \in T$ should be chosen if x is observed. Decisions are chosen according to r and the chosen distribution is $P_{q_i} r$ which depends on $q_i \in \Theta$.

DEFINITION 2.6.10: Risk functions are expected losses. Thus the risk under q_i from using decision rule r is $E_{P_{q_i,r}} L_{q_i,r} = \int L_{q_i,r}(t)(P_{q_i,r})(dt)$ provided this integral exists.

Now with our arsenal of background information we can study Le Cam's measure of deficiency:

DEFINITION 2.6.11: The deficiency distance can be defined as follows: let $E=(X, A; P_{q_i}: q_i \in \Theta)$ and $F=(Y, B; Q_{q_i}: q_i \in \Theta)$ be two experiments defined on the same parameter set. E is ε -deficient with respect to F ($\varepsilon > 0$), if for every finite decision space (T, L) , for every bounded loss function $L_{q_i,r}(t)$ and for every decision rule S in F there exists a decision rule r in E such that for every $q_i \in \Theta$ the following inequality between risk functions is valid:

$$E_{P_{q_i,r}} L_{q_i,r} \leq E_{Q_{q_i,S}} L_{q_i,r} + \varepsilon_{q_i} \left\{ \sup |L_{q_i,r}| \right\}.$$

The deficiency $d(E, F)$ of E with respect to F is the smallest $\varepsilon_{q_i} > 0$ for which E is ε_{q_i} -deficient with respect to F . The deficiency distance, $\Delta(E, F)$, is the largest of the numbers $d(E, F)$ and $d(F, E)$.

Another quantity called insufficiency was introduced by Le Cam (1974, p. 630-649). Insufficiency measures the distance between an experiment E and some subexperiment $E' = (X, A'; P'_{q_i}: q_i \in \Theta)$ (where $A' \subset A$) and under certain regularity conditions is defined as follows:

DEFINITION 2.6.12: The insufficiency of E' with respect to E is:

$$h(E', E) = \inf \sup_{q_i \in Q} \|P_{q_i} - P'_{q_i}\|$$

where the infimum is taken over all families of measures $(P'_{q_i} : q_i \in \Theta)$ for which A' is sufficient and P_{q_i} and P'_{q_i} agree on A . The norm $\|P_{q_i} - P'_{q_i}\|$ is the total variation of $P_{q_i} - P'_{q_i}$.

Both the deficiency and insufficiency can be very difficult to calculate and at times only bounds can be given. This is discussed specifically in Mammen (1986, p. 666). A more tractable method would be quite desirable.

Deficiency in a sense measures the degree of sufficiency of a subfield and insufficiency measures the lack of sufficiency of a subfield. Torgersen (1991) gives a recent and thorough survey on the theory of the comparison of experiments. He particularly emphasizes Le Cam's work throughout the survey.

Chapter 3

A New Approach

Most of the existing methods for comparing experiments have limitations, either due to the assumptions that must be made or the difficult numerical calculations. A new method will be presented shortly. This method is effective even when the conditions are not met that are necessary for other measures of information to be useful. The calculations involved in this new method are not difficult given the present computing capabilities. And it will be possible not only to order our experiments by the amount of information each provides, but also to determine how much more or less information one experiment provides as compared to another.

3.1 Preliminary Definitions

As part of this approach to compare two or more experiments or statistics an experiment will be performed. Call this statistical experiment E as in Definition 2.6.6. So X is the set of possible outcomes, or the sample space. \mathcal{A} is a σ -field of subsets of X . Θ is the parameter set, or the set of possible values of q . Dichotomous parameter sets will be considered; i.e. $\Theta = (q_1, q_2)$. Any finite parameter set can be studied by restricting attention to two point subsets of the parameter set. P_{q_1} is the probability distribution of the outcome when q_1 is the true parameter. So $E = (X, \mathcal{A}; P_{q_i} : q_i \in \Theta)$ is a family of probability measures on the measurable space (X, \mathcal{A}) . Let set $A \subseteq \mathcal{A}$. The following definition comes from Billingsley (1986, p. 54).

DEFINITION 3.1.1: If a property holds for all $A \subseteq \mathcal{A}$ outside a set of measure 0 with respect to P_{q_i} then it will be said that the property holds almost everywhere, abbreviated a.e.

Now recall the definition of sufficiency in terms of parameters (Lehmann(1986,p.19)):

DEFINITION 3.1.2: We say a statistic $T(X)$ is sufficient for $q_i \in \Theta$ if the conditional distribution, $P_{q_i}(X|T(X)=t)$, is independent of $q_i \in \Theta$ a.e.

One interpretation of this definition is the following: if we know only the sufficient statistic $T(X)$, and not the original outcome X , we could construct (randomly) an outcome X' that possesses a distribution identical to X , without any knowledge of the original distribution of X (Bahadur (1954 pp. 424-425)). Then if $T(X)$ is a sufficient statistic, no other statistic can provide more information about Θ than $T(X)$.

Consider now the problem of comparing two experiments E and F defined on the same parameter set, Θ . These experiments will result in two different statistics, T and W , in a way detailed in section 3.2. One would like to answer the question of which of these experiments is preferable. If both are equally costly, then the more preferable experiment might be the one that gives the most information about the parameter space. In this situation, a simple ordering of the experiments is all that is necessary to aid in the decision. However, suppose the cost of the experiment is an issue, and the least costly experiment results in the least informative statistic. How much will be 'lost' by basing conclusions on this experiment? In other words, how efficient is the least informative experiment compared to the most informative experiment?

When we are comparing card-counting statistics in our blackjack example, we will be able to determine how much information each of the statistics provides. Obviously, a simpler statistic is easier to remember, but how much information might be sacrificed by using this simpler statistic?

As part of our new technique for comparing statistics we will be computing distances between distributions. This distance will be a metric. The following definition comes from Lehmann (1986, p. 571).

DEFINITION 3.1.3: A space P is a metric space if for every p_1, p_2 , and p_3 in P :

- i. $d(p_1, p_2) \geq 0$
- ii. $d(p_1, p_2) = 0$ if and only if $p_1 = p_2$
- iii. $d(p_1, p_2) = d(p_2, p_1)$
- iv. $d(p_1, p_2) + d(p_2, p_3) \geq d(p_1, p_3)$

The metric we will be using in this paper is statistical distance.

DEFINITION 3.1.4: Let $P = \{ P_{q_i} : q_i \in \Theta, i=1,2 \}$ be a family of probability distributions over (X, A) . Then statistical distance is defined

$$d(P_{q_1}, P_{q_2}) = 2 \left(\sup_{A \in A} |P_{q_1}(A) - P_{q_2}(A)| \right).$$

A related metric is Hellinger distance which can be defined as

$$d(P_{q_1}, P_{q_2}) = \left[\int \left(\left(P_{q_1}(A) \right)^{\frac{1}{2}} - \left(P_{q_2}(A) \right)^{\frac{1}{2}} \right)^2 \right]^{\frac{1}{2}}.$$

Torgersen (1991 pp. 6, 32-34) notes that statistical distance and Hellinger distance are equivalent metrics.

3.2 The New Method

Recall definition 2.6.6 for the experiments $E = (X, A; P_{q_i} : q_i \in \Theta)$ and $F = (Y, B; Q_{q_i} : q_i \in \Theta)$.

Suppose that these two experiments yield two statistics, $T(X)$ and $W(Y)$, respectively, each of

which is some function of the observed data. One would like to measure the amount of information contained in each statistic. The literature suggests that one measure of information here might be based on sufficiency. Clearly a sufficient statistic contains more information than a non-sufficient one, but can one easily quantify the amount? And how can one compare two non-sufficient statistics?

More specifically, consider the comparison of the two statistics, $T(X)$ and $W(Y)$. If we restrict our attention to $T(X)$ then let X^* be the set of possible outcomes of $T(X)$ and let A^* be a σ -field of subsets of X^* . The parameter set is still $\Theta = (q_1, q_2)$ and now $P_{q_i}^*$ is the probability distribution of the outcome of $T(X)$ when q_i is the true parameter. So now when speaking in terms of the statistic $T(X)$, $E^* = (X^*, A^*; P_{q_i}^* : q_i \in \Theta)$ is a family of probability measures on the measurable space (X^*, A^*) . Similarly, in reference to the statistic $W(Y)$, $F^* = (Y^*, B^*; Q_{q_i}^* : q_i \in \Theta)$ is a family of probability measures on the measurable space (Y^*, B^*) .

In order to compare $T(X)$ and $W(Y)$ it will first be necessary to determine two things: 1) how efficient is $T(X)$ in distinguishing between q_1 and q_2 and as compared to a sufficient statistic; and 2) how efficient is $W(Y)$ in distinguishing between q_1 and q_2 as compared to a sufficient statistic? A method of measuring this efficiency will be introduced. Following are some definitions outlining the new method. Several nice properties of this new method will be presented and then this method will be demonstrated in some simple examples 3.4.1 and 4.1.

More specifically, let X be all of the observed data or another statistic sufficient for the family (P_{q_1}, P_{q_2}) . How efficient is $T(X)$ in distinguishing between q_1 and q_2 as compared to X ? For a given value t of a statistic $T(X)$, define the efficiency as such:

DEFINITION 3.2.1: The efficiency of $T(X)$ in distinguishing between q_1 and q_2 as compared to X is:

$$\text{eff}_{1,2}(T) = \frac{d(P_{q_1}, P_{q_2}) - d(P_{q_1}^{X/T(X)=t}, P_{q_2}^{X/T(X)=t})}{d(P_{q_1}, P_{q_2})}$$

where $P_{q_i}^{X/T(X)=t}$ is the conditional distribution of X given some observed value t of $T(X)$ when q_i is the true parameter.

Now we must address the problem that worried Kullback, the difficulty of comparing two distributions with different supports, or two distributions that are not absolutely continuous with respect to one another (see Definitions 2.4.1 and 2.4.2).

DEFINITION 3.2.2: Define $d(P_{q_1}^{X/T(X)=t}, P_{q_2}^{X/T(X)=t}) \equiv 0$ when t is not in the support of P_{q_1} or P_{q_2} .

In other words, if a value of $T(X)=t$ is observed that is possible under only one of the distributions, say P_{q_1} , then $T(X)=t$ provides perfect information about the parameter space and $\text{eff}_{1,2}(T)=1$.

Notice that the efficiency, $\text{eff}_{1,2}(T)$, is a random variable, therefore it is only natural to take its expected value.

DEFINITION 3.2.3: The expected efficiency of $T(X)$ with respect to X is the following:

$$E_{q_i}(\text{eff}_{1,2}(T)) = \frac{d(P_{q_1}, P_{q_2}) - E_{q_i} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\}}{d(P_{q_1}, P_{q_2})}$$

This expected efficiency is what we will use to compare statistics for different $q_i \in \Theta$. Under some circumstances it may be preferable to calculate a weighted average of this efficiency.

DEFINITION 3.2.4: The weighted average of the expected efficiency of $T(X)$ with respect to X , taken with respect to q_1 and q_2 , is the following:

$$\frac{W_1 \left\{ E_{q_1} [\text{eff}_{1,2}(T)] \right\} + W_2 \left\{ E_{q_2} [\text{eff}_{1,2}(T)] \right\}}{W_1 + W_2}$$

Here the weights W_1 and W_2 could be:

- i) Prior probabilities associated with P_{q_1} and P_{q_2} , respectively, if they are available; or
- ii) Utility functions. (See Good (1983 pp. 6-8, 189-91) for a discussion of utility functions).

3.3 Characterization of Sufficiency

There exist several ways of defining a sufficient statistic (see Definition 3.1.2). We will give several more existing definitions and then a new characterization of sufficiency. (Zacks (1971, p. 38)).

DEFINITION 3.3.1: Sufficient statistic: A statistic $T: (X, \mathcal{A}) \rightarrow (X^*, \mathcal{A}^*)$ is sufficient for $\{P_{q_i} : q_i \in \Theta\}$ if for every borel set $A \in \mathcal{A}$ and $A^* \in \mathcal{A}^*$,

$$P[A \cap T^{-1}(A^*)] = \int_{A^*} P[A|T=t] P^T(dt) \text{ where } P[A|T=t] \text{ is independent of the}$$

family $\{P_{q_i} : q_i \in \Theta\}$ and $P^T(A^*) = P(T^{-1}(A^*))$ for $A^* \in \mathcal{A}^*$.

Bahadur (1954, p. 430) notes that a statistic T is sufficient for the family $\{P_{q_i} : q_i \in \Theta\}$ if and only if the subfield \mathcal{T} induced by T is sufficient for $\{P_{q_i} : q_i \in \Theta\}$. (Refer to Definition 2.6.4.)

Another method of defining a sufficient statistic is based on the factorization criterion. (See Lehmann (1986, p. 20) for the discrete case and Billingsley (1986, p. 472) for the continuous case.)

DEFINITION 3.3.2: Factorization Criterion: Suppose that X is discrete. Then $T(X)$ is sufficient for the family $\{P_{q_i} : q_i \in \Theta\}$ if and only if there exists a factorization $P_{q_i}(X) = g_{q_i}[T(X)]h(X)$ where $g_{q_i}[T(X)]$ may depend on q_i and $T(X)$ only and $h(X)$ does not depend on q_i . A similar definition exists for the continuous case. In the case of the subfield \mathcal{T} induced by T , \mathcal{T} is sufficient if and only if the density p_{q_i} of P_{q_i} can be written $p_{q_i} = g_{q_i}h$, where g_{q_i} is \mathcal{T} -measurable (see Definition 2.6.2).

The following theorem provides a new way to characterize a sufficient statistic. If $T(X)$ is sufficient for $\{P_{q_i} : q_i \in \Theta\}$ then it should provide as much information about the parameter space as X will. In other words, the data, X , can be reduced by forming the statistic $T(X)$ and if $T(X)$ is sufficient then no information will be lost by basing a decision on $T(X)$ instead of on X . Or more formally:

THEOREM 3.3.1: For any parameter space Θ assume that $q_1 \neq q_2$. Consider the following measure of expected efficiency:

$$E_{q_i}(\text{eff}_{1,2}(T)) = \frac{d(P_{q_1}, P_{q_2}) - E_{q_i} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\}}{d(P_{q_1}, P_{q_2})}$$

Then $E_{q_i}(\text{eff}_{1,2}(T)) = 1$ for all pairs $(q_1, q_2) \in \Theta$ if and only if T is a sufficient statistic for the family $\{P_{q_i} : q_i \in \Theta\}$.

Proof: Assume that T is sufficient for the family $\{P_{q_i} : q_i \in \Theta\}$, or the conditional distribution,

$P_{q_i}^{X/T(X)=t}$ is independent of q_i . In other words, $P_{q_1}^{X/T(X)=t} = P_{q_2}^{X/T(X)=t}$. Then,

$$\begin{aligned} E_{q_i}(\text{eff}_{1,2}(T)) &= \frac{d(P_{q_1}, P_{q_2}) - E_{q_i} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\}}{d(P_{q_1}, P_{q_2})} \\ &= \frac{d(P_{q_1}, P_{q_2}) - 0}{d(P_{q_1}, P_{q_2})} \\ &= 1 \end{aligned}$$

Conversely, suppose that $E_{q_i}(\text{eff}_{1,2}(T)) = 1$ for all $q_i \in \Theta$. That implies:

$$\begin{aligned} E_{q_i}(\text{eff}_{1,2}(T)) &= \frac{d(P_{q_1}, P_{q_2}) - E_{q_i} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\}}{d(P_{q_1}, P_{q_2})} \\ &= 1 \end{aligned}$$

$$\text{or } E_{q_i} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\} = 0.$$

Since $\left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\}$ is a metric (see Definition 3.1.3), $\left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\} \geq 0$ for all t

a.e. Hence, $E_{q_i} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\} \geq 0$. So $E_{q_i} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\} = 0$ implies

$\left\{d\left(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}\right)\right\}=0$ for all t a.e. By definition of a metric, $\left\{d\left(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}\right)\right\} = 0$ a.e. if and only if $P_{q_1}^{X/T(X)=t} = P_{q_2}^{X/T(X)=t}$ a.e. Hence, $E_{q_i}(\text{eff}_{1,2}(T))=1$ for all $q_i \in \Theta$ implies that $P_{q_1}^{X/T(X)=t} = P_{q_2}^{X/T(X)=t}$ a.e. which implies that $P_{q_1}^{X/T(X)=t}$ is independent of q a.e. This is true if and only if T is sufficient for the family $\{P_{q_i} : q_i \in \Theta\}$.

◻

Since no other statistic can provide more information about the parameter space than a sufficient statistic one would expect that this measure of efficiency is never greater than 1.

LEMMA 3.3.1: If $P_{q_1} \neq P_{q_2}$ then:

$$E_{q_i}(\text{eff}_{1,2}(T)) = \frac{d(P_{q_1}, P_{q_2}) - E_{q_i}\left\{d\left(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}\right)\right\}}{d(P_{q_1}, P_{q_2})} \leq 1$$

$$\begin{aligned} \text{Proof: } E_{q_i}(\text{eff}_{1,2}(T)) &= \frac{d(P_{q_1}, P_{q_2}) - E_{q_i}\left\{d\left(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}\right)\right\}}{d(P_{q_1}, P_{q_2})} \\ &= 1 - \frac{E_{q_i}\left\{d\left(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}\right)\right\}}{d(P_{q_1}, P_{q_2})} \end{aligned}$$

Since $\left\{d\left(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}\right)\right\}$ and $\left\{d(P_{q_1}, P_{q_2})\right\}$ are both metrics, they are non-negative quantities

(as is $E_{q_i}\left\{d\left(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}\right)\right\}$) hence their ratio $\frac{E_{q_i}\left\{d\left(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}\right)\right\}}{d(P_{q_1}, P_{q_2})}$ is non-negative. So

$$1 - \frac{E_{q_i}\left\{d\left(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}\right)\right\}}{d(P_{q_1}, P_{q_2})} \leq 1$$

$$\text{or } E_{q_i}(\text{eff}_{1,2}(T)) = \frac{d(P_{q_1}, P_{q_2}) - E_{q_i} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\}}{d(P_{q_1}, P_{q_2})} \leq 1$$

The counterpart to a sufficient statistic is a noninformative statistic or an ancillary statistic. The following two definitions are from Lehmann (1986, pp. 141-144, 542-543)).

DEFINITION 3.3.3: Suppose we have a set of observations X that are distributed according to $\{P_{q_i} : q_i \in \Theta\}$. A statistic T is ancillary if its distribution does not depend on $q_i \in \Theta$. In other words, if one observes $T(X)$ and then, given $T(X)$, observes a quantity X with distribution $P_{q_i}^{X/T(X)=t}$, X will be distributed according to the original distribution P_{q_i} a.e. $T(X)$ is ancillary if it provides no information about the distribution of X .

DEFINITION 3.3.4: A family of distributions $\{P_{q_i} : q_i \in \Theta\}$ is boundedly complete if for all bounded functions f , $E_{q_i}[f(X)] = 0$ for all P_{q_i} implies $f(X) = 0$ a.e.

An ancillary statistic provides no information about the parameter space before it has been observed and therefore its expected efficiency compared to a sufficient statistic should be very small. This expected efficiency is, in fact, equal to zero as we shall soon prove. To assist us in our proof we will need to use a theorem known as Basu's Theorem (Encyclopedia of Statistics, vol. 1, pp 193-196).

THEOREM 3.3.2 Basu's Theorem: Let X be sufficient and boundedly complete. Then a statistic T is ancillary only if it is conditionally independent of X for each q_i .

If X and T are conditionally independent then $P_{q_i}^{X|T(X)=t} = P_{q_i}$ a.e. Now let's state the theorem related to an ancillary statistic.

THEOREM 3.3.3: If T is an ancillary statistic and if X is sufficient and boundedly complete then $E_{q_i}(\text{eff}_{1,2}(T)) = 0$ for all $q_i \in \Theta$.

Proof: Suppose that T is ancillary. Then the distribution of T does not depend on q_i , and by Theorem 3.3.2, $P_{q_i}^{X|T(X)=t} = P_{q_i}$ a.e. Hence,

$$\begin{aligned} E_{q_i}(\text{eff}_{1,2}(T)) &= \frac{d(P_{q_1}, P_{q_2}) - E_{q_i} \left\{ d(P_{q_1}^{X|T(X)=t}, P_{q_2}^{X|T(X)=t}) \right\}}{d(P_{q_1}, P_{q_2})} \\ &= \frac{d(P_{q_1}, P_{q_2}) - d(P_{q_1}, P_{q_2})}{d(P_{q_1}, P_{q_2})} \\ &= 0. \end{aligned}$$

3.4 Further Observations

What advantages does the expected efficiency have over Fisher's information and the Kullback information function? The following two statements will be demonstrated with example 3.4.1:

1) It is possible for the Fisher's information in a nonsufficient statistic to equal the Fisher's information in a sufficient statistic. In example 3.4.1, X is a sufficient statistic for $\{P_{q_i}\}$ and T is not and $I(f(X, q_i)) = I(f(T, q_i))$ for every q_i . In our measure of information, the expected efficiency will equal one if and only if X is sufficient (Theorem 3.3.1). Hence, when this new

technique is used, the efficiency of a nonsufficient statistic will always be strictly less than that for a sufficient statistic.

2) Recall that the Kullback information function requires that P_{q_1} and P_{q_2} be absolutely continuous with respect to one another (definition 2.4.1). Otherwise the function is not a useful measure of information. In example 3.4.1 we wish to discriminate between P_{q_1} and P_{q_2} and the Kullback information function will in fact be equal to ∞ for q_1 . (Dr. Good suggests that it might be possible to avoid this problem of infinite Kullback information. See the discussion in section 2.4.) The new technique, however, will allow us to measure the efficiency of many statistics as compared to a sufficient statistic, without making any assumptions about the distribution except that it be known.

EXAMPLE 3.4.1: Suppose $X = \{x_1, x_2, x_3, x_4\}$ and we make one observation, X , where

$$P_{q_1}(X = x_1) = \frac{q_1}{4} \quad P_{q_1}(X = x_2) = \frac{q_1}{4} \quad P_{q_1}(X = x_3) = \frac{q_1}{4} \quad P_{q_1}(X = x_4) = \frac{4 - 3q_1}{4}$$

Let $T(x_1) = T(x_2) = t_1$, $T(x_3) = t_2$, $T(x_4) = t_3$ where $t_1 \neq t_2$, $t_1 \neq t_3$, $t_2 \neq t_3$. Then

$$P_{q_1}(T = t_1) = \frac{q_1}{4} + \frac{q_1}{4} = \frac{q_1}{2} \quad P_{q_1}(T = t_2) = \frac{q_1}{4} \quad P_{q_1}(T = t_3) = \frac{4 - 3q_1}{4}$$

Let us calculate Fisher's information for X :

$$\begin{aligned} I(f(X, q_1)) &= \sum_{x_j} \left(\frac{1}{q_1} \log P_{q_1}(X = x_j) \right)^2 P_{q_1}(X = x_j) \\ &= \left(\frac{1}{q_1} \log \frac{q_1}{4} \right)^2 \left(\frac{q_1}{4} \right) + \left(\frac{1}{q_1} \log \frac{q_1}{4} \right)^2 \left(\frac{q_1}{4} \right) + \left(\frac{1}{q_1} \log \frac{q_1}{4} \right)^2 \left(\frac{q_1}{4} \right) + \left(\frac{1}{q_1} \log \frac{4 - 3q_1}{4} \right)^2 \left(\frac{4 - 3q_1}{4} \right) \\ &= 3 \left(\frac{4}{q_1} \cdot \frac{1}{4} \right)^2 \frac{q_1}{4} + \left(\frac{4}{4 - 3q_1} \cdot \frac{-3}{4} \right)^2 \frac{4 - 3q_1}{4} \end{aligned}$$

$$= \frac{3}{4q_1} + \frac{9}{4(4-3q_1)}$$

Now let us calculate Fisher's information for T:

$$\begin{aligned} I(f(T, q_1)) &= \sum_{t_k} \left(\frac{1}{1q} \log P_{q_1}(T = t_k) \right)^2 P_{q_1}(T = t_k) \\ &= \left(\frac{1}{1q} \log \frac{q_1}{2} \right)^2 \left(\frac{q_1}{2} \right) + \left(\frac{1}{1q} \log \frac{q_1}{4} \right)^2 \left(\frac{q_1}{4} \right) + \left(\frac{1}{1q} \log \frac{4-3q_1}{4} \right)^2 \left(\frac{4-3q_1}{4} \right) \\ &= \left(\frac{2}{q_1} \cdot \frac{1}{2} \right)^2 \frac{q_1}{2} + \left(\frac{4}{q_1} \cdot \frac{1}{4} \right)^2 \frac{q_1}{4} + \left(\frac{4}{4-3q_1} \cdot \frac{-3}{4} \right)^2 \frac{4-3q_1}{4} \\ &= \frac{3}{4q_1} + \frac{9}{4(4-3q_1)} \end{aligned}$$

So $I(f(X, q_1)) = I(f(T, q_1))$ yet T is not sufficient for the family $\{P_{q_1}, P_{q_2}\}$ (as will be shown shortly).

Let the distribution for q_2 be the following:

$$P_{q_2}(X = x_1) = \frac{q_2}{2} \quad P_{q_2}(X = x_2) = 0 \quad P_{q_2}(X = x_3) = 0 \quad P_{q_2}(X = x_4) = \frac{2-q_2}{2}$$

Again $T(x_1) = T(x_2) = t_1$, $T(x_3) = t_2$, $T(x_4) = t_3$ where $t_1 \neq t_2$, $t_1 \neq t_3$, $t_2 \neq t_3$. Then

$$P_{q_2}(T = t_1) = \frac{q_2}{2} + 0 = \frac{q_2}{2} \quad P_{q_2}(T = t_2) = 0 \quad P_{q_2}(T = t_3) = \frac{2-q_2}{2}$$

$$\text{Then } I(f(X, q_2)) = I(f(T, q_2)) = \frac{1}{2q_2} + \frac{1}{2(2-q_2)}$$

Again we see that $I(f(X, q_2)) = I(f(T, q_2))$ when T is not sufficient for the family $\{P_{q_1}, P_{q_2}\}$ and X is. (T's lack of sufficiency will be proven shortly.)

Now let's calculate the expected efficiency of T with respect to X using the theory introduced in section 3.2:

$$E_{q_i}(\text{eff}_{1,2}(T)) = \frac{d(P_{q_1}, P_{q_2}) - E_{q_i} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\}}{d(P_{q_1}, P_{q_2})}$$

Recall we are using statistical distance so $d(P_{q_1}, P_{q_2}) = 2 \left(\sup_{A \in \mathcal{A}} |P_{q_1}(A) - P_{q_2}(A)| \right)$

Or in the discrete case, $d(P_{q_1}, P_{q_2}) = 2 \left(\sum_{x_j} |P_{q_1}(X = x_j) - P_{q_2}(X = x_j)| \right)$

First we will calculate the nonzero probabilities for the conditional part:

$$E_{q_i} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\} \quad i = 1, 2.$$

$$P_{q_1}(X = x_1/T = t_1) = \frac{1}{2}$$

$$P_{q_2}(X = x_1/T = t_1) = 1$$

$$P_{q_1}(X = x_2/T = t_1) = \frac{1}{2}$$

$$P_{q_2}(X = x_2/T = t_1) = 0$$

$$P_{q_1}(X = x_3/T = t_2) = 1$$

$$P_{q_2}(X = x_3/T = t_2) = 0$$

$$P_{q_1}(X = x_4/T = t_3) = 1$$

$$P_{q_2}(X = x_4/T = t_3) = 1$$

All other conditional probabilities are equal to zero.

Recall the definition of a sufficient statistic: $T=T(X)$ is sufficient for the family of distributions $\{P_{q_i} : q_i \in \Theta\}$ if and only if the conditional distribution of X given $T=t$ is independent of q_i a.e. Notice that in this example the conditional distribution of X given T depends on q_i . Hence T is not sufficient. So we have for $i=1,2$ $j=1,2,3,4$ $k=1,2,3$:

$$E_{q_i} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\} = E_{q_i} \left\{ 2 \left(\sum_{x_j} |P_{q_1}(X = x_j/T) - P_{q_2}(X = x_j/T)| \right) \right\}$$

and $d(P_{q_1}^{X/T(X)=t}, P_{q_2}^{X/T(X)=t}) \equiv 0$ if t is not in the support of T under either q_1 or under q_2 .

$$\begin{aligned}
& \text{So } E_{q_1} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\} \\
&= 2 \left\{ \left| P_q(X=x_1/T=t_1) - P_{q_2}(X=x_1/T=t_1) \right| P_q(T=t_1) \right. \\
&+ \left| P_q(X=x_2/T=t_1) - P_{q_2}(X=x_2/T=t_1) \right| P_q(T=t_1) \\
&+ \left| P_q(X=x_3/T=t_2) - P_{q_2}(X=x_3/T=t_2) \right| P_q(T=t_2) \\
&+ \left. \left| P_{q_1}(X=x_4/T=t_3) - P_{q_2}(X=x_4/T=t_3) \right| P_{q_1}(T=t_3) \right\} \\
&= 2 \left\{ \left| \frac{1}{2} - 1 \right| \frac{q_1}{2} + \left| \frac{1}{2} - 0 \right| \frac{q_1}{2} + 0 + \left| 1 - 1 \right| \left(\frac{4 - 3q_1}{4} \right) \right\} \\
&= 2 \left\{ \frac{q_1}{2} \right\}
\end{aligned}$$

$$\begin{aligned}
& \text{and } E_{q_2} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\} \\
&= 2 \left\{ \left| P_{q_1}(X=x_1/T=t_1) - P_{q_2}(X=x_1/T=t_1) \right| P_{q_1}(T=t_1) \right. \\
&+ \left| P_q(X=x_2/T=t_1) - P_{q_2}(X=x_2/T=t_1) \right| P_{q_1}(T=t_1) \\
&+ \left| P_q(X=x_3/T=t_2) - P_{q_2}(X=x_3/T=t_2) \right| P_{q_2}(T=t_2) \\
&+ \left. \left| P_{q_1}(X=x_4/T=t_3) - P_{q_2}(X=x_4/T=t_3) \right| P_{q_2}(T=t_3) \right\} \\
&= 2 \left\{ \left| \frac{1}{2} - \frac{2}{2} \right| \frac{1}{2} + \left| \frac{1}{2} - 0 \right| \frac{2}{2} + \left| 1 - 1 \right| \left(\frac{-2}{2} \right) \right\} \\
&= 2 \frac{q_2}{2}
\end{aligned}$$

Now calculating $P_{q_1} P_{q_2}$ will be a bit more complicated.

$$\begin{aligned}
d(P_{q_1}, P_{q_2}) &= 2 \left(\sum_{x_j} |P_{q_1}(X = x_j) - P_{q_2}(X = x_j)| \right) = 2 \left\{ |P_{q_1}(X = x_1) - P_{q_2}(X = x_1)| \right. \\
&+ |P_{q_1}(X = x_2) - P_{q_2}(X = x_2)| + |P_{q_1}(X = x_3) - P_{q_2}(X = x_3)| + |P_{q_1}(X = x_4) - P_{q_2}(X = x_4)| \Big\} \\
&= 2 \left(\left| \frac{q_1}{4} - \frac{q_2}{2} \right| + \left| \frac{q_1}{4} - 0 \right| + \left| \frac{q_1}{4} - 0 \right| + \left| \frac{4 - 3q_1}{4} - \frac{2 - q_2}{2} \right| \right) \\
&= 2 \left(\left| \frac{q_1 - 2q_2}{4} \right| + \frac{q_1}{2} + \left| \frac{3q_1 - 2q_2}{4} \right| \right)
\end{aligned}$$

Then we have

$$d(P_{q_1}, P_{q_2}) = \begin{cases} 2 \left(\frac{3q_1}{2} - q_2 \right) & \text{if } q_1 \geq 2q_2 \\ -2 \left(\frac{q_1}{2} - q_2 \right) & \text{if } 3q_1 \leq 2q_2 \\ 2q_1 & \text{if } q_1 < 2q_2 \text{ and } 3q_1 > 2q_2 \end{cases}$$

$$\begin{aligned}
\text{So } E_{q_1}(\text{eff}_{1,2}(T)) &= \frac{d(P_{q_1}, P_{q_2}) - E_{q_1} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\}}{d(P_{q_1}, P_{q_2})} \\
&= 1 - \frac{E_{q_1} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\}}{d(P_{q_1}, P_{q_2})}
\end{aligned}$$

$$\text{Then if } q_1 \geq 2q_2, \quad E_{q_1}(\text{eff}_{1,2}(T)) = 1 - \frac{\frac{q_1}{2}}{\left(\frac{3q_1}{2} - q_2 \right)} = 1 - \frac{q_1}{3q_1 - 2q_2}$$

$$\text{and } E_{q_2}(\text{eff}_{1,2}(T)) = 1 - \frac{\frac{q_2}{2}}{\left(\frac{3q_1}{2} - q_2 \right)} = 1 - \frac{q_2}{3q_1 - 2q_2}$$

$$\text{If } 3q_1 \leq 2q_2, \quad E_{q_1}(\text{eff}_{1,2}(T)) = 1 - \frac{\frac{q_1}{2}}{\left(q_2 - \frac{q_1}{2}\right)} = 1 - \frac{q_1}{2q_2 - q_1}$$

$$\text{and } E_{q_2}(\text{eff}_{1,2}(T)) = 1 - \frac{\frac{q_2}{2}}{\left(q_2 - \frac{q_1}{2}\right)} = 1 - \frac{q_2}{2q_2 - q_1}.$$

$$\text{If } q_1 < 2q_2 \text{ and } 3q_1 > 2q_2 \text{ then } E_{q_1}(\text{eff}_{1,2}(T)) = 1 - \frac{\frac{q_1}{2}}{q_1} = 1 - \frac{1}{2} = \frac{1}{2}$$

$$\text{and } E_{q_2}(\text{eff}_{1,2}(T)) = 1 - \frac{\frac{q_2}{2}}{q_1} = 1 - \frac{q_2}{2q_1}$$

Notice that our expected efficiencies are less than one, hence T is not a sufficient statistic by Theorem 3.3.1. So using our measure of efficiency we find that the efficiency of X for the model $\{P_{q_1}, P_{q_2}\}$ is strictly greater than the efficiency of T . According to Fisher's measure of information, these two statistics contain the same amount of information!

While Kullback information for this example is ∞ for q_1 (in the denominator we have $f(X, q_2) = 0$ for $X=x_2$ and $X=x_3$), our new technique does provide us with information. But what does a value of $E_{q_1}(\text{eff}_{1,2}(T))$ by itself mean? This will be discussed in chapter 4.

3.5 Summary of the New Method

In this chapter we have developed a new method for comparing experiments. This method measures the expected efficiency of any statistic at discriminating between distributions as compared to a sufficient statistic. The new technique works by measuring distances between distributions. The larger the distance, the better a statistic “separates”, or discriminates,

“separates” distributions as compared to a sufficient statistic.

We have also discussed how the new method can be superior to existing techniques at or distribution, the new method can provide useful information where the Fisher and Kullback provides, as in the methods of comparing risk or power functions. However, we also can obtain quantifying information about how much more or less information one statistic contains of the new method involve intermediate probability and computing skills (and a little practice). This makes the new method not only more practical to use, but easier to understand and

Chapter 4

Calibration

Now that we have defined the new method of comparing experiments, it's time to discuss its interpretability. As mentioned in Chapter 3, a value of $E(\text{eff}(T))$ by itself has little meaning. Its scale is not linear and does, in fact, depend on the application. And it is necessary to find a common scale against which to compare different statistics.

In studying different types of efficiencies in the most are functions of the sample size (see Nikitin (1995) for a survey of asymptotic (1954, p. 227) refers to sequences X_n which consist of increasingly informative observations.

suppose our experiment consists of taking ten observations, $X = (x_1, x_2, \dots, x_{10})$. One sufficient statistic would be these ten observations, X_{10} , and its expected efficiency as

$X_1, X_2, X_3, \dots, X_9$, consisting of sequences $(x_1), (x_1, x_2), (x_1, x_2, x_3), \dots, (x_1, \dots, x_9)$, respectively, and calculate the expected efficiency of each of these. These expected efficiencies form a framework against which other statistics can be compared. In other words, if the expected efficiency for any statistic T is close to the expected efficiency for X_7 , then the observation can be made that T contains approximately as much information as $\frac{7}{10}$ or 70% of the sample. This concept will be demonstrated in example 4.1. It will be used in the blackjack problem also.

EXAMPLE 4.1: A woman has been playing the slot machines on 3 randomly selected days. She either played 3 games each of the days or 4 games each of the days. If she played 3 games a day, her chance of winning is .25. If she played 4 games a day, her chance of winning is .5. Her husband asks her how many games she won on day i , x_i , in order to estimate the number of games she plays each day and her probability of winning. This is a binomial experiment with unknown n and p . The sufficient statistic that will be used for this problem is the trivial statistic, $X_3 = (x_1, x_2, x_3)$.

- Let x_i = number of games won on day i .
- $x_i \sim B(3, .25)$ or $x_i \sim B(4, .5)$
- Let $X = X_3 = (x_1, x_2, x_3)$
- Then $P(X = (x_1, x_2, x_3)) = \prod_{i=1}^3 \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}$

(Notice that these two distributions are not absolutely continuous with respect to each other. Hence, the Kullback information function would not be useful. In fact, the Kullback information would be infinite for $n=4$.)

First, suppose her husband asks her how many games she won on the first day. (So $T = x_1$.) How efficient is that information in estimating the number of games played and the probability of winning as opposed to asking her how many games she won on all three days?

$$\text{Recall } E_{q_i}(\text{eff}_{1,2}(T)) = \frac{d(P_{q_1}, P_{q_2}) - E_{q_i} \left\{ d(P_{q_1}^{X/T(X)}, P_{q_2}^{X/T(X)}) \right\}}{d(P_{q_1}, P_{q_2})}$$

$$\text{where } d(P_{q_1}, P_{q_2}) = \sum_{x_3=0}^4 \sum_{x_2=0}^4 \sum_{x_1=0}^4 \left| \prod_{i=1}^3 \binom{3}{x_i} (.25)^{x_i} (.75)^{3-x_i} - \prod_{i=1}^3 \binom{4}{x_i} (.5)^{x_i} (.5)^{4-x_i} \right| = 1.5321$$

Values for $E_{q_i}(\text{eff}_{1,2}(T))$ are given in Table 4.1 for T equal to $(x_1), (x_1 + x_2), (x_1, x_2), (x_1 + x_2 + x_3)$ and (x_1, x_2, x_3) .

TABLE 4.1 Expected Efficiencies for Example 4.1

T	$E_{n=3, p=.25}(\text{eff}_{1,2}(T))$	$E_{n=4, p=.5}(\text{eff}_{1,2}(T))$
$X_1 = (x_1)$	0.1045	0.1604
$(x_1 + x_2)$	0.3058	0.2958
$X_2 = (x_1, x_2)$	0.3065	0.3905
$(x_1 + x_2 + x_3)$	0.9659	0.7624
$X_3 = (x_1, x_2, x_3)$	1.0	1.0

One can see in this example that the sum of two observations provides slightly less information than the two observations themselves. The sum is a one-dimensional number. Perhaps it is easier to remember this sum than the two separate numbers. In that case, the loss in information will not be too great if one chooses to use $(x_1 + x_2)$ over (x_1, x_2) .

Now consider the expected efficiencies for $(x_1 + x_2 + x_3)$. By Theorem 3.3.1, this statistic is not sufficient for the family of distributions since these expected efficiencies are not equal to one. Indeed, the only sufficient statistics for estimating both n and p in a binomial distribution are the trivial statistic, (x_1, x_2, x_3) , and the order statistic. And one can see that the sum of these three observations is less efficient in estimating n and p than (x_1, x_2, x_3) but more efficient than (x_1, x_2) .

Notice also that $E_{q_i}(\text{eff}_{1,2}(x_1)) \leq E_{q_i}(\text{eff}_{1,2}(x_1, x_2)) \leq E_{q_i}(\text{eff}_{1,2}(x_1, x_2, x_3))$ for $i = 1, 2$, as would be expected since (x_1) , (x_1, x_2) , and (x_1, x_2, x_3) are increasingly informative sequences.

Chapter 5

Application of the New Method

5.1 Introduction

Games of chance can be traced back to prehistoric times. Colored pebbles and astragali (heel bones) were used in some of the earliest gaming. Eventually dice evolved from these implements and several thousand years later playing cards were invented. However, it wasn't until the sixteenth century that gambling began to be described in probabilistic terms. Since that time man has attempted to calculate for any game of chance the odds of winning, the degree of certainty of any decision made during the game, the effect of information on the odds. Much of early probability theory evolved from these attempts.

Modern casino games are generally straightforward betting games. Probabilities of winning are predetermined by the house and there is no way that any knowledge of the bettor can change the odds. One exception to these games is Blackjack. Since its introduction in the late nineteenth century, various betting strategies, computer simulations and card-counting statistics associated with Blackjack have been developed. As Epstein states, “Blackjack is the only form of public gaming available where the player is able to obtain a mathematical expectation greater than zero.” (Epstein (1995 p. 215).

5.2 Rules for Blackjack

The rules for Blackjack vary from casino to casino, so we will follow the general rules stated in Epstein (1995 pp. 215-251). There is a dealer, representing the house, and one to six players. Bets are placed before the deal. Anywhere from one to eight decks of cards are shuffled together and used in the game. Two cards are dealt to each person, the players' are

both facing down while one of the dealer's is facing down and one is facing up. Each card is assigned a numerical value, an ace takes on a value of 1 or 11; tens, jacks, queens and kings take on the value 10; all other cards take on their face value. A player may draw additional cards with the goal of obtaining a card total that exceeds the dealer's total but does not exceed a total of 21. Some different situations might occur:

- If a player's total exceeds 21 he loses the bet.
- If a player's total is less than or equal to 21 and the dealer's total exceeds 21, the player wins the bet.
- If both totals are less than or equal to 21, the total closest to 21 wins the bet.
- If the totals are the same, it's a tie and the bet is canceled.
- If the first two cards are an ace and a ten card, this combination is called a blackjack and wins over all other combinations equaling 21.
- Blackjack in the dealer's hand and the player's hand is a tie and the bet is canceled.

At the end of the round, the dealt cards are exposed and put aside and the next round is drawn from the remaining cards. After 60% to 80% of the cards have been played at the end of a round the entire deck is reshuffled.

A player would have a definite advantage over the house if he could remember exactly every card that had been played in previous rounds and placed his bet accordingly. Or instead of remembering every card, since suits are of no consequence, he could simply remember the number of aces, the number of twos, the number of ten's, etc. However, as the number of rounds increases, it becomes increasingly difficult to keep track of this information. In lieu of this difficulty, many card-counting statistics have been developed. We will compare several of these statistics using the methodology developed in the preceding sections.

Some of the simpler card-counting statistics were evaluated by determining the mathematical expectation as a result of drawing a card and not drawing a card. This expectation needed to be calculated for the different totals of the player's hand and for different values of the dealer's up-card. The statistics could then be compared based on these expectations. But for more complicated statistics these expectations can be difficult to calculate. Perhaps an easier way to compare these card counting statistics would be to measure the amount of information that each one provides.

5.3 Deck Composition

There may be many different types of deck compositions among which we wish to distinguish. The deck composition will determine our betting strategy. For the sake of simplicity, we use an example that Epstein (1995, pp. 242-243) uses: we have a complete 52-card deck. It is possible that one card was removed, either by an unscrupulous dealer or some error. It is in our best interest to determine the composition of the remaining cards. For example, our betting strategy will differ considerably depending on whether an ace is missing from the deck or whether a five is missing. So our parameter space consists of eleven possible deck compositions, a 52-card deck, a 51-card deck with one ace missing, a 51-card deck with one two missing, etc.

More formally, suppose that k cards have been dealt. Let $X_k = (x_1, x_2, \dots, x_k)$ and let y_1 =number of aces that have been dealt, y_2 = number of twos that have been dealt, ... , y_{10} =number of tens that have been dealt. Notice that $\sum_{i=1}^{10} y_i = k$. The parameter set $\Theta = (q_0, q_1, q_2, \dots, q_{10})$ where P_{q_0} = the probability distribution of the remaining undealt cards given no cards are missing (we started with a 52-card deck), P_{q_1} = the probability distribution of the remaining undealt cards given one ace is missing from the original deck (i.e. only three aces are left in a 51-card deck), P_{q_2} = the probability distribution of the remaining undealt cards

given one two is missing from the original deck (i.e. only three twos are left in a 51-card deck), etc., and $P_{q_{10}}$ = the probability distribution of the remaining undealt cards given one ten is missing from the original deck (and only 15 tens are left in a 51-card deck).

The statistic X in this example is a sufficient statistic for providing information about the family of different distributions. However, as mentioned previously, it is rather cumbersome to remember X for large values of k . Several popular card-counting statistics will be studied. Their names and counts for ace through ten are given in Table 5.3.1.

Notice that none of these card-counting statistics is sufficient. Indeed, a great deal of information about the deck composition is lost by using any of these statistics. So which statistic is best? Which statistic allows one to make the most informed decision during the next round? Notice that the Hiopt I statistic is simpler to work with than any of the others. How much worse off will one be using the Hiopt I if it turns out that it is the least informative?

TABLE 5.3.1 Point counts for Some Card-Counting Statistics

Statistic	A	2	3	4	5	6	7	8	9	10
Ten Count	4	4	4	4	4	4	4	4	4	-9
Hiopt I	0	0	1	1	1	1	0	0	0	-1
Braun +-	-1	1	1	1	1	1	0	0	0	-1
Hiopt II	0	1	1	2	2	1	1	0	0	-2
Uston's APC	0	1	2	2	3	2	2	1	-1	-3
Griffin 4	0	1	2	3	4	3	3	1	-1	-4

Before analyzing these card-counting statistics it is necessary to calibrate our measure of efficiency with respect to the family of distributions. We will calibrate our measure by calculating the expected efficiency given that one of our k cards dealt is known, given that 2 of our k cards dealt are known, etc., and finally given that k of our k cards are known. This will

give us a scale of reference when we calculate the expected efficiencies of the card-counting statistics.

5.4 Calibration

Recall, when we calculate the expected efficiency of a statistic we are comparing it to a sufficient statistic. Our expected efficiency will be a number between 0 and 1. What exactly does an efficiency of .85 mean? In this example does it mean our statistic provides the information equivalent to knowing 85% of the cards that have been played? Or does it mean our statistic provides the information equivalent to knowing, for example, 50% of the cards that have been played?

In the Blackjack example our sufficient statistic $X_k = (x_1, x_2, \dots, x_k)$ represents all of the cards that have been dealt. If k cards have been dealt, let X_l represent l of those k cards. We would like to calculate the expected efficiency of X_l for $l=0, \dots, k$. (The expected efficiency for $l=0$ is zero and for $l=k$ is one.) This will give us a scale of comparison for any other card-counting statistic.

A computer program was used to calculate these efficiencies. The program code is included in Appendix C. A 52 card deck was used, $k=8$ cards were dealt, and the expected efficiencies for $X_0, X_1, X_2, \dots, X_8$ were calculated for each pair (q_0, q_i) $i=1, \dots, 10$. Those efficiencies are given in Table 5.4.1 and Table 5.4.2. For example, $\text{eff}_{0,1}(X_4)$ represents how well knowing four of the cards that have been dealt helps distinguish between a full deck and a deck with an ace missing, as compared to knowing all of the cards that have been dealt. $E_{q_1}(\text{eff}_{0,1}(X_4))$ is the expected value of this efficiency when q_1 is the true parameter (one ace is missing from the deck). So, we are calculating the expected efficiency of X_l in

discriminating between all pairs P_{q_0} and P_{q_i} , for $i=1,\dots,10$, $l=0,\dots,8$. The expected efficiencies for all pairs consisting of P_{q_0} and any of P_{q_1} through P_{q_9} are all the same.

TABLE 5.4.1 Calibration of Expected Efficiency for P_{q_0} and P_{q_i} , $i=1,\dots,9$

X_l $l = 0, \dots, 8$	$E_{q_0}(\text{eff}_{0,i}(X_l))$ <i>the same for $i = 1, \dots, 9$</i>	$E_{q_i}(\text{eff}_{0,i}(X_l))$ <i>the same for $i = 1, \dots, 9$</i>
X_0 , zero cards known	0.0000	0.0000
X_1 , one card known	0.0396	0.0426
X_2 , two cards known	0.0984	0.1031
X_3 , three cards known	0.1788	0.1842
X_4 , four cards known	0.2833	0.2883
X_5 , five cards known	0.4147	0.4186
X_6 , six cards known	0.5759	0.5781
X_7 , seven cards known	0.7699	0.7706
X_8 , eight cards known	1.0000	1.0000

Notice in this example the efficiencies are monotonically increasing between zero and one. Next, if we calculate the efficiency of a card counting statistic and find it has an efficiency of, for example 0.57 then we can say that this card-counting statistic provides about the same information as remembering approximately 80% of the cards that have been dealt.

A Table 5.4.2 is given for distinguishing between P_{q_0} and $P_{q_{10}}$. In comparing Table 5.4.1 and Table 5.4.2 we see some very interesting results. Observe the expected efficiencies for X_7 , when seven cards are known. They are much larger in Table 5.4.1 when we are trying to

distinguish between a full deck and a deck with say, an ace missing, as opposed to Table 5.4.2, when we are trying to distinguish between a full deck and a deck with a ten missing. This is logical, since it is possible to observe four aces in seven cards and know definitely which deck composition we have. However, in the situation where we are comparing a full deck to a deck with a ten missing, we are still left with 15 tens. Hence, it is not possible to know definitely, from only seven cards, which is the true deck composition.

TABLE 5.4.2 Calibration of Expected Efficiency for P_{q_0} and $P_{q_{10}}$

$X_l \quad l = 0, \dots, 8$	$E_{q_0}(\text{eff}_{0,10}(X_l))$	$E_{q_{10}}(\text{eff}_{0,10}(X_l))$
X_0 , zero cards known	0.0000	0.0000
X_1 , one card known	0.0954	0.0955
X_2 , two cards known	0.1508	0.1527
X_3 , three cards known	0.1697	0.1715
X_4 , four cards known	0.2765	0.2772
X_5 , five cards known	0.4031	0.4067
X_6 , six cards known	0.4173	0.4203
X_7 , seven cards known	0.5754	0.5767
X_8 , eight cards known	1.0000	1.0000

Now let's compare the efficiencies for the card-counting statistics to these efficiencies. Tables 5.4.3, 5.4.4 and 5.4.5 contain the efficiencies for the card-counting statistics.

TABLE 5.4.3 Expected Efficiencies for Ten Count and Hiopt I

P_{q_0} vs. P_{q_i}	Ten Count		Hiopt I	
	$E_{q_0}(\text{eff}_{0,i}(T))$	$E_{q_i}(\text{eff}_{0,i}(T))$	$E_{q_0}(\text{eff}_{0,i}(T))$	$E_{q_i}(\text{eff}_{0,i}(T))$
$i=1$	0.0345	0.0345	0.0044	0.0048
$i=2$	0.0345	0.0345	0.0044	0.0048
$i=3$	0.0345	0.0345	0.1069	0.1208
$i=4$	0.0345	0.0345	0.1069	0.1208
$i=5$	0.0345	0.0345	0.1069	0.1208
$i=6$	0.0345	0.0345	0.1069	0.1208
$i=7$	0.0345	0.0345	0.0044	0.0048
$i=8$	0.0345	0.0345	0.0044	0.0048
$i=9$	0.0345	0.0345	0.0044	0.0048
$i=10$	0.9998	0.9998	0.4975	0.5001

When trying to distinguish between a full deck and a ten-depleted deck, the ten count statistic provides as much information as knowing almost eight out of eight cards. In this same situation, the Hiopt I, Hiopt II, Uston's APC and Griffin-4 statistics provide as much information as knowing about six out of eight cards. Braun+- statistic provides less information than knowing five out of eight cards. So these statistics provide a good amount of information when we need to distinguish between a full deck and a ten-depleted deck. However, when we are trying to distinguish between a full deck and other deck compositions, these statistics provide very little information. The amount of information provided ranges from the equivalent of knowing two out of eight cards to knowing zero out of eight cards!

TABLE 5.4.4 Expected Efficiencies for Braun + - and Hiopt II

P_{q_0} vs. P_{q_i}	Braun + -		Hiopt II	
	$E_{q_0}(\text{eff}_{0,i}(T))$	$E_{q_i}(\text{eff}_{0,i}(T))$	$E_{q_0}(\text{eff}_{0,i}(T))$	$E_{q_i}(\text{eff}_{0,i}(T))$
$I=1$	0.0893	0.1005	0.0048	0.0053
$I=2$	0.0893	0.1005	0.0365	0.0409
$I=3$	0.0893	0.1005	0.0365	0.0409
$i=4$	0.0893	0.1005	0.1164	0.1320
$i=5$	0.0893	0.1005	0.1164	0.1320
$i=6$	0.0893	0.1005	0.0365	0.0409
$i=7$	0.0045	0.0050	0.0365	0.0409
$i=8$	0.0045	0.0050	0.0048	0.0053
$i=9$	0.0045	0.0050	0.0048	0.0053
$i=10$	0.3619	0.3698	0.6025	0.6036

These tables contain an overwhelming amount of information. It would be useful to condense this information into a smaller subset of numbers. Recall, in section 3.2 we mentioned that there may be instances when we prefer to use a weighted average of the expected efficiency (see Definition 3.2.5). For example, in our blackjack problem, depending on our hand, it may be very important to discriminate between a deck with an ace missing and a deck with a five missing. However, we may have less need to discriminate between a deck with a four missing and one with a five missing. So we will combine the expected efficiencies, weighting them accordingly. This technique is discussed in the next section.

TABLE 5.4.5 Expected Efficiencies for Uston's APC and Griffin-4

P_{q_0} vs. P_{q_i}	Uston's APC		Griffin-4	
	$E_{q_0}(\text{eff}_{0,i}(T))$	$E_{q_i}(\text{eff}_{0,i}(T))$	$E_{q_0}(\text{eff}_{0,i}(T))$	$E_{q_i}(\text{eff}_{0,i}(T))$
$i=1$	0.0067	0.0075	0.0056	0.0062
$i=2$	0.0186	0.0209	0.0124	0.0139
$i=3$	0.0609	0.0677	0.0369	0.0414
$i=4$	0.0609	0.0677	0.0734	0.0820
$i=5$	0.1164	0.1316	0.1150	0.1303
$i=6$	0.0609	0.0677	0.0734	0.0820
$i=7$	0.0609	0.0677	0.0734	0.0820
$i=8$	0.0186	0.0209	0.0124	0.0139
$i=9$	0.0230	0.0254	0.0141	0.0158
$i=10$	0.5762	0.5800	0.5784	0.5813

5.5 Utility

Utility theory is defined as "...the study of quantitative representations of people's preferences and choices."(Encyclopedia of Statistical Sciences (1989, p. 449)). Typically utilities are a means of applying subjective weights to probabilities. This is a common practice in gambling theory, where decisions are monetarily motivated. For instance, in Blackjack one may wish to wager more or less, depending on the composition of the deck. In this example, there are an infinite number of possible utility functions that could be applied. For our purposes we will chose an objective utility function. (Epstein (1995, p. 222) makes this assumption, also.)

Epstein (1995, p. 243) has calculated the effect on player expectation G of removing a card from a 52-card deck. Those results are given in Table 5.5.1. The full deck expected gain is .10%. We can see that the player benefits if a low card (excluding aces) is removed and the house benefits if an ace or ten is removed from the deck.

The weights we will use in calculating our weighted averages of the expected efficiencies in the Blackjack example will be

$$W_i = \frac{|G_i - G_0|}{\sum_{i=0}^{10} |G_i - G_0|}.$$

The weight for the 52 card deck will be $W_0 = .00$, $W_1 = \frac{|-.48-.10|}{4.24} = .137$ will be the weight for the 51-card deck with an ace missing, etc. Notice that $\sum W_i = 1$.

TABLE 5.5.1 Player Expectation for a Depleted Deck

Card Removed= i	% Expectation (G_i)	W_i $i=0, \dots, 10$
1	-.48	.137
2	.50	.094
3	.56	.108
4	.70	.142
5	.86	.179
6	.56	.108
7	.41	.073
8	.12	.005
9	-.06	.038
10	-.39	.116
None=0	.10	.00

For the calibration efficiencies and the card-counting statistics, the weighted average will be calculated as:

$$\frac{\sum_{j=1}^{10} (W_j E_{q_j}(\text{eff}_{0,j}(X_l)))}{\sum_{j=1}^{10} (W_j)}$$

Table 5.5.2 gives the weighted averages for the calibration problem.

TABLE 5.5.2 Weighted Averages for Calibration Problem

X_l	Weighted Average
X_0 , zero cards known	0.0000
X_1 , one card known	0.0440
X_2 , two cards known	0.1045
X_3 , three cards known	0.1839
X_4 , four cards known	0.2881
X_5 , five cards known	0.4184
X_6 , six cards known	0.5739
X_7 , seven cards known	0.7655
X_8 , eight cards known	1.0000

Table 5.5.3 gives the weighted averages for the card-counting statistics.

TABLE 5.5.3 Weighted Averages for Card-Counting Statistics

Statistic	Weighted Average
Ten Count	0.1499
Hiopt I	0.1245
Braun + -	0.1207
Hiopt II	0.1290
Uston's APC	0.1241
Griffin 4	0.1245

5.5 Conclusion

Though the Ten Count statistic provides the most information about the deck composition based on our weighted measure of efficiency, the amount of information provided is less than knowing three out of the eight cards that have been dealt! Why not go with the simplest statistic in this example? One can argue that the Hiopt I statistic is much simpler to work with, and no information will be lost in this example by using the Hiopt I instead of Griffin-4. In fact, in this example, it would be easier to simply remember the first three cards that have been dealt!

Notice that the blackjack problem is another example of an application where many of the existing methods of measuring information are not useful. Kullback information will be either zero (if we take the expected value with respect to q_1, q_2, \dots , or q_{10}) or it will be ∞ (if we take the expected value with respect to q_0). (See section 2.4 for a discussion of an alternate approach to calculating Kullback information.) Also, Fisher's information does not exist since Condition A in section 2.1 is not met. None of the other methods discussed in Chapter 2 can be easily applied, if at all

Chapter 6

Conclusion

In this paper we have studied various methods for measuring information in statistics; weight of evidence, Fisher's information, Kullback information function, comparison of risk functions, comparison of power functions, deficiency and insufficiency. We have discussed the possible shortcomings or disadvantages of each of these methods. There are areas where these methods can be applied effortlessly, but there are also problems we cannot solve using these standard techniques.

We have developed a new method to measure information in statistics. This method is more general in its application: we can now study information in applications where previously we couldn't. This new method has nice properties that make it easy to use and easy to calibrate. We don't need to make any assumptions about distributions or independence. This is a versatile, self-contained measure of information that should have far-reaching applications.

Chapter 7

Future Research

The new technique introduced in the preceding pages opens the door to a new field of research in the comparison of experiments. Though the technique has broad and far-reaching applications, it had specific uses in this paper. This was sufficient to demonstrate the significant contribution this new method will make, but where do we go from here?

One interesting avenue of research would be to explore different metrics. (See Definition 3.1.3.) We used statistical distance as the metric in our measure of efficiency. How does the expected efficiency behave under different metrics?

In this paper we worked with finite parameter sets. In the blackjack example in Chapter 5 the parameter set consisted of eleven different distributions. The comparison of statistics required the calculation of the expected efficiency for designated pairs of distributions and then the weighted sum over values in the parameter set. Another topic suggesting future research would be the calculation of expected efficiency for infinite parameter sets.

Another topic that could be an entire project in itself is the calculation of the weighted average. As was mentioned in section 3.2 we may wish to use prior probabilities or some other utility. What is appropriate depends entirely on the application. And how could this weighted average be calculated for infinite parameter sets?

The most ambitious project that is derived from this paper involves the calibration procedure. In the examples and applications in this paper, the expected efficiency was an increasing function of the sequence X_l , where $l \leq k$ =sample size (refer to Tables 4.1, 5.4.1, and 5.4.2). Common sense tells us that larger sample sizes give more information about the

parameter space. Hence, the expected efficiency should possess a monotonicity property. More generally, if $T_1(X)$ is a function of $T_2(X)$ then we wish to prove that $E_{q_i}(\text{eff}_{1,2}(T_1(X))) \leq E_{q_i}(\text{eff}_{1,2}(T_2(X)))$. As of yet, the proof of this property has been elusive, but it will be the subject of intensive future research.

In the Blackjack Problem it would be interesting to investigate other sample sizes. For instance, if 26 cards were dealt instead of eight, and the card-counting statistics were calculated, would they still contain less information than 3/8 of the 26 cards?

A new area of research deals with the question of how to handle very large data sets. How can this new method be applied to this area? Can we use it in data mining? Can we use it to determine the best way to group large data sets to make them easier to analyze? This new method unveils a list of questions for every question answered. It will be a rich area of research in the future.

References

- Ash, R. (1972). *Real Analysis and Probability*. Academic Press, Inc. New York.
- Bahadur, R. R. (1954). "Sufficiency and statistical decision functions." *Ann. Math. Statist.*
Vol. 25, pp.423-62.
- Billingsley, P. (1986). *Probability and Measure*. John Wiley and Sons, Inc, New York.
- Blackwell, D. (1951). "Comparison of experiments." In *Proc. Second Berkeley Symposium on Mathematical Statistics and Probability*, Univ. of California Press, Berkeley, pp. 93-102.
- Denbigh, and Denbigh (1985). *Entropy in Relation to Incomplete Knowledge*, Cambridge University Press, New York.
- Encyclopedia of Statistical Sciences*, Vols. 1 and 9, (1989) John Wiley and Sons, New York.
- Epstein, R.A. (1995). *The Theory of Gambling and Statistical Logic*, Academic Press, New York.
- Fisher, R. A. (1925). "Theory of statistical estimation." *Proc. Camb. Phil. Soc.*, Vol. 22, pp. 700-725.
- Good, I. J. (1950). *Probability and the Weighing of Evidence*, Charles Griffin, London.
- Good, I. J. (1983). *Good Thinking*, University of Minnesota Press, Minnesota.
- Kullback, S. and Leibler, R. A. (1951). "On information and sufficiency." *Ann. Math. Statist.*,
Vol. 22, pp. 79-86.
- Kullback, S. (1968). *Information Theory and Statistics*, Dover Publications, Inc., New York.
- Le Cam, L. (1964). "Sufficiency and approximate sufficiency." *Ann. Math. Statist.*, Vol. 35,
pp. 1419-1455.

- Le Cam, L. (1974). "On the information contained in additional observations." *Ann. Statist.*, Vol. 2, pp. 630-649.
- Lee, Peter M. (1989). *Bayesian Statistics: An Introduction*, Oxford University Press, New York.
- Lehmann, E. L. (1986). *Testing Statistical Hypothesis*, John Wiley and Sons, New York.
- Mammen, E. (1986). "The statistical information contained in additional observations." *Ann. Statist.* Vol. 14, pp. 665-678.
- Nikitin, Yakov. (1995). *Asymptotic Efficiency of Nonparametric Tests*, Cambridge University Press, New York.
- Osteyee, D. B., and Good, I. J. (1974) "Information, Weight of Evidence, the Singularity between Probability Measures and Signal Detection." *Lecture Notes in Mathematics*. Vol. 376. Springer-Verlag, Germany.
- Savage, L. J. (1954). *The Foundations of Statistics*, John Wiley and Sons, New York.
- Shannon, C. E. (1948). "A mathematical theory of communication." *Bell System Tech. J.* Vol. 27, pp. 379-423.
- Rohatgi, V. K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*, John Wiley and Sons, Inc., New York.
- Torgersen, E. (1991). *Comparison of Statistical Experiments*, Cambridge University Press, Cambridge.
- Zacks, S. (1971). *The Theory of Statistical Inference*, John Wiley and Sons, Inc., New York.

Appendix A

Program Notes

Programs are included for calculating the expected efficiencies for the Hiopt II statistic, Uston's APC statistic, and the calibration components. The programs are written in FORTRAN. Several checks are built into each program to make sure they are running properly. The checks consisted of summing several of the variables over all possible hands and making sure they sum to one. A third program was used to verify the value of the unconditional differences for the calibration program. This program is not included.

Several problems were encountered while running the calibration program. Due to the large number of nested DO loops, the program uses a great deal of time and money to run for large values of K and C. (Recall, K=number of cards dealt and C=number of cards known.) Also, the rounding error rate is fairly large (third and fourth decimal place) for large values of K. Hence, a value of K=8 seemed a fair compromise. Given more resources, it would be extremely interesting to run these programs for larger values of K, say K=26.

Appendix B

Hiopt and Uston Program

This program can be used for any pair of card-counting statistics.

```
C  THIS PROGRAM WILL CALCULATE THE EXPECTED EFFICIENCIES OF TWO
C  CARD COUNTING STATISTICS. TO USE ANOTHER STATISTIC SIMPLY REPLACE THE
C  FORMULAS FOR THE USTON AND HIOPT VARIABLES, AND ADJUST TO BOUNDS OF THE
C  ARRAY TO ACCOMMODATE THE NEW STATISTICS. WE WILL ASSUME 8 CARDS HAVE
C  BEEN DEALT FROM A DECK COMPOSITION WE WISH TO DETERMINE.
C  THE FIRST DISTRIBUTION CONSISTS OF A 52 CARD DECK.
C  THE SECOND DISTRIBUTION CONSISTS OF A 51 CARD DECK (A CARD IS
C  MISSING).
C  (A) WE WILL CALCULATE THE STATISTICAL DISTANCE
C  BETWEEN THESE TWO DISTRIBUTIONS FOR EACH POSSIBLE HAND AND SUM
C  THAT DISTANCE OVER ALL POSSIBLE HANDS. THIS IS THE UNCONDITIONAL
C  DISTANCE.
C  (B) WE WILL CALCULATE THE CONDITIONAL DISTRIBUTION OF EACH
C  POSSIBLE HAND, GIVEN THE CARD COUNTING STATISTIC EQUALS SOME
C  VALUE. THEN WE WILL CALCULATE THE STATISTICAL DISTANCE BETWEEN
C  THESE CONDITIONAL DISTRIBUTIONS.
C  (C) NEXT WE WILL CALCULATE THE EXPECTED VALUE OF THIS
C  CONDITIONAL DISTRIBUTION, GIVEN THE VALUE OF THE CARD COUNTING
C  STATISTIC, AND SUM OVER ALL VALUES OF THE CARD COUNTING STATISTIC.
C  (D) LASTLY WE WILL CALCULATE THE EXPECTED EFFICIENCY OF EACH
C  CARD COUNTING STATISTIC UNDER EACH DISTRIBUTION BY THE FOLLOWING
C  FORMULA:
C
C  EFF=( (UNCONDITIONAL - EXPECTED CONDITIONAL) / UNCONDITIONAL)
C
C  ## NOTE: ANYTIME TENS ARE REFERRED TO INCLUDES TENS AND FACE
C  CARDS.
C  VARIABLES USED IN THIS PROGRAM ARE:
C  BNC(N,K)=AN ARRAY CONTAINING BINOMIAL COEFFICIENTS
```

C K=THE NUMBER OF CARDS DEALT
 C USTON=THE VALUE OF USTON'S APC FOR A GIVEN SET OF CARDS
 C
$$=(\# \text{ TWOS}+\# \text{ EIGHTS})+2(\# \text{ THREES}+\# \text{ FOURS}+\# \text{ SIXES}+\# \text{ SEVENS})$$

 C
$$+3(\# \text{ FIVES})-(\# \text{ NINES})-3(\# \text{ TENS})$$

 C HIOPT=THE VALUE OF HIOPT II FOR A GIVEN SET OF CARDS
 C
$$=(\# \text{ TWOS}+\# \text{ THREES}+\# \text{ SIXES}+\# \text{ SEVENS})+2(\# \text{ FOURS}+\# \text{ FIVES})$$

 C
$$-2(\# \text{ TENS})$$

 C PROBHA1=PROBABILITY OF THE HAND GIVEN A 52 CARD DECK
 C PROBHA2=PROBABILITY OF THE HAND GIVEN A 51 CARD DECK
 C DIFF=ABSOLUTE VALUE OF DIFFERENCE BETWEEN PROBHA1 AND
 C PROHA2
 C DIFF2=ABSOLUTE VALUE OF THE CONDITIONAL DIFFERENCE FOR THE HIOPT
 C DIFF3=ABSOLUTE VALUE OF THE CONDITIONAL DIFFERENCE FOR USTON'S
 C PRO1HI(HIOPT)=AN ARRAY CONTAINING PROBABILITIES OF THE HIOPT
 C STATISTIC GIVEN 52 CARDS
 C PRO2HI(HIOPT)=AN ARRAY CONTAINING PROBABILITIES OF THE HIOPT
 C STATISTIC GIVEN 51 CARDS
 C PRO1US(USTON)=AN ARRAY CONTAINING PROBABILITIES OF THE USTON
 C STATISTIC GIVEN 52 CARDS
 C PRO2US(USTON)=AN ARRAY CONTAINING PROBABILITIES OF THE USTON
 C STATISTIC GIVEN 51 CARDS
 C CHECK1=SUM OF PROBHA1, SHOULD EQUAL 1
 C CHECK2=SUM OF PROBHA2, SHOULD EQUAL 1
 C UNCOND=SUM OF UNCONDITIONAL DIFFERENCES
 C CONDHI(HIOPT)=AN ARRAY CONTAINING THE ABSOLUTE VALUES OF THE
 C DIFFERENCES IN CONDITIONAL DISTRIBUTIONS FOR EACH VALUE
 C OF THE HIOPT
 C CONDUS(USTON)=AN ARRAY CONTAINING THE ABSOLUTE VALUES OF
 C THE DIFFERENCES IN CONDITIONAL DISTRIBUTIONS FOR EACH
 C VALUE OF USTON
 C EXP1HI=SUM OF EXPECTED VALUE OF THE CONDITIONAL DISTANCES
 C GIVEN THE HIOPT, GIVEN 52 CARDS
 C EXP2HI=SUM OF EXPECTED VALUE OF THE CONDITIONAL DISTANCES

```

C      GIVEN THE HIOPT, GIVEN 51 CARDS
C      EXP1US=SUM OF EXPECTED VALUE OF THE CONDITIONAL DISTANCES
C      GIVEN THE USTON, GIVEN 52 CARDS
C      EXP2US=SUM OF EXPECTED VALUE OF THE CONDITIONAL DISTANCES
C      GIVEN THE USTON, GIVEN 51 CARDS
C      EFF1HI=EXPECTED EFFICIENCY OF THE HIOPT GIVEN 52 CARDS
C      EFF2HI=EXPECTED EFFICIENCY OF THE HIOPT GIVEN 51 CARDS
C      EFF1US=EXPECTED EFFICIENCY OF USTON'S GIVEN 52 CARDS
C      EFF2US=EXPECTED EFFICIENCY OF USTON'S GIVEN 51 CARDS
C      I1=NUMBER OF ACES, I2=NUMBER OF TWOS,
C      I3=NUMBER OF THREES, I4=NUMBER OF FOURS,
C      I5=NUMBER OF FIVES, I6=NUMBER OF SIXES,
C      I7=NUMBER OF SEVENS, I8=NUMBER OF EIGHTS,
C      I9=NUMBER OF NINES, I10=NUMBER OF TENS AND FACE CARDS
C
C LET'S DEFINE VARIABLES AND ARRAYS
C
      DIMENSION BNC(0:52,0:52),PRO1HI(-30:30),PRO2HI(-30:30)
      DIMENSION PRO1US(-30:30),PRO2US(-30:30),CONDHI(-30:30)
      DIMENSION CONDUS(-30:30)
      DOUBLE PRECISION PROHA1,PROHA2,DIFF,UNCOND,EXP1HI,
$      EXP2HI,EXP1US,EXP2US,EFF1HI,EFF2HI,EFF1US,EFF2US,
$      CHECK1,CHECK2,DIFF2,DIFF3
      INTEGER USTON,HIOPT
C
C NOW LET'S CALCULATE THE BINOMIAL COEFFICIENTS AND STORE THEM IN
C ARRAY BNC(N,K). WE WILL USE THESE TO CALCULATE PROBHAND.
C
      DO 2 N=0,52
        DO 1 K=1,52
          BNC(N,K)=0.0
1      CONTINUE
2      CONTINUE

```

```

        DO 3 N=0,52
            BNC(N,0)=1.0
            BNC(1,1)=1.0
3       CONTINUE
        DO 5 N=2,52
            DO 4 K=1,N
                BNC(N,K)=BNC(N-1,K)+BNC(N-1,K-1)
4       CONTINUE
5       CONTINUE
C
C NOW INITIALIZE VARIABLES
C
        K=8
        CHECK1=0
        CHECK2=0
        UNCOND=0
        EXP1HI=0
        EXP2HI=0
        EXP1US=0
        EXP2US=0
        EFF1HI=0
        EFF2HI=0
        EFF1US=0
        EFF2US=0
C
        DO 100 I=-30,30
            PRO1HI(I)=0
            PRO2HI(I)=0
            PRO1US(I)=0
            PRO2US(I)=0
            CONDHI(I)=0
            CONDUS(I)=0
100    CONTINUE

```


C

C NOW LOOP THROUGH ALL POSSIBLE COMBINATIONS OF CARDS, KEEPING ONLY
C K CARD COMBINATIONS, CALCULATING PROBABILITIES FOR VALUES OF THE
C STATISTICS

DO 190 I1=0,4

DO 180 I2=0,4

DO 170 I3=0,4

DO 160 I4=0,4

DO 150 I5=0,4

DO 140 I6=0,4

DO 130 I7=0,4

DO 120 I8=0,4

DO 110 I9=0,4

DO 1000 I10=0,16

C

C USE ONLY COMBINATIONS OF K CARDS

C

T=I1+I2+I3+I4+I5+I6+I7+I8+I9+I10

IF (T.EQ.K) THEN

C

C CALCULATE USTON'S

C

USTON=I2+I8+((I3+I4+I6+I7)*2)+(I5*3)+(I9*(-1))+(I10*(-3))

C

C CALCULATE HIOPT

C

HIOPT=I2+I3+I6+I7+2*(I4+I5)-2*(I10)

C

C CALCULATE THE PROBABILITY OF THE HAND GIVEN 52 CARDS

C

PROHA1=(BNC(4,I1)*BNC(4,I2)*BNC(4,I3)*BNC(4,I4)*

\$ BNC(4,I5)*BNC(4,I6)*BNC(4,I7)*BNC(4,I8)*

\$ BNC(4,I9)*BNC(16,I10))/BNC(52,K)

```

C
C  CALCULATE THE PROBABILITY OF THE HAND GIVEN A CARD IS MISSING. IF
C  USING THE DISTRIBUTION WITH AN ACE MISSING SET INEQUALITY TO: IF
C  (I1.LE.3) AND SET THE BINOMIAL PART FOR I1 TO BNC(3,I1). REPEAT FOR 2
C  THROUGH 9. IF USING THE DISTRIBUTION WITH A TEN MISSING, SET
C  INEQUALITY TO: IF (I10.LE.15) AND SET BINOMIAL PART FOR I10 TO BNC(15,I10).
C
    IF (I4.LE.3) THEN
C
    PROHA2=(BNC(4,I1)*BNC(4,I2)*BNC(4,I3)*BNC(3,I4)*
$      BNC(4,I5)*BNC(4,I6)*BNC(4,I7)*BNC(4,I8)*
$      BNC(4,I9)*BNC(16,I10))/BNC(51,K)
C
    ELSE
    PROHA2=0
    ENDIF
C
    CHECK1=CHECK1+PROHA1
    CHECK2=CHECK2+PROHA2
C
C  STORE PROBABILITIES FOR EACH VALUE OF HIOPT AND USTON UNDER BOTH
C  DISTRIBUTIONS TO USE IN CALCULATING THE CONDITIONAL PROBABILITIES.
C
    PRO1HI(HIOPT)=PRO1HI(HIOPT)+PROHA1
    PRO1US(USTON)=PRO1US(USTON)+PROHA1
    PRO2HI(HIOPT)=PRO2HI(HIOPT)+PROHA2
    PRO2US(USTON)=PRO2US(USTON)+PROHA2
C
    ELSE
    GO TO 200
200  CONTINUE
    ENDIF
C

```

```

1000    CONTINUE
110     CONTINUE
120     CONTINUE
130     CONTINUE
140     CONTINUE
150     CONTINUE
160     CONTINUE
170     CONTINUE
180     CONTINUE
190     CONTINUE

```

C

C NOW LOOP THROUGH ALL POSSIBLE COMBINATIONS OF CARDS, KEEPING ONLY
C K CARD COMBINATIONS, CALCULATING THE DIFFERENCES IN DISTRIBUTIONS

C

```

DO 19 I1=0,4
DO 18 I2=0,4
DO 17 I3=0,4
DO 16 I4=0,4
DO 15 I5=0,4
DO 14 I6=0,4
DO 13 I7=0,4
DO 12 I8=0,4
DO 11 I9=0,4
DO 10 I10=0,16

```

C

C USE ONLY COMBINATIONS OF K CARDS

C

```

T=I1+I2+I3+I4+I5+I6+I7+I8+I9+I10
IF (T.EQ.K) THEN

```

C

C CALCULATE USTON'S

C

```

USTON=I2+I8+((I3+I4+I6+I7)*2)+(I5*3)+(I9*(-1))+(I10*(-3))

```

C

C CALCULATE HIOPT

C

HIOPT=I2+I3+I6+I7+2*(I4+I5)-2*(I10)

C

C CALCULATE THE PROBABILITY OF THE HAND GIVEN 52 CARDS

C

PROHA1=(BNC(4,I1)*BNC(4,I2)*BNC(4,I3)*BNC(4,I4)*

\$ BNC(4,I5)*BNC(4,I6)*BNC(4,I7)*BNC(4,I8)*

\$ BNC(4,I9)*BNC(16,I10))/BNC(52,K)

C

C CALCULATE THE PROBABILITY OF THE HAND GIVEN A CARD IS MISSING.

C

IF (I4.LE.3) THEN

C

PROHA2=(BNC(4,I1)*BNC(4,I2)*BNC(4,I3)*BNC(3,I4)*

\$ BNC(4,I5)*BNC(4,I6)*BNC(4,I7)*BNC(4,I8)*

\$ BNC(4,I9)*BNC(16,I10))/BNC(51,K)

C

ELSE

PROHA2=0

ENDIF

C

DIFF=ABS(PROHA1-PROHA2)

IF (I4.LE.3) THEN

DIFF2=ABS(PROHA1/PRO1HI(HIOPT)-PROHA2/PRO2HI(HIOPT))

DIFF3=ABS(PROHA1/PRO1US(USTON)-PROHA2/PRO2US(USTON))

ELSE

DIFF2=0

DIFF3=0

ENDIF

C

UNCOND=UNCOND+DIFF

```

CONDHI(HIOPT)=CONDHI(HIOPT)+DIFF2
CONDUS(USTON)=CONDUS(USTON)+DIFF3
C
ELSE
GO TO 20
20  CONTINUE
    ENDIF
C
10  CONTINUE
11  CONTINUE
12  CONTINUE
13  CONTINUE
14  CONTINUE
15  CONTINUE
16  CONTINUE
17  CONTINUE
18  CONTINUE
19  CONTINUE
C
C NOW MULTIPLY ARRAYS TO GET EXPECTED VALUES, USING ONLY VALUES OF
C THE STATISTICS POSSIBLE UNDER BOTH DISTRIBUTIONS.
C
DO 21 I=-16,16
    EXP1HI=CONDHI(I)*PRO1HI(I)+EXP1HI
    EXP2HI=CONDHI(I)*PRO2HI(I)+EXP2HI
21  CONTINUE
C
DO 22 I=-24,20
    EXP1US=CONDUS(I)*PRO1US(I)+EXP1US
    EXP2US=CONDUS(I)*PRO2US(I)+EXP2US
22  CONTINUE
C
C NOW CALCULATE THE EFFICIENCIES

```

C

```
EFF1HI=(UNCOND-EXP1HI)/UNCOND  
EFF2HI=(UNCOND-EXP2HI)/UNCOND  
EFF1US=(UNCOND-EXP1US)/UNCOND  
EFF2US=(UNCOND-EXP2US)/UNCOND
```

C

C NOW PRINT THE RESULTS

C

```
WRITE(9,*)'P0 VS P4'  
WRITE(9,*)'EFF1HIOP=',EFF1HI,' EFF2HIOP=',EFF2HI  
WRITE(9,*)'EFF1UST=',EFF1US,' EFF2UST=',EFF2US  
WRITE(9,*)'CHECK1=',CHECK1,' CHECK2=',CHECK2
```

C

END

Appendix C

Calibration Program

C THIS PROGRAM WILL CALCULATE THE EXPECTED EFFICIENCIES FOR
C CALIBRATION IN THE BLACKJACK EXAMPLE. WE WILL ASSUME THAT K
C CARDS HAVE BEEN DEALT. THE FIRST DISTRIBUTION CONSISTS OF A 52 CARD
C DECK. THE SECOND DISTRIBUTION CONSISTS OF A 51 CARD DECK (ONE CARD
C MISSING).
C
C (A) WE WILL CALCULATE THE STATISTICAL DISTANCE
C BETWEEN THESE TWO DISTRIBUTIONS FOR EACH POSSIBLE HAND AND SUM
C THAT DISTANCE OVER ALL POSSIBLE HANDS. THIS IS THE UNCONDITIONAL
C DISTANCE.
C
C (B) WE WILL CALCULATE THE CONDITIONAL DISTRIBUTION OF EACH
C POSSIBLE HAND, GIVEN THAT C=I OF THE K CARDS ARE KNOWN. THEN WE
C WILL CALCULATE THE STATISTICAL DISTANCE BETWEEN THESE CONDITIONAL
C DISTRIBUTIONS. (THIS PROGRAM IS RERUN FOR DIFFERENT VALUES OF C.)
C
C (C) NEXT WE WILL CALCULATE THE EXPECTED VALUE OF THIS CONDITIONAL
C DISTRIBUTION, GIVEN THAT C=I OF THE K CARDS ARE KNOWN, AND
C SUM OVER ALL VALUES OF THE C CARDS.
C
C (D) LASTLY WE WILL CALCULATE THE EXPECTED EFFICIENCY OF EACH
C VALUE OF C UNDER EACH DISTRIBUTION BY THE FOLLOWING
C FORMULA:
C $EFF = (UNCONDITIONAL - EXPECTED\ CONDITIONAL) / UNCONDITIONAL$
C
C ## NOTE: ANYTIME TENS ARE REFERRED TO INCLUDES TENS AND FACE

C CARDS.

C

C VARIABLES USED IN THIS PROGRAM ARE:

C BNC(N,K)=AN ARRAY CONTAINING BINOMIAL COEFFICIENTS

C PROBH1=PROBABILITY OF THE K CARDS BEING GIVEN A 52 CARD DECK

C PROBH2=PROBABILITY OF THE K CARDS BEING GIVEN A 51 CARD DECK

C WITH A 2,3,4,5,6,7,8, OR 9 MISSING

C PROBH10=PROBABILITY OF THE K CARDS BEING GIVEN A 51 CARD DECK

C WITH A 10 MISSING.

C DIF2=CONDITIONAL DIFFERENCES BETWEEN 52 CARD DISTRIBUTION AND

C 51 CARD DISTRIBUTION WITH 2,3,4,5,6,7,8, OR 9 MISSING

C DIF10=CONDITIONAL DIFFERENCES BETWEEN 52 CARD DISTRIBUTION

C AND 51 CARD DISTRIBUTION WITH 10 MISSING

C UNCO1=SUM OF DIF2

C UNCO10=SUM OF DIF10

C EXP1=EXPECTED VALUE OF CONDITIONAL DIFFERENCES BETWEEN 52

C CARD DECK AND 51 CARD DECK WITH 2,3,4,5,6,7,8, OR 9 MISSING WITH

C RESPECT TO 52 CARD DECK.

C EXP2=EXPECTED VALUE OF CONDITIONAL DIFFERENCES BETWEEN 52

C CARD DECK AND 51 CARD DECK WITH 2,3,4,5,6,7,8, OR 9 MISSING WITH

C RESPECT TO 51 CARD DECK.

C EXP10A=EXPECTED VALUE OF CONDITIONAL DIFFERENCES BETWEEN 52

C CARD DECK AND 51 CARD DECK WITH 10 MISSING WITH

C RESPECT TO 52 CARD DECK.

C EXP10B=EXPECTED VALUE OF CONDITIONAL DIFFERENCES BETWEEN 52

C CARD DECK AND 51 CARD DECK WITH 10 MISSING WITH RESPECT TO

C 51 CARD DECK.

C EFF1=EXPECTED EFFICIENCY,52 CARD DECK VS 51 CARD DECK WITH A

C NON- TEN MISSING, WRT 52 CARD DECK.

C EFF2=EXPECTED EFFICIENCY,52 CARD DECK VS 51 CARD DECK WITH A

C NON- TEN MISSING WRT 51 CARD DECK.

C EFF10A=EXPECTED EFFICIENCY,52 CARD DECK VS 51 CARD DECK WITH

C TEN MISSING, WRT 52 CARD DECK.


```

C     EFF10B=EXPECTED EFFICIENCY,52 CARD DECK VS 51 CARD DECK WITH
C     TEN MISSING, WRT 51 CARD DECK.
C     PROB1, PROB2, PROB10=PROBABILITY OF THE K-C CARDS GIVEN
C     RESPECTIVE DISTRIBUTIONS
C     PR1CH, PR2CH, PR10CH=SUM OF PROB1, PROB2, PROB10 OVER ALL
C     HANDS. SHOULD SUM TO 1.
C     PROT1, PROT2, PROT10=PROBABILITY OF C CARDS FOR THE RESPECTIVE
C     DISTRIBUTIONS
C     CHECK1, CHECK2, CHCK10=SUMS OF PROBABILITIES FOR C CARDS OVER
C     ALL POSSIBLE HANDS FOR GIVEN DISTRIBUTION, EQUALS 1
C     RV1, RV10=SUM OF DIF2 AND DIF10 FOR A GIVEN VALUE OF C
C     K=NUMBER OF CARDS DEALT
C     C=NUMBER OF K CARDS KNOWN
C     I1=NUMBER OF ACES KNOWN, I2=NUMBER OF TWOS KNOWN,
C     I3=NUMBER OF THREES KNOWN, I4=NUMBER OF FOURS KNOWN,
C     I5=NUMBER OF FIVES KNOWN, I6=NUMBER OF SIXES KNOWN,
C     I7=NUMBER OF SEVENS KNOWN, I8=NUMBER OF EIGHTS KNOWN,
C     I9=NUMBER OF NINES KNOWN,
C     I10=NUMBER OF TENS AND FACE CARDS KNOWN
C
C     J1=NUMBER OF I1 ACES UNKNOWN, J2=NUMBER OF I2 TWOS UNKNOWN,
C     J3=NUMBER OF I3 THREES UNKNOWN, J4=NUMBER OF I4 FOURS UNKNOWN,
C     J5=NUMBER OF I5 FIVES UNKNOWN, J6=NUMBER OF I6 SIXES UNKNOWN,
C     J7=NUMBER OF I7 SEVENS UNKNOWN, J8=NUMBER OF I8 EIGHTS UNKNOWN,
C     J9=NUMBER OF I9 NINES UNKNOWN,
C     J10=NUMBER OF I10 TENS AND FACE CARDS UNKNOWN
C
C LET'S DEFINE VARIABLES AND ARRAYS
      DIMENSION BNC(0:52,0:52)
      DOUBLE PRECISION PROH1,CHECK1,EXP1,CH1,CH2,
$     EXP2,EFF1,EFF2,DIF2,PROB1,PROB2,RV1,
$     PR1CH,PR2CH,PROT1,PROT2,PROH2,CHECK2,
$     UNCO1,UNCO10,CHCK10,PROT10,PROB10,PROH10,

```

```

$   EXP10A,EXP10B,EFF10A,EFF10B,RV10,CH10,
$   PR10CH,DIF10
INTEGER K,C
C
C NOW LET'S CALCULATE THE BINOMIAL COEFFICIENTS AND STORE THEM IN
C ARRAY BNC(N,K). WE WILL USE THESE TO CALCULATE PROBABILITIES.
C
      DO 2 N=0,52
        DO 1 K=1,52
          BNC(N,K)=0.0
1        CONTINUE
2      CONTINUE
      DO 3 N=0,52
        BNC(N,0)=1.0
        BNC(1,1)=1.0
3      CONTINUE
      DO 5 N=2,52
        DO 4 K=1,N
          BNC(N,K)=BNC(N-1,K)+BNC(N-1,K-1)
4      CONTINUE
5      CONTINUE
C
C NOW INITIALIZE VARIABLES
C
      K=8
      C=7
      UNCO1=0
      UNCO10=0
      RV1=0
      RV10=0
      CHECK1=0
      CHECK2=0
      CHCK10=0

```

```

PR1CH=0
PR2CH=0
PR10CH=0
EXP1=0
EXP2=0
EXP10A=0
EXP10B=0
EFF1=0
EFF2=0
EFF10A=0
EFF10B=0
CH1=0
CH2=0
CH10=0
C
  WRITE(9,*)'K=',K,'C=',C
C
C NOW LOOP THROUGH ALL POSSIBLE COMBINATIONS OF CARDS, KEEPING ONLY
C C CARD COMBINATIONS
C
  DO 19 I1=0,4
    DO 18 I2=0,4
      DO 17 I3=0,4
        DO 16 I4=0,4
          DO 15 I5=0,4
            DO 14 I6=0,4
              DO 13 I7=0,4
                DO 12 I8=0,4
                  DO 11 I9=0,4
                    DO 10 I10=0,16
C
C USE ONLY COMBINATIONS OF C CARDS
C

```

```

T=I1+I2+I3+I4+I5+I6+I7+I8+I9+I10
IF (T.EQ.C) THEN
C
C CALCULATE THE PROBABILITY OF THE C CARDS UNDER EACH DISTRIBUTION
C
  PROT10=(BNC(4,I1)*BNC(4,I2)*BNC(4,I3)*BNC(4,I4)*
$    BNC(4,I5)*BNC(4,I6)*BNC(4,I7)*BNC(4,I8)*
$    BNC(4,I9)*BNC(15,I10))/(BNC(51,C) )
C
  CHCK10=CHCK10+PROT10
C
  PROT1=(BNC(4,I1)*BNC(4,I2)*BNC(4,I3)*BNC(4,I4)*
$    BNC(4,I5)*BNC(4,I6)*BNC(4,I7)*BNC(4,I8)*
$    BNC(4,I9)*BNC(16,I10))/(BNC(52,C) )
C
  CHECK1=CHECK1+PROT1
C
  IF (I1.LE.3) THEN
C
    PROT2=(BNC(3,I1)*BNC(4,I2)*BNC(4,I3)*BNC(4,I4)*
$    BNC(4,I5)*BNC(4,I6)*BNC(4,I7)*BNC(4,I8)*
$    BNC(4,I9)*BNC(16,I10))/(BNC(51,C) )
C
    ELSE
    PROT2=0
    ENDIF
C
    CHECK2=CHECK2+PROT2
C
C NOW CALCULATE THE EXPECTED CONDITIONAL DIFFERENCES BY LOOPING
C THROUGH
C ALL POSSIBLE K-C COMBINATIONS.
C

```

```

DO 29 J1=0,4-I1
DO 28 J2=0,4-I2
DO 27 J3=0,4-I3
DO 26 J4=0,4-I4
DO 25 J5=0,4-I5
DO 24 J6=0,4-I6
DO 23 J7=0,4-I7
DO 22 J8=0,4-I8
DO 21 J9=0,4-I9
DO 20 J10=0,16-I10
C
      H=J1+J2+J3+J4+J5+J6+J7+J8+J9+J10
      IF (H.EQ.K-C) THEN
C
C  CALCULATE THE PROBABILITIES OF THE REMAINING K-C CARDS,
C  GIVEN ONE OF THE DISTRIBUTIONS
C
      PROB10=(BNC(4-I1,J1)*BNC(4-I2,J2)*BNC(4-I3,J3)*
$ BNC(4-I4,J4)*BNC(4-I5,J5)*BNC(4-I6,J6)*BNC(4-I7,J7)
$ *BNC(4-I8,J8)*BNC(4-I9,J9)*BNC(15-I10,J10))/
$(BNC(51-C,K-C))
C
      PROH10=PROB10*PROT10
C
      PROB1=(BNC(4-I1,J1)*BNC(4-I2,J2)*BNC(4-I3,J3)*
$ BNC(4-I4,J4)*BNC(4-I5,J5)*BNC(4-I6,J6)*BNC(4-I7,J7)
$ *BNC(4-I8,J8)*BNC(4-I9,J9)*BNC(16-I10,J10))/
$(BNC(52-C,K-C) )
C
      PROH1=PROB1*PROT1
C
      IF (I1+J1.LE.3) THEN
C

```

```

PROB2=(BNC(3-I1,J1)*BNC(4-I2,J2)*BNC(4-I3,J3)*
$ BNC(4-I4,J4)*BNC(4-I5,J5)*BNC(4-I6,J6)*BNC(4-I7,J7)
$ *BNC(4-I8,J8)*BNC(4-I9,J9)*BNC(16-I10,J10))/
$(BNC(51-C,K-C) )
C
ELSE
PROB2=0
ENDIF
C
PROH2=PROB2*PROT2
C
C CALCULATE CHECKS TO MAKE SURE PROGRAM IS RUNNING CORRECTLY
C
PR1CH=PR1CH+PROB1
PR2CH=PR2CH+PROB2
PR10CH=PR10CH+PROB10
CH1=CH1+PROH1
CH2=CH2+PROH2
CH10=CH10+PROH10
C
IF (I1.LE.3) THEN
DIF2=ABS((PROB1)-(PROB2))
ELSE
DIF2=0
ENDIF
C
UNCO1=UNCO1+(ABS(PROH1-PROH2))
RV1=RV1+DIF2
C
DIF10=ABS((PROB1)-(PROB10))
C
UNCO10=UNCO10+(ABS(PROH1-PROH10))
RV10=RV10+DIF10

```

```

C
    ELSE
    GO TO 100
100  CONTINUE
    ENDIF

C
20    CONTINUE
21    CONTINUE
22    CONTINUE
23    CONTINUE
24    CONTINUE
25    CONTINUE
26    CONTINUE
27    CONTINUE
28    CONTINUE
29    CONTINUE

C
C  CALCULATE EXPECTED CONDITIONAL DIFFERENCES
C
    EXP1=EXP1+(RV1 *PROT1)
    EXP2=EXP2+(RV1 *PROT2)
    EXP10A=EXP10A+(RV10 *PROT1)
    EXP10B=EXP10B+(RV10 *PROT10)
    RV1=0
    RV10=0

C
    ELSE
    GO TO 200
200  CONTINUE

C
    ENDIF

C
10    CONTINUE

```

11 CONTINUE

12 CONTINUE

13 CONTINUE

14 CONTINUE

15 CONTINUE

16 CONTINUE

17 CONTINUE

18 CONTINUE

19 CONTINUE

C

C NOW CALCULATE THE EFFICIENCIES

C

$EFF1 = (UNCO1 - EXP1) / UNCO1$

$EFF2 = (UNCO1 - EXP2) / UNCO1$

$EFF10A = (UNCO10 - EXP10A) / UNCO10$

$EFF10B = (UNCO10 - EXP10B) / UNCO10$

C

C NOW PRINT THE RESULTS

C

 WRITE(9,*)'PR1CH=',PR1CH,' PR2CH=',PR2CH

 WRITE(9,*)'CHECK1=',CHECK1,' CHECK2=',CHECK2

 WRITE(9,*)'PR10CH=',PR10CH,' CHCK10=',CHCK10

 WRITE(9,*)'CH1=',CH1,' CH2=',CH2

 WRITE(9,*)'CH10=',CH10

 WRITE(9,*)'EFF1=',EFF1,' EFF2=',EFF2

 WRITE(9,*)'EFF10A=',EFF10A,' EFF10B=',EFF10B

 WRITE(9,*)'UNCO1=',UNCO1,' UNCO10=',UNCO10

C

END

VITA

Patricia Lee Kitchin, daughter of Thurman Delna Kitchin and Judith Reynolds Stover, was born May 26, 1965, in Orlando, Florida. She earned an Associates of Arts degree from Santa Fe Community College in 1985. She continued her studies at the University of South Alabama in Mobile, Alabama, where she received a Bachelor of Science degree in Mathematics in 1987 and a Master of Science degree in Mathematics in 1989. She then enrolled at Virginia Polytechnic Institute and State University in 1989, where she received a Master of Science degree in Statistics in 1991. While working on her Doctor of Philosophy degree she worked as a part-time statistician for Corning, Inc. in Blacksburg, Virginia, and a part time instructor for Virginia Polytechnic Institute and State University and Virginia Western Community College. She also spent a year working as a full-time statistician for Bayer Pharmaceuticals in West Haven, Connecticut, and a year working as a full-time bookkeeper for Blue Ridge Outdoors in Blacksburg, Virginia. She has one daughter, Jennifer Lee Jones, born March 25, 1983. In May, 1997, she was remarried to John Mark Leach. She accepted a position as a Visiting Assistant Professor in the Management Science Department in the R. B. Pamplin School of Business at Virginia Polytechnic Institute and State University beginning August 15, 1997.