A possible foundation for Blackwell's Equivalence

Rafael B. Stern and Carlos A. de B. Pereira

Citation: AIP Conf. Proc. **1073**, 90 (2008); doi: 10.1063/1.3039027
View online: http://dx.doi.org/10.1063/1.3039027
View Table of Contents: http://proceedings.aip.org/dbt/dbt.jsp?KEY=APCPCS&Volume=1073&Issue=1
Published by the American Institute of Physics.

A possible foundation for Blackwell's Equivalence

Rafael B. Stern and Carlos A. de B. Pereira

Instituto de Matemática e Estatística - Universidade de São Paulo - Brasil
cpereira@ime.usp.br, rstern@ime.usp.br

Abstract. The Sufficiency Principle states that it is equal for the Statistician to observe the full data or a sufficient statistic. Nevertheless, it can only be used as a criterion for the comparison of random functions which are coupled on the same measure space. On the other hand, when analyzing experiments usually only their marginal distributions are given. Thus the Sufficiency Principle alone, in general, is not enough to compare them. In this article we show that Blackwell's Equivalence relation between experiments is equal to the Sufficiency Principle and the Coupling Invariance Principle which are, in some manner, weaker than the Likelihood Principle.

Keywords: Experimental Design; Blackwell Equivalence; Sufficiency Principle.
PACS: 02.50.-r, 02.50.Ga, 02.50.Tt.

1. INTRODUCTION

A principle which is usually accepted by all kinds of Statisticians is the Sufficiency Principle. It states that if there is a parameter of interest, θ, and data X is collected then if $T(X)$ is a sufficient statistic for X in respect to θ and two points x_1 and x_2 are such that $T(x_1) = T(x_2)$, all inference made on θ observing any one of these points should be the same. It follows from this principle that it is equivalent for the Statistician to observe the full data or just to observe the sufficient statistic.

However, when one is faced with the decision of choosing an experiment to be performed, he usually is only presented with the marginal distribution of each of these experiments. Since their joint distribution is unknown, it is not possible to use the sufficiency principle alone in order the determine which experiments are equivalent. Thus we believe that in order to provide the logical framework for comparison of experiments some extra principle must be added.

As a first step we consider that it should be possible to compare two experiments only knowing their marginal distributions. This happens because, usually, when we choose one of them, it is as if the other had not even existed. But once we consider that it is possible to compare two experiments only knowing their marginal distribution it also seems reasonable that, for any coupling existent between them, the comparison should be the same.

In a more narrow sense, we consider the Coupling Invariance Principle. It states that if there exists a coupling of the marginal distributions of two experiments such that both are sufficient one for the other, then both experiments are equally suitable for the inference on the interest parameter, no matter the way in which they are actually coupled.

On the other hand, a known way of determining whether two experiments bring the

CP1073, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering—28th International Workshop*
edited by M. de Souza Lauretto, C. A. de Bragança Pereira, and J. M. Stern

same information about θ is the Blackwell Equivalence [2] relation. It states that if X_1 and X_2 are random functions on θ, the parameter of interest, then they are equivalent in regard to θ if there is a randomization of X_1 with the same distribution as X_2 and a randomization of X_2 with the same distribution as X_1.

The objective of this article is to explore the consequences of the belief in both the Sufficiency and the Coupling Invariance Principle, specially in regard to the Blackwell Equivalence relation. To do so, in Section 3 we present some basic results of Markov Chains which will be necessary in the following. Next, in Section 4 we prove that if two experiments are Blackwell Equivalent then there exists a very strong coupling between them. Finally, in Section 5 we discuss the relationship between the Sufficiency Principle, the Coupling Invariance Principle, Blackwell's Equivalence relation and the Likelihood Principle.

Regarding this paper, all the sample spaces considered will be finite and endowed with the discrete σ-algebra. This way, we will consider that these spaces have a complete order, whenever this is necessary. All the results could be expanded in a straightforward way to enumerable spaces but this has not been done in matter to achieve simplicity. We believe that it might be possible to further generalize the results for some non-enumerable spaces but then the Markov Chains framework would have to be substituted for a more general kind of ergodic theorem.

2. EXAMPLES

In order to contextualize the arguments given in the Introduction, in this section we provide a few simple examples.

Example 1: Let us consider two random variables X and Y which, conditionally to the parameter of interest θ, have the same distribution. Firstly, it is trivial that X is Blackwell Equivalent to Y. Nevertheless, it is impossible to compare X and Y by the usual sufficiency concept since only their marginal distribution is given.

In this example it is easy to solve the question if there is some joint distribution for X and Y such that, in the usual sense, one is sufficient for the other. For example, if $X = Y$ almost surely, this condition is satisfied. On the other hand, not for all joint distributions are they reciprocally sufficient. For example, if X and Y were conditionally independent given θ, then they would not be reciprocally sufficient.

Example 2: Another simple example compares the usual definition of a sufficient statistic with that of Blackwell Sufficiency. Let X be a random vector, θ the parameter of interest and $T(X)$ a sufficient statistic of X.

Firstly, $T(X)$ is Blackwell Equivalent to X. The sufficiency of X is trivial and to prove that of $T(X)$ one must only remember that if x and y are points such that $T(x) = T(y)$ then $P(X = x|\theta)/P(X = y|\theta)$ is constant on θ and represents the odds of x against y.

In addition, $T(X)$ and X exist on the same measure space. They yield points with proportional likelihood functions with probability one. This is a specific case of the result which will be presented in section 4.

Example 3: A more interesting example can be found in [1]. The distributions of the considered random variables (X and Y) are Bernoulli with parameters respectively θ and $k\theta$. In the cited article it is shown that if $k \leq 1$ then X is Blackwell sufficient for Y. An easy way to prove such a statement is to imagine a random variable Z with Bernoulli distribution of parameter k and then show that the random variable ZX has the same law as Y.

Using the results in this paper it can be proved that Y is not Blackwell Sufficient for X in a straightforward way. On one hand, points 0 and 1 yield likelihood functions which are not proportional (on the variable X). On the other hand, ZX is a randomization of X such that $P(ZX = 0|X = x) \neq 0$ for any value of x. From the main result in this paper it will then be possible to conclude that Y is not Blackwell Sufficient for X.

In [1] it is shown that since X is Blackwell Sufficient for Y then sampling without replacement is Blackwell Sufficient for sampling with replacement. From this example it follows that sampling with replacement is not Blackwell Sufficient for sampling without replacement.

3. MARKOV CHAINS

In order to conclude the main demonstration of this article it will be necessary to make use of some results related to the theory of Markov Chains. In this section we briefly present all that will be needed in the following.

LEMMA 1: Let there be a Markov Chain (x_n) on a finite space χ with a transition matrix A. If A is irreducible then it has an invariant measure and it is unique.

Proof follows directly from the Perron-Frobenius Theorem [4] for non negative matrices.

LEMMA 2: Let there be a Markov Chain on a finite space χ with a transition matrix A and no transient states. Let A have irreducible components $C(1)$, ..., $C(n)$. Then there exists an unique set of probability functions $p_1(i)$, $i \in \{1,\ldots,|C(1)|\}$, $\ldots$, $p_n(i)$, $i \in \{1,\ldots,|C(n)|\}$, such that all invariant measures (μ) of A can be written as the following:

If $c_{k,i}$ is the element of number i of $C(k)$ then $\mu(c_{k,i}) = p_k(i).q_k$ and q is a probability function in $\{1,\ldots,n\}$.

This result can be found in [3]. A way to interpret it is to consider the sub-matrix A_k associated to $C(k)$. Since this matrix is irreducible, by Lemma 1 it only has one invariant measure which is p_k. Now suppose that at the initial position (x_0) of the Chain each component $C(k)$ has probability q_k of being chosen. As n goes to infinity the law of x_n converges to the one provided by Lemma 2.

4. MAIN RESULT

In this section we show that if two experiments are Blackwell Equivalent then there exists a very strong coupling between them. The existence of this coupling will make possible to discuss the relationship between Blackwell Equivalence, Sufficiency Principle and Coupling Invariance Principle in the next section.

THEOREM 1: Let X_1 and X_2 be two experiments with probability distributions respectively $f_1(.|\theta)$, with domain in χ_1, and $f_2(.|\theta)$, with domain in χ_2, then X_1 and X_2 are Blackwell Equivalent if and only if it is possible to couple both experiments in such a way that $\forall\ x_1 \in \chi_1$, $x_2 \in \chi_2$, if $P(X_1 = x_1, X_2 = x_2|\theta) > 0$ for some θ, then $f_1(x_1|.) \propto f_2(x_2|.)$.

Proof:
($\Leftarrow$) First, it is easy to prove that a sufficient statistic is always Blackwell Equivalent to the the data. This happens because the odds between points with proportional likelihoods are θ-free. Next, in the coupling provided, X_1 and X_2 are both Sufficient Statistics for (X_1, X_2). Thus it follows that X_1 and X_2 are Blackwell Equivalent.

($\Rightarrow$) Since X_1 is Blackwell Sufficient for X_2 then there exists a θ-free transition matrix, P, such that:

$$Pf_1(.|\theta) = f_2(.|\theta), \forall\theta \in \Theta$$

On the other hand, X_2 is also Blackwell Sufficient for X_1 and, similarly, there is a θ-free transition matrix, Q, which satisfies:

$$Qf_2(.|\theta) = f_1(.|\theta), \forall\theta \in \Theta$$

By using both these relations we know that there exist two other θ-free transition matrices $A = QP$ and $B = PQ$, such that:

$$Af_1(.|\theta) = f_1(.|\theta), \forall\theta \in \Theta$$

$$Bf_2(.|\theta) = f_2(.|\theta), \forall\theta \in \Theta$$

We will adopt as a strategy to continue the demonstration to imagine that A is a transition matrix for a Markov Chain on χ_1. This way, we have that $f_1(.|\theta)$ is an invariant measure for A, $\forall\theta \in \Theta$.

Firstly, without loss of generality, we may assume that A has no transient states. If there were such a state, for all θ, it would have null measure. On one side, since the theorem only imposes restrictions to states which have positive measure for some θ, the demonstration can be easily adapted as to satisfy the existence of transient states. On the other side, if an experiment has an outcome which never happens, the sample space could be rewritten so to disregard it.

Now, by using Lemma 2, we know that if $C(1),\ldots,C(n)$ are irreducible components of A and c(k,i) is the element of number i of $C(k)$ then $f_1(c(k,i)|\theta) = q_{k,\theta}.p_k(i)$. Consequently $f_1(c(k,i)|\theta) = f_1(c(k,j)|\theta).(p_k(i)/p_k(j))$. Finally, we have that if two states are in the same irreducible component then their likelihood functions are proportional. The same proof applies to matrix B.

We now consider the minimal sufficient statistic for X_1, X_1', which assumes states in χ_1', a set of subsets of χ_1 which elements have proportional likelihood, and also the minimal sufficient statistic X_2, X_2', in χ_2', defined in the same way. We can imagine χ_1' and χ_2' as partitions given by the equivalence relation "proportional likelihoods".

Since X_1' is Blackwell Equivalent to X_1, X_1 is Blackwell Equivalent to X_2 and X_2 is Blackwell Equivalent to X_2', then X_1' is Blackwell Equivalent to X_2'. Thus if X_1' has its distribution given by f_1' and X_2' has its distribution given by f_2' then there exist transition matrices P' and Q' such that $P'f_1' = f_2'$ and $Q'f_2' = f_1'$.

We define that the element of number i in χ_1' is connected to the one of number j in χ_2' if $p'(i,j) > 0$. In a similar manner we define that the element of number j in χ_2' is connected to the one of number i in χ_1' if $q'(j,i) > 0$.

Firstly we note that every state in χ_1' is connected to at least one state in χ_2' and every state in χ_2' is connected to at least one state in χ_1' since P' and Q' are transition matrices.

After that, it is easy to see that if one state x_1' in χ_1' is connected to x_2' in χ_2' then x_2' only is connected to x_1'. This happens because if x_2' were connected to another state in χ_1' then there would be an irreducible component in $Q'P'$ which contained states with likelihood functions which were not proportional. Similarly, if a state x_2' in χ_2' is connected to a state x_1' in χ_1' then x_1' is only connected to x_2'.

From the above it is proven that each state x_1' in χ_1' is connected only to one state x_2' in χ_2' and x_2' is only connected to x_1'. This implies that the likelihood of x_1' is proportional to that of x_2'.

Finally, let X_1 be a random function with probability distribution given by f_1 and let U be a random variable uniform in $[0,1]$. Let us consider the minimal sufficient statistic of X_1, X_1'. It has been shown that if we construct X_2', the minimal sufficient statistic of X_2, as the randomization $F(X_1',U)$ induced by P' then if $P(X_1' = x_1', X_2' = x_2'|\theta) > 0$, then $f_1(x_1'|\theta) \propto f_2(x_2'|\theta)$, $\forall\, x_1' \in \chi_1', x_2' \in \chi_2'$. Constructing X_2 by the natural randomization on X_2' the theorem is proven.

5. CONCLUSIONS

From the previous section it is now possible to conclude that if two experiments X_1 and X_2 are Blackwell Equivalent and a person believes in the Sufficiency Principle and in the Coupling Invariance Principle then he must believe both experiments are equally suitable.

Because of the $\Rightarrow$ passage in Theorem 1 it is possible to couple X_1 and X_2 such that both are sufficient statistics for (X_1,X_2). This way, if the experiments were actually coupled in this manner, by the sufficiency principle any one of them would lead to the same inference on θ. In addition, by the Coupling Invariance Principle, since there exists a coupling in which X_1 and X_2 are equally suitable for the inference on θ then this result is general. This way, the Sufficiency and Coupling Invariance Principles induce Blackwell Equivalence.

On the other hand, both principles follow when one believes in Blackwell's Equivalence relation. Since it has been proven that any sufficient statistic is Blackwell Equivalent to the whole data, the Sufficiency Principle is a consequence of Blackwell's Equivalence. The Coupling Invariance Principle also follows directly from Blackwell's Equivalence by the usage of $\Leftarrow$ passage in Theorem 1.

This way, the Coupling Invariance Principle and the Sufficiency Principle are equivalent to Blackwell's Equivalence relation.

Further on, since in the $\Rightarrow$ passage of Theorem 1 the demonstration stated a general transition matrix P' it is possible to conclude that, $\forall\theta \in \Theta$, if two experiments are Blackwell Equivalent then each normalized likelihood function (the data that will wield it) has equal probability of being observed. This way, if one believes that the second condition is enough for two experiments to be considered equivalent (which seems to be slightly related to the Likelihood Principle), then he also must believe in the Blackwell Equivalence Relation.

Finally, the demonstration of the passage $\Rightarrow$ of Theorem 1 also shows an interesting fact. Whenever randomization on some experiment is such that there exists mixture of datum with likelihood functions which are not proportional, then the resulting distribution can't be Blackwell Equivalent to the former. This might indicate that, in these situations, information about the parameter of interest is being lost.

ACKNOWLEDGMENTS

We are grateful to Antonio Galves, Estéfano Alves de Souza, Fernando V. Bonassi, Luis G. Esteves, Julio M. Stern, Paulo C. Marques and Sergio Wechsler for the insightful discussions and suggestions. The authors of this paper have benefited from the support of CNPq and FAPESP.

REFERENCES

1. Basu, D. and Pereira, C.A. de B. (1990) Blackwell Sufficiency and Bernoulli Experiments *Brazilian Journal of Probability and Statistics Experiments*, **4**, pp 137-145.
2. Blackwell, D. (1969) Comparison of experiments, *Proceedings of the 2nd Berkeley Symposium*, pp 93-101.
3. Ferrari, P. A. and Galves, A. (2000) *Coupling and regeneration for stochastic processes*. Sociedad Venezolana de Matematicas, Caracas, 2000, pp 13-31.
4. Kemeny, J. and Snell, L. (1960) *Finite Markov Chains*. Van Nostrand Reinhold, New York.
5. Birnbaum, A. (1962) On the Foundations of Statistical Inference, *Journal of the American Statistical Association*, **57**, 269Ű326.