A Simple Proof of Blackwell's "Comparison of Experiments" Theorem*

JACQUES CRÈMER

Department of Economics, University of Pennsylvania,
Philadelphia, Pennsylvania 19104

Received January 27, 1981; revised May 29, 1981

This note presents a simple proof of Blackwell's well-known theorem on the equivalence of two orderings of experiments, one based on the statistical concept of sufficiency, and the other on the economic concept of the value to the decision makers. Journal of Economic Literature, Classification Number: 026, 211.

I. INTRODUCTION

In [1], Blackwell established the equivalence of two quasi-orderings on the set of observations—one based on the statistical concept of sufficiency and another based on the economic criterion of the value to a decision maker. Most of the existing proofs of this theorem are long and involved (see [2] for an exposition). In this note, I present a much shorter proof. It is presented in a discrete framework, but the basic ideas should carry to more general frameworks.

II. BLACKWELL'S THEOREM

We consider a decision maker whose utility function $u(a, \theta_t)$ is defined over $A \times \theta$, where $A$ is the set of possible actions, and $\theta$ the set of possible states of nature $(\theta_1, \theta_2, ..., \theta_n)$. $\theta$ will remain fixed throughout, and we call $t$ the data $(u, A)$ which describe the decision maker. Given a vector

* This paper is a byproduct of research conducted jointly with A. Zaman. Its production was supported in part by NSF Grant SOC 78–06157. A referee provided interesting comments.

1 Ponssard [3] also presents a nice proof of Blackwell's theorem, with somewhat different techniques.
\( p = (p(1), \ldots, p(l), \ldots, p(n)) \) \((\sum_i p(i) = 1, p(i) \geq 0 \text{ for all } i)\) of probabilities of the states of nature, the decision maker solves problem \( \mathcal{P}(p; t) \):

\[
\max_{a \in A} \sum_i p(i) u(a, \theta_i).
\]

Let \( a(p; t) \) be a solution of this problem. \( V(p; t) \), the value of problem \( \mathcal{P}(p; t) \), is defined by:

\[
V(p; t) = \sum_i p(i) u(a(p; t), \theta_i).
\]

If \( \mathcal{P}(p; t) \) has a solution for every \( p \), we say that \( t = (u, A) \) is admissible.

Let us now reinterpret \( p \) as a (subjective) prior probability distribution, the same throughout this paper, and, without loss of generality, assume that \( p(i) > 0 \) for all \( i \in I = \{1, \ldots, n\} \).

Before acting, the decision-maker can observe one of two random variables \( x^1 \) and \( x^2 \). For \( s = 1, 2 \), \( x^s \) can take values in the set \( K^s = \{1, 2, \ldots, k^s, \ldots, m^s\} \). When \( \theta_i \) is the state of nature, the conditional probability distribution of \( x^s \) is \( \pi^s(\cdot \mid i) = (\pi^s(1 \mid i), \ldots, \pi^s(k^s \mid i), \ldots, \pi^s(m^s \mid i)) \). Without loss of generality, for all \( k^s \in K^s \) there exists an \( i \in I \) such that \( \pi^s(k^s \mid i) > 0 \). As a consequence \( \pi^s(k^s) = \sum_{i \in I} \pi^s(k^s \mid i) p(i) \) is also strictly positive.

When the decision maker observes \( k^s \), he recomputes his subjective probability distribution over \( \theta \). The new distribution is described by the \( n \)-vector \( p^s(\cdot \mid k^s) \) whose \( i \)-th component is \( p^s(i \mid k^s) = \pi^s(k^s \mid i) \cdot p(i)/\pi^s(k^s) \). Then, the decision maker takes action \( a^s(p^s(\cdot \mid k^s); t) \); his expected utility, conditional on his observation, is \( V(p^s(\cdot \mid k^s); t) \). The expected value of the sequence “observe \( x^s \), choose an action” is \( S(\bar{x}^s; t) = \sum_{k^s \in K^s} \pi^s(k^s) \).

The following quasi-ordering on observations (i.e., random variables) seems natural for an economist:

**Definition 1.** \( \bar{x}^1 \) is more informative than \( \bar{x}^2 \) (\( \bar{x}^1 \succ \bar{x}^2 \)) if \( S(\bar{x}^1; t) \geq S(\bar{x}^2; t) \) for all admissible \( t \).

From a probabilistic point of view, on the other hand, it seems natural to say that \( \bar{x}^1 \) is more informative than \( \bar{x}^2 \) if \( \bar{x}^2 \) is constructed by “adding noise” to \( \bar{x}^1 \). Formally:

\[ ^2 \text{Actually, we could replace max by sup in the definition of problem } \mathcal{P}(p; t) \text{ and only assume that such a sup exist.} \]
DEFINITION 2. \( x' \) is sufficient for \( x^2 \) (\( x' \preceq x^2 \)) if there exist \( m^2 \times m^1 \) positive real numbers \( b_{k^2 k^1} \) \((k^2 \in \{1, \ldots, m^2\}, k^1 \in \{1, \ldots, m^1\})\) such that:

\[
\sum_{k^2} b_{k^2 k^1} = 1 \quad \text{for all } k^1,
\]

\[
\pi^2(k^2 \mid i) = \sum_{k^1} b_{k^2 k^1} \pi^1(k^1 \mid i).
\]

We can now prove Blackwell's theorem [1]:

**THEOREM.** \( x' \preceq x^2 \) if and only if \( x' \Rightarrow x^2 \).

III. PROOF OF THE THEOREM

(a) *Preliminary Computations*

Let \( B \) be the set of families of \( k^2 \times k^1 \) real positive numbers \( \{b_{k^2 k^1}\} \) such that \( \sum_{k^2} b_{k^2 k^1} = 1 \) for all \( k^1 \in K^1 \). Let \( \pi^2(\cdot \mid \cdot) \) be the \( m^2 \times n \) vector whose \((k^2 + m^1(i-1))\)th component is \( \pi^2(k^2 \mid i) \), and let \( D \) be the set of vectors \( d \in \mathbb{R}^{m^2 \times n} \) such that \( \sum_{k^1} b_{k^2 k^1} \pi^1(k^1 \mid i) = d_{k^2 i} \) for some family \( \{b_{k^2 k^1}\} \) in \( B \) (\( d_{k^2 i} \) is the \((k^2 + (m^2 - 1)i)\)th component of \( d \)).

S1 to S5 below are equivalent:

S1: \( x' \Rightarrow x^2 \).

S2: \( \pi^2(\cdot \mid \cdot) \in D \).

S3: For all \( q \in \mathbb{R}^{m^2 \times n} \), there exists \( \{b_{k^2 k^1}\} \in B \) such that:

\[
\sum_{k^2, i} q_{k^2 i} \pi^2(k^2 \mid i) \leq \sum_{k^2, i} q_{k^2 i} \sum_{k^1} b_{k^2 k^1} \pi^1(k^1 \mid i) = \sum_{k^1} \sum_{k^2} b_{k^2 k^1} \sum_{i} q_{k^2 i} \pi^1(k^1 \mid i).
\]

S4: For all \( q \in \mathbb{R}^{m^2 \times n} \):

\[
\sum_{k^2, i} q_{k^2 i} \pi^2(k^2 \mid i) \leq \sum_{k^1} \left[ \max_{k^2} \sum_{i} q_{k^2 i} \pi^1(k^1 \mid i) \right].
\]

To see that this definition does indeed translate the intuitive notion of "adding noise" first reinterpret \( b_{k^2 k^1} = \pi(k^2 \mid k^1) \) (this makes \( \pi(k^2 \mid k^1) \) independent of \( i \)). Then, \( x^2 \) is obtained by "scrambling" the information contained in \( x' \). The general definition says that \( x' \Rightarrow x^2 \) if \( x^2 \) has, conditional on any \( \theta \in \theta \), the same distribution than a random variable obtained by scrambling the information contained in \( x' \).
S5: For all \( \phi \in R^{m_2 \times n} \):

\[
\sum_{k_2} \pi^2(k_2) \sum_i \phi_{k_2 i} p^2(i | k_2) \leq \sum_{k_1} \pi^1(k_1) \max_{k_2} \left[ \sum_i \phi_{k_2 i} p^1(i | k_1) \right].
\]

S1 \( \Leftrightarrow \) S2 by definition of sufficiency.

S2 \( \Leftrightarrow \) S3 by the separating hyperplane theorem. (If S3 did not hold, \( \pi^2(\cdot | \cdot) \) could be separated from \( D \) by an hyperplane, which would contradict S2.)

S3 \( \Leftrightarrow \) S4 because \( \sum_{k_1} \max_{k_2} | \sum_i q_{k_2 i} \pi^1(k_1 | i) | \) is the maximum over \( B \) of \( \sum_{k_1} \sum_{k_2} b_{k_2 k_1} \sum_i q_{k_2 i} \pi^1(k_1 | i) \).

S4 \( \Leftrightarrow \) S5 by the change of variables \( q_{k_2 i} = p(i) \phi_{k_2 i} \).

We prove this implication by contradiction. Assume that \( \tilde{x}^1 \) is not sufficient for \( \bar{x}^* \), then S5 does not hold and there exists \( \phi \in R^{m_2 \times n} \) such that:

\[
\sum_{k_1} \pi^1(k_1) \max_{k_2} \left[ \sum_i \phi_{k_2 i} p^1(i | k_1) \right]
\leq \sum_{k_2} \pi^2(k_2) \sum_i \phi_{k_2 i} p^2(i | k_2)
\leq \sum_{k_2} \pi^2(k_2) \max_{\lambda^2 \in K^2} \left[ \sum_i \phi_{\lambda_2 i} p^2(i | k_2) \right].
\]

Let \( t = (u, A) \) with \( A = K^2 \) and \( u(k^2, \theta_i) = \phi_{k_2 i} \). Then:

\[ S(\tilde{x}^1; t) < S(\bar{x}^2; t), \]

which establishes the contradiction.

(c) \( \tilde{x}^1 \supseteq \bar{x}^2 \Rightarrow \tilde{x}^1 \supseteq \bar{x}^2 \)

For any \( t = (u, A) \), let \( t' = (u, A') \), where \( A' \) is the set of \( a \in A \) such that \( a = a(p^2(\cdot | k^2); t) \) for some \( k^2 \in K^2 \). Then \( S(\bar{x}^2; t') = S(\bar{x}^2; t) \) and \( S(\tilde{x}^1; t') \leq S(\tilde{x}^1; t) \).

Let \( \phi_{k_2 i} = u(a(p^2(\cdot | k^2); t), \theta_i) = u(a(p^2(\cdot | k^2); t'), \theta_i) \), then S5 implies:

\[ S(\tilde{x}^2; t') \leq S(\tilde{x}^1; t') \]

and we have:

\[ S(\tilde{x}^2; t) = S(\tilde{x}^2; t') \leq S(\tilde{x}^1; t') \leq S(\tilde{x}^1; t), \]

which establishes the result.
REFERENCES


2. R. E. Kihlstrom, A "Bayesian" exposition of Blackwell's Theorem on the comparison of experiments, mimeo.