A simple Bayes factor for testing measurement-theoretic axioms

Thierry Marchant*

Abstract

A lot of models used to represent preferences or other relations have been characterized in the framework of measurement theory. To assess the empirical validity of one of these models, we can either test the model itself or test the axioms characterizing it. When a model is rejected, the latter approach has the advantage of indicating why it is rejected. The available statistical techniques for testing measurement-theoretic axioms are complex and not always satisficing. In this paper, we present a new statistical technique based on a Bayes factor, very simple and flexible.

Key words : Measurement theory, empirical test

1 Introduction

Measurement is a fundamental and ubiquitous operation in most modern sciences and has usually been considered as unproblematic until the end of the nineteenth century. Hölder [1901] was one of the first scientists to question the very nature of measurement and made the first steps towards the development of a theory of measurement. In the second half of the twentieth century, measurement-theory made enormous progress [Krantz et al., 1971, Suppes et al., 1989, Luce et al., 1990] and helped us to understand what measurement is and which hypotheses (called axioms) underly the various measurement techniques. Simultaneously with these theoretical developments, some researchers started to cast doubts on the empirical validity of some axioms [e.g. Allais, 1953, Tversky, 1969]. For instance, Tversky [1969] used an experiment to produce data in contradiction with transitivity. Unfortunately, these data were not easy to use in a statistical analysis and the first sound

*Ghent University, Dunantlaan 1, 9000 Ghent, Belgium. Thierry.Marchant@UGent.be

analysis of Tversky's data appeared in Iverson and Falmagne [1985]. The approach followed by Iverson and Falmagne was based on a likelihood ratio and was very complex, making its use with large data sets or other axioms than transitivity very difficult or even impossible. Actually, since 1985, this technique was almost never used. Since then, a lot of empirical research took place, usually producing difficult to analyze data, but no significant advance occured in the statistical analysis of experimental data, until very recently, with different papers by Iverson, Karabatsos and Myung [e.g. Karabatsos, 2005, Myung et al., 2005].

In Karabatsos [2005], it is assumed that, at the time of the experiment, the subject chooses at random one binary relation according to some probability distribution. Based on the data, the probability of each possible binary relation is estimated and the hypothesis that is actually tested is whether the total probability, of all relations satisfying the axiom under scrutiny, is larger than $1/2$. A drawback of this method is that it can lead to the acceptance of an axiom with strange probability distributions. Suppose for instance that we test transitivity, on four elements a, b, c, d, and that the probability distribution governing the subject's choices is $P[\text{order } a \succ b \succ c \succ d] = .26$, $P[\text{order } d \succ c \succ b \succ a] = .26$, $P[\text{hamiltonian cycle } c \succ b \succ a \succ d \succ c] = .48$ and $P[\text{any other relation}] = 0$. The total probability of the relations satisfying transitivity is $.52 > 1/2$. So, according to Karabatsos [2005], transitivity holds. Yet, the relation with the largest (by far) probability is not at all transitive. Furthermore, the only two transitive relations with a positive probability are contradictory. This does not speak in favor of transitivity.

Myung et al. [2005] assume that the subject has in mind a probability p_{ab} for each pair (a, b) and, when presented with the pair (a, b), selects a with probability p_{ab} or b with probability $1 - p_{ab}$. Using a very sophisticated Bayesian technique, based on posterior parameter samples, they then test a stochastic recasting of the axiom under scrutiny. For example, instead of transitivity, they test weak stochastic probability, i.e. $p_{ab} \geq .5$ and $p_{bc} \geq .5$ implies $p_{ac} \geq .5$. This approach has two weaknesses: (1) for each parameter p_{ab}, they assume a prior Beta distribution on $[0, 1]$ which is very convenient for the computations but difficult to motivate and (2) as they say in the footnote on p. 207, "The approach based on posterior model probabilities may be theoretically more appealing than that based on posterior parameter samples ... "

In this paper, we propose an approach based on a Bayes factor, thus based on posterior model probabilities. It is a generalization of the approach proposed by Desimpelaere and Marchant [2006]. In Section 2, we present the probabilistic model, supposed to generate the data, that we will test, and, in Section 3, we show how to test the model, using a Bayes factor.

2 Probabilistic model

2.1 Notation and model

Let $X = \{a, b, c, \ldots\}$ be the set of objects or stimuli that we will use in a forced choice experiment ($\#X = m$). We suppose that the subject has a complete binary relation on X in mind. This relation, called true relation and denoted by $\succsim_T$, is of course unknown. The symmetric (respectively asymmetric) part of this relation is denoted by $\sim_T$ (resp. $\succ_T$). When we ask the subject to choose between two objects a and b from X, we assume that he chooses according to $\succsim_T$. But, since humans are not always consistent, the subject makes some errors. So, if $a \succ_T b$, the subject chooses a with probability p_{ab} ($\geq 1/2$ and unknown) or b with probability $1 - p_{ab}$ (we assume $p_{ab} = p_{ba}$ but this can be relaxed). If $a \sim_T b$, the subject chooses a or b with equal probability ($p_{ab} = 1/2$).

Let C be the set of all complete binary relations on X and A be the set of all relations in C that satisfy a given axiom (also called A). The hypothesis that we will test is that $\succsim_T$ belongs to A. Let D be the set of all unordered pairs of distinct objects in X. Since we will follow a Bayesian approach, we consider that $\succsim_T$ is a random variable taking its values in C and, attached to each relation R in C, there is a vector of parameters $\mathbf{p}(R) = (p_{ab}(R))_{(a,b)\in D}$. This is in contrast with Myung et al. [2005]: they consider that the parameters p_{ab} are continuous random variables with values in $[0, 1]$. Nevertheless, in our model, once the value of $\succsim_T$ is fixed, the parameters $\mathbf{p}$ are fixed and satisfy weak stochastic transitivity, just like in Myung et al. [2005].

2.2 Likelihood

Suppose each of the $m(m-1)/2$ pairs in D is presented N_{ab} times ($N_{ab} \geq 0$) to a subject and he chooses n_{ab} times a (this implies that he chooses $N_{ab} - n_{ab}$ times b). Our data consist then of the vector $\mathbf{n} = (n_{ab})_{(a,b)\in D}$. Suppose we know $\succsim T$ and $\mathbf{p}(\succsim_T)$. The likelihood is then

$$P[\mathbf{n}|\succsim_T, \mathbf{p}(\succsim_T), \mathbf{N}] = \prod_{(a,b): a\succ_T b} p_{ab}^{n_{ab}} (1 - p_{ab})^{N_{ab}-n_{ab}} \prod_{(a,b): a\sim_T b} (1/2)^{N_{ab}}.$$

If we know $\succsim_T$ but do not know $\mathbf{p}(\succsim_T)$, we can estimate it. We have $\hat{p}_{ab}(\succsim_T) = \max(n_{ab}/N_{ab}, 1/2)$ if $a \succ_T b$ and $\hat{p}_{ab}(\succsim_T) = p_{ab}(\succsim_T) = 1/2$ if $a \sim_T b$. It is then possible to estimate the likelihood $P[\mathbf{n}|\succsim_T, \mathbf{N}]$ by $P[\mathbf{n}|\succsim_T, \hat{\mathbf{p}}(\succsim_T), \mathbf{N}]$.

3 Bayes factor

In this section, we will use the Bayes factor

$$\frac{P[\succsim_T \in A|\mathbf{n},\mathbf{N}]}{P[\succsim_T \in A^*|\mathbf{n},\mathbf{N}]},$$

where $A^* = C \setminus A$ in order to weigh the evidence in favour of A against the evidence in favour of A^*. The Bayes factor can be estimated by

$$\frac{P[\mathbf{n}|\succsim_T \in A,\mathbf{N}]}{P[\mathbf{n}|\succsim_T \in A^*,\mathbf{N}]}\,\frac{P[\succsim_T \in A]}{P[\succsim_T \in A^*]}, \tag{1}$$

where $P[\succsim_T \in A]$ is the prior probability of A and $P[\succsim_T \in A^*] = 1 - P[\succsim_T \in A]$ is the prior probability of A^*. Using standard probability rules, we have

$$P[\mathbf{n}|\succsim_T \in A,\mathbf{N}] = \frac{\sum_{R\in A}\left(P[\mathbf{n}|\succsim_T = R,\mathbf{N}]\,P[\succsim_T = R]\right)}{P[\succsim_T \in A]} \tag{2}$$

and a similar expression for $P[\mathbf{n}|\succsim_T \in A^*,\mathbf{N}]$. Using these, we can rewrite the Bayes factor (1) as

$$\frac{\sum_{R\in A}\left(P[\mathbf{n}|\succsim_T = R,\mathbf{N}]\,P[\succsim_T = R]\right)}{\sum_{R\in A^*}\left(P[\mathbf{n}|\succsim_T = R,\mathbf{N}]\,P[\succsim_T = R]\right)}. \tag{3}$$

If we know the priors $P[\succsim_T = R]$, for all $R \in A$ and all $R \in A^*$, it is possible to compute (3). If we have no prior information, we can use Laplace's principle and assume that all relations in A have the same probability (i.e. $P[\succsim_T = R] = P[\succsim_T \in A]/\#A$, for $R \in A$) and that all relations in A^* have the same probability (i.e. $P[\succsim_T = R] = P[\succsim_T \in A^*]/\#A^*$, for $R \in A^*$). The Bayes factor then becomes

$$\frac{P[\succsim_T \in A]}{P[\succsim_T \in A^*]}\,\frac{\#A^*}{\#A}\,\frac{\sum_{R\in A} P[\mathbf{n}|\succsim_T = R,\mathbf{N}]}{\sum_{R\in A^*} P[\mathbf{n}|\succsim_T = R,\mathbf{N}]}. \tag{4}$$

Here, again, in absence of prior information, we can assume that $P[\succsim_T \in A] = P[\succsim_T \in A^*] = 1/2$ and the Bayes factors simplifies to

$$\frac{\#A^*}{\#A}\,\frac{\sum_{R\in A} P[\mathbf{n}|\succsim_T = R,\mathbf{N}]}{\sum_{R\in A^*} P[\mathbf{n}|\succsim_T = R,\mathbf{N}]}. \tag{5}$$

Exact values of expressions (3–5) can easily be computed for $m \leq 5$ but are intractable for large m. Desimpelaere and Marchant [2006] showed that a Monte Carlo approximation converges in a reasonable time for $m \leq 10$. Using fast computers, optimized code and good approximation algorithms, it is probably possible to go until $m = 11$ or 12 but not much further. This is a weakness of the approach we present but, since almost all empirical studies of measurement-theoretic axioms involve less than 10 objects, this weakness does not seem very important.

When we have prior information, it is sometimes possible to speed up the computations. Suppose for example that the objects in X are binary lotteries of the form $(x, q; y)$, where x obtains with probability q and y with probability $1 - q$, with x and y real numbers. We may assume (or we may have previously tested) that preferences over loteries are monotone in the following sense: $x > x'$ and $y > y'$ implies $(x, q; y) \succ_T (x', q; y')$. We then have $P[\succsim_T = R] = 0$ for any relation R violating monotonicity (denoted by M). Expression (3) can then be rewritten as

$$\frac{\sum_{R \in A \cap M} \left(P[\mathbf{n} | \succsim_T = R, \mathbf{N}]\ P[\succsim_T = R]\right)}{\sum_{R \in A^* \cap M} \left(P[\mathbf{n} | \succsim_T = R, \mathbf{N}]\ P[\succsim_T = R]\right)}. \tag{6}$$

The sums being now over $A \cap M$ (or $A^* \cap M$), they can be computed faster.

4 Conclusion

We presented a simple technique to test any deterministic measurement-theoretic axiom involving a binary relation. It is based on a Bayes factor and can handle prior information that we have about the axiom under scrutiny or about other characteristics of the relation. The computation of the Bayes factor requires a lot of time, thereby limiting the size of the problems we can analyze. Our approach can easily be adapted to measurement techniques involving a k-ary relation instead of a binary one.

References

M. Allais. Le comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'école américaine. *Econometrica*, 21:503–46, 1953.

C. Desimpelaere and T. Marchant. An empirical test of some measurement-theoretic axioms for fuzzy sets. 2006.

O. Hölder. Die Axiome der Quantität und die Lehre vom Mass. *Ber. Verh. Sächs. Ges. Wiss. Leipzig, Math. Phys. Cl.*, 53:1–64, 1901.

G. Iverson and J.-C. Falmagne. Statistical issues in measurement. *Mathematical Social Sciences*, 10:131–153, 1985.

G. Karabatsos. The exchangeable multinomial model as an approach to testing deterministic axioms of choice and measurement. *Journal of Mathematical Psychology*, 49: 51–69, 2005.

D. H. Krantz, R. D. Luce, P. Suppes, and A. Tversky. *Foundations of measurement: Additive and polynomial representations*. Academic Press, London, 1971.

R. D. Luce, D. H. Krantz, P. Suppes, and A. Tversky. *Foundations of measurement*, vol. 3: *Representation, axiomatisation and invariance*. Academic Press, New York, 1990.

J. I. Myung, G. Karabatsos, and G. Iverson. A Bayesian approach to testing desicion making axioms. *Journal of Mathematical Psychology*, 49:205–225, 2005.

P. Suppes, D. H. Krantz, R. D. Luce, and A. Tversky. *Foundations of Measurement*, volume 2: Geometrical, threshold and probabilistic representations. Academic Press, New York, 1989.

A. Tversky. Intransitivity of preferences. *Psychological review*, 76:31–48, 1969.