A Statistical Test in Choice Data of the Assumption that

Repeated Choices are Independently and Identically Distributed

Michael H. Birnbaum

California State University, Fullerton

Running head: Testing iid assumptions in choice

*Contact Information:
Prof. Michael Birnbaum
Dept. of Psychology, CSUF H-830M
Box 6846
Fullerton, CA 92834-6846
USA

Phone: (657) 278-2102

Email: mbirnbaum@fullerton.edu

I thank William Batchelder for suggesting the permutation method and helpful discussions; Michel Regenwetter kindly provided data for reanalysis and useful discussions of these issues. Thanks are also due Kathleen Preston for helpful suggestions. This work was supported in part by a grant from the National Science Foundation, SES DRMS-0721126.

Date: Aug 1, 2011

Abstract

This paper develops a test of independence in choice data collected with small samples. The method builds on the approach of Smith and Batchelder (2008). The technique is intended to distinguish cases where a person is systematically changing “true” preferences (from one group of trials to another) from cases in which a person is following a random preference mixture model with independently and identically distributed sampling in each trial. Preference reversals are counted between all pairs of repetitions. The variance of these preference reversals between all pairs of repetitions is then calculated. The distribution of this statistic is simulated by a Monte Carlo procedure in which the data are randomly permuted and the statistic is calculated in each simulated sample. A second test computes the correlation between the mean number of preference reversals and the difference between replicates, which is also simulated by Monte Carlo. Data of Regenwetter, Dana, and Davis-Stober (2011) are reanalyzed by this method. Eight of the 18 participants showed significant deviations from the independence assumptions by one or both of these tests, which is significantly more than expected by chance.

Regenwetter, Dana, and Davis-Stober (2011) proposed a solution to the problem of testing whether noisy choice data satisfy or violate structural properties of decision making models such as transitivity of preference. Their stochastic choice model assumes that on a given trial, a person behaves as if she or he samples randomly from a mixture of transitive preferences. The model was used to analyze a replication of Tversky's (1969) study that had reported systematic violations of transitivity of preference (Regenwetter, et al., 2010, 2011). Reanalysis via this stochastic mixture model of the Tversky data and new data concluded that transitivity can be retained.

Birnbaum (2011) agreed with much of their paper, including their conclusions regarding transitivity, but criticized the method in part because it assumes that responses by the same person to repeated choices are independent and identically distributed (iid). If this assumption is violated, the method of Regenwetter et al. (2011) might lead to wrong conclusions regarding the tests of structural properties.

In the true and error model, a rival stochastic representation that can also be used to test structural properties such as transitivity in mixture models, independence can be violated when a person has a mixture of true preferences and changes true preferences systematically during the course of the study. Birnbaum (2011) showed how the Regenwetter et al. method might lead to wrong conclusions when iid is violated, and described methods for testing between these two rival stochastic models of choice. These methods allow tests of assumptions of the Regenwetter, et al. (2011) approach against violations that might be indicative of the true and error model. He described conventional statistical tests that require conventional sized samples. Hypothetical examples illustrated cases in which the method of

Regenwetter, et al. (2011) might lead to the conclusion that transitivity was satisfied, even when a more detailed analysis showed that the data contained systematic violations.

The methods described by Birnbaum (2011) to test independence would require large numbers of trials, however, and might be difficult to implement. The experiment of Tversky (1969), which Regenwetter et al. (2011) replicated, does not have sufficient data to allow the full analyses proposed by Birnbaum (2011). Nevertheless, this note shows that by building on the approach of Smith and Batchelder (2008), it is possible to test the iid assumptions even in small studies such as that of Regenwetter, et al. (2011).

Testing IID Assumptions in Small Studies of Choice

Suppose a person is presented with *m* choices, and each choice is presented *n* times. For example, each of these *m* choices might be intermixed with filler choices and presented in a random sequence such that each choice from the experimental design is separated by several fillers. These choices might also be blocked such that all *m* choices are presented in each of *n* trial blocks, but blocking is not necessary to this test. Let $x(i, j)$ represent the response to choice j on the ith presentation of this same choice.

Define matrix $\boldsymbol{z}$ with entries as follows:

$$z(i,k) = \sum_{l=1}^{m} [x(i,l) - x(k,l)]^2, \qquad (1)$$

where $\boldsymbol{z}$ is an n by n matrix showing the distance between each pair of rows of the original data matrix. If responses are coded with successive integers, 0 and 1, for example, representing the choice of first or second stimulus, then $z(i, k)$ is a simply a count of the number of preference reversals between repetitions i and k, that is, between two rows of $\boldsymbol{x}$. In this case, the entries of $z(i, k)$ would have a minimum of 0, when a person made exactly the same responses on all m choices in two repetitions, and a maximum of m, when a person made exactly opposite choices on all m trials.

Smith and Batchelder (2008) show that random permutations of the original data matrix allow one to simulate the distribution of data that might have been observed under the null hypothesis. According to iid, it should not matter how the data of $\boldsymbol{x}$ are permuted within each column. That is, it should not matter if we switch two values in the same column of $\boldsymbol{x}$; they are two responses to the same choice on different repetitions by the same person. Therefore, it should not matter whether we assign one response to the first repetition block and the other to the last, or vice versa.

Assuming iid, the off-diagonal entries of matrix $\boldsymbol{z}$ should be homogeneous, apart from random variation. However, if a person has systematically changed preferences during the study, then there can be some entries of $\boldsymbol{z}$ that are small and others that are much larger. That is, there can be a relatively larger variance of the entries in $\boldsymbol{z}$ when iid is violated.

Therefore, one can compute the variance of the entries in $\boldsymbol{z}$ of the original data matrix, $\boldsymbol{x}$, and then compare this observed variance with the distribution of simulated variances generated from random permutations of the data matrix. If independence holds, then random permutations of the columns will lead to variances that are comparable to that of the original data, but if the data violate independence, then the original data might have a higher variance than that of most random permutations. The proportion of random permutations leading to a simulated variance that is greater than or equal to that observed in the original data, taken from a large number of Monte Carlo simulations, is the p_V value for this test of iid.

A second statistic that can be calculated from the matrix of z is to compute the correlation between the mean number of preference reversals and the absolute difference in replications. If a person changes gradually from one set of preferences to another, behavior will be more similar between replicates that are closer together in time than between those that are farther apart (Birnbaum, 2011). This statistic can also be simulated via Monte Carlo methods, and the proportion of cases for which the absolute value of the simulated correlation is greater than or equal to the absolute value original correlation is the estimate of the p_r value for the correlation coefficient. (Using absolute values makes this a two-tailed test).

Reanalysis of Regenwetter, et al. (2011)

Applying this approach to the data of Regenwetter, et al. (2011), the estimated p_V and p_r values based on 100,000 simulations are shown in Table 1 for those cases in which $p < 0.2$. Four of the p_V values are “significant” at the $\alpha = 0.05$ level. The binomial probability to find four or more participants out of 18 significant at this level by chance is 0.01. Fifteen of the 18 correlation coefficients are positive, and 6 of them are significantly different from 0 by this two-tailed test. The probability to find 6 or more significant at this level is 0.005. Eight of the 18 participants have significant deviations by one or both of these tests. In summary, one can reject the hypothesis that these data satisfy the assumptions of independence and identical distribution.

Insert Tables 1 and 2 about here.

Table 2 shows data for Participant #2. These data show relatively more responses of “0” at the beginning of the study than at the end. Therefore, the first three or four repetitions resemble each other more than they do the next dozen repetitions, which in turn resemble each other more than they do the first repetitions. Random permutations of the data distribute the "0" values more evenly among rows, which resulted none (zero) of 100,000 random permutations of the data having larger variance than that in the original data. Figure 2 shows the estimated distribution of the variance statistic under the null hypothesis.

Insert Figure 2 here.

For Participant #2, the correlation between mean number of preference reversals and the difference in repetitions in the original data is .91, which is also significant by the p_r simulation.

These tests of iid are not guaranteed to find cases where a person might change true preferences. For example, if a person had exactly two true preference patterns in the mixture that differed in only one choice, it would not produce violations of iid. But in cases where iid is violated, it could be misleading to average data across repetitions to compute marginal choice probabilities (Smith & Batchelder, 2008; Birnbaum, 2011).

Each of these methods (variance or correlation) simplifies the z matrix into a single statistic that can be used to test a particular idea of non-independence. The variance method would indicate cases in which a person randomly sampled from a set of preference patterns in each block of trials, as in the true and error model. The correlation method detects violations of iid in the z matrix that follow a sequential pattern; for example, a positive correlation would be consistent with a case in which a person sticks with one true preference pattern until something causes a change to another pattern. Violations of either would be consistent with the hypothesis that there are systematic changes in "true" preference during the course of the study (Birnbaum, 2011; Birnbaum & Schmidt, 2008). However, there may be more information in the data (and the z matrix) beyond what one or two indices could represent; for example, one might explore the z matrix via multidimensional scaling in order to gain additional insight into the pattern of violation of iid.

In summary, it is possible to test assumptions of iid in small samples, and when these tests are applied, it appears that these assumptions are not consistent with data of Regenwetter et al. (2011). A larger study such as described in Birnbaum (2011) would have greater power and would certainly be a better way to identify and study violations of iid, but this note shows how these assumptions can also be tested with small samples.

References

Birnbaum, M. H. (2011). Testing mixture models of transitive preference: Comments on Regenwetter, Dana, and Davis-Stober (2011). *Psychological Review*, in press.

Birnbaum, M. H., & Schmidt, U. (2008). An experimental investigation of violations of transitivity in choice under uncertainty. *Journal of Risk and Uncertainty, 37,* 77-91.

Regenwetter, M., Dana, J. & Davis-Stober, C. (2010). Testing Transitivity of Preferences on Two- Alternative Forced Choice Data. *Frontiers in Psychology*, *1*, 148. doi: 10.3389/fpsyg.2010.00148.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of Preferences. *Psychological Review*, *118*, 42-56.

Regenwetter, M., Dana, J., Davis-Stober, C. P., and Guo, Y. (in press). Parsimonious testing of transitive or intransitive preferences: Reply to Birnbaum (2011). *Psychological Review*, in press.

Smith, J. B., & Batchelder, W. H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin & Review*, *15*, 713-731. doi: 10.3758/PBR.15.4.713

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76*, 31-48.

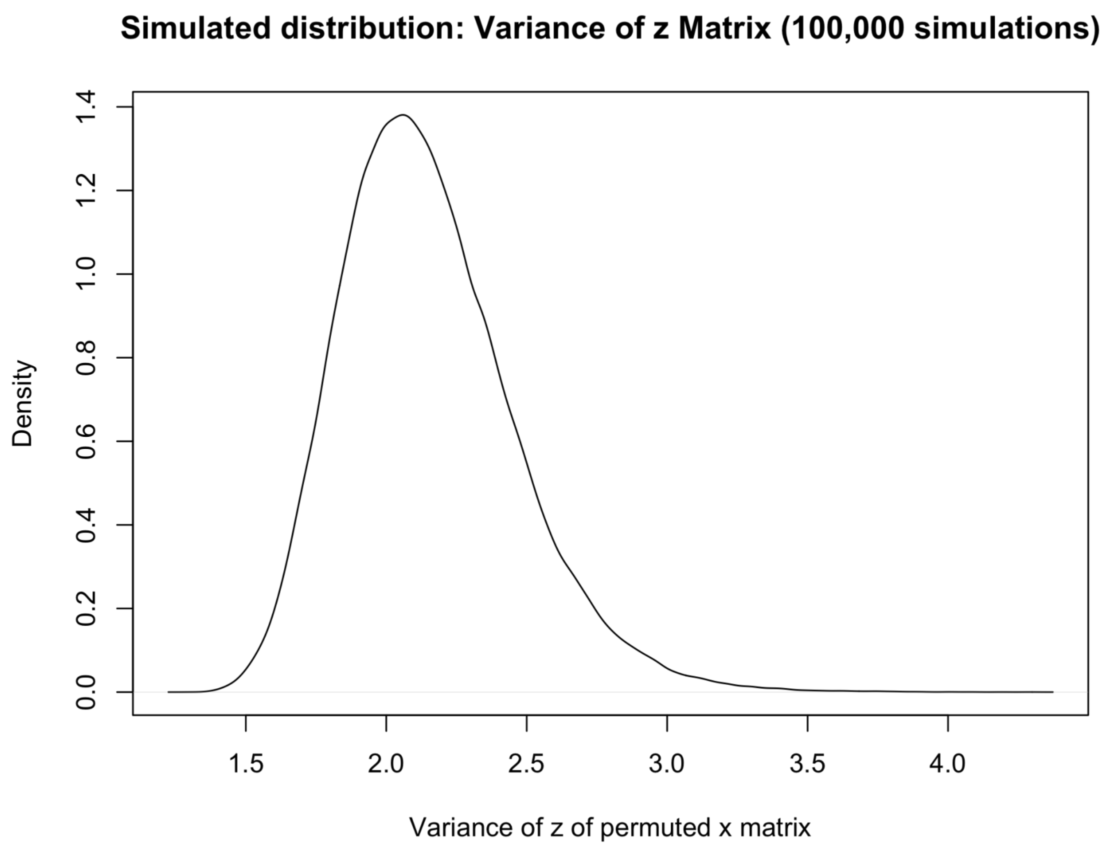
Table 1. Results of Monte Carlo Simulations. The mean of z is the mean number of disagreements between two repetitions out of 10 choices. The variance of z is the variance of these entries, averaged over all entries in z, for the original data. The p_v-values are the proportion of random permutations of the data for which the calculated variance of z in the permuted data is greater than or equal that of the original data (each is based on 100,000 simulated samples). Correlations between mean rate of preference reversals and difference between repetitions are shown in the last column. Corresponding p_r-values are shown in the last column. Value of p greater than 0.2 are not shown.

Participant	Mean of z	Variance of z	p_v	r	p_r
1	3.59	2.89	-	0.61	0.086
2	2.72	4.29	0.000	0.91	0.000
3	0.19	0.16	-	0.61	-
4	2.72	2.61	0.077	0.01	-
5	0.65	1.06	0.011	-0.82	0.104
6	2.67	2.18	0.114	0.67	0.048
7	1.70	1.75	-	0.88	0.005
8	0.35	0.26	-	0.12	-
9	4.33	3.12	-	-0.53	0.107
10	1.23	1.51	0.047	0.45	-
11	0.53	0.48	-	0.11	-
12	3.79	2.68	-	0.44	-
13	4.27	3.47	0.181	0.71	0.011
14	0.19	0.16	-	0.71	-
15	3.48	4.25	0.000	0.82	0.003
16	1.29	0.67	-	0.77	0.024
17	4.32	3.11	-	0.34	-
18	4.09	2.76	-	-0.38	-

Table 2. Individual data for Participant #2. Each row shows results for one repetition of the study.

	V12	V13	V14	V15	V23	V24	V25	V34	V35	V45
1	1	0	0	0	1	1	1	0	1	1
2	0	0	0	0	1	0	1	0	0	1
3	0	1	0	0	0	0	1	1	0	1
4	0	0	1	1	1	1	1	0	1	0
5	1	1	1	0	1	0	1	1	1	1
6	1	1	1	1	1	1	1	1	0	1
7	1	1	1	1	1	1	1	1	1	1
8	0	1	1	1	1	1	1	1	1	1
9	0	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	0	1	1
11	1	1	0	1	1	1	1	0	1	1
12	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1
14	0	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1
17	1	1	1	1	1	1	1	1	1	1
18	0	1	1	1	1	1	1	1	1	0
19	1	1	1	1	1	1	1	1	1	0
20	0	1	1	1	1	0	1	0	1	1

Figure 1. Estimated distribution of variance of the entries in z matrices generated from random permutations of the original data matrix, based on 100,000 Monte Carlo simulations. None of the random simulations exceeded the value observed in the original data, 4.29.



Listing of the program in R.

```
#  This is R code to analyze independence in choice data for each subject
nchoices<-10      # nchoices is the number of choices (columns)
nreps<-20         # nreps is the number of repetitions of the study (rows)
nsubs<-1          # nsubs is the number of subjects.  (set to 18)
nruns<-100000        # nruns is the number of random permutations of the
data (set to 10,000)
outfile="results2.txt"   # outfile is where the results will be printed

 files1<-
c("reg_01.txt","reg_02.txt","reg_03.txt","reg_04.txt","reg_05.txt","reg_06.
txt","Reg_700a.txt","Reg_800a.txt","reg_09.txt","reg_10.txt","Reg_1100a.txt
","reg_12.txt","reg_13.txt","reg_14.txt","reg_15.txt","reg_16.txt","reg_17.
txt","reg_18.txt")

for (iii in 1:nsubs) {

file1<-files1[iii]
mm=read.table(file1)    # read in the data for one subject
x <- mm                 # x (same as mm) is a matrix of the original data

z=array(0,c(nreps,nreps))     # Here arrays are initialized
zz=array(0,c(nreps*nreps))
xperm=array(0,c(nreps,nchoices))
zperm=array(0,c(nreps,nreps))
zzperm=array(0,c(nreps*nreps))
vardist=array(0,c(nruns))
cordist=array(0,c(nruns))
repdif=array(0,c(nreps,nreps))
rrdif = array(0,c(nreps*nreps))
zzap=array(0,c(nreps-1))
sum=array(0,c(nreps-1))

#  These are calculations on the original data
#  z is the matrix of disagreements between reps in original data
for (i in 1:nreps) { for (j in 1:nreps)
{ for (k in 1:nchoices) { z[i,j] = z[i,j]+ (x[i,k]-x[j,k])^2 }
repdif[i,j]<-abs(i-j)
}}
zz<-c(z)
a <- mean(zz)
b <- var(zz)

# here we calculate the correlation between rep difference and distance
nn<-nreps-1
for (id in 1:nn) {
sum[id]<-0
ni<-nreps - id
for (i in 1:ni) {
j<-(i+id)

sum[id]<- sum[id]+ z[i,j]     }
zzap[id]<-sum[id]/(nreps-id) }
repdif2<-c(1:19)
c<-cor(zzap,repdif2)
```

```
# Here begin the calculations on permuted data.  Note that data are
permuted across rows within columns.  This leads to tests of iid
independence.
# xperm is a permutation of the data
# zperm is the matrix of disagreements between reps in the permuted data
# totvar is the number of cases where the variance of permuted data exceeds
the variance in the original data.

totvar=0.0
totcor=0.0
for (kk in 1:nruns) {
for (ii in 1:nreps){
for (jj in 1:nchoices) {xperm[,jj]<-x[sample(nreps,nreps),jj]} }

for (it in 1:nreps) {
for(jt in 1:nreps) {zperm[it,jt]=0} }

for (i in 1:nreps) {
for (j in 1:nreps) {
for (k in 1:nchoices) { zperm[i,j] = zperm[i,j]+ (xperm[i,k]-xperm[j,k])^2
} }}

zzperm<-c(zperm)
a1<-mean(zzperm)
b1<-var(zzperm)
vardist[kk]=b1      # vardist is a vector containing sampled variances of
zperm.
if (b1 >= b) {totvar=totvar+1}

# here we calculate the correlation between rep difference and distance in
permuted data
nn<-nreps-1
for (id in 1:nn) {
sum[id]<-0
ni<-nreps - id
for (i in 1:ni) {
j<-(i+id)
sum[id]<- sum[id]+ zperm[i,j]     }
zzap[id]<-sum[id]/(nreps-id) }
repdif2<-c(1:19)
c1<-cor(zzap,repdif2)
cordist[kk]<-c1
if (abs(c1) >= abs(c)) {totcor=totcor+1}
}

p=totvar/nruns                 # p is the p-value of the test of iid
assumptions
p2=totcor/nruns                # p2 is the p-value of the correlation
o1=c(file1,a,b,p,c,p2,nruns)  # This is the list for printout

sink(outfile,append=TRUE)
print(o1)           # Here the results are printed to the output file
sink()
#  hist(vardist) this would display the histogram sampling distb. under
null hypothesis
# plot(density(vardist))  this would display the density of above histogram
}
```