URN MODELS FOR MARKOV EXCHANGEABILITY

by

Arif Zaman
Purdue University

Technical Report #82-4

Statistics Department
Purdue University

February 1982

URN MODELS FOR MARKOV EXCHANGEABILITY

by

Arif Zaman
Purdue University

Summary

Markov exchangeability, a generalization of exchangeability proposed by de Finetti, requires that a probability on a string of letters be constant on all strings which have the same initial letter and the same transition counts. The set of Markov exchangeable measures forms a convex set. A graph theoretic, and an urn interpretation of the extreme points of this convex set is given.

AMS 1970 subject classifications. Primary 60J10; secondary 62A15,05C35.

Key words and phrases. Extreme point representation, Eulerian paths on digraphs.

1. Introduction

A proability on finite strings of letters is said to be Markov exchangeable if it assigns the same probability to strings which have the same initial letter and the same transition counts (e.g. abbaab, abaabb, aabbab, or aababb). Diaconis and Freedman (1980) consider the problem of expressing the extreme points of the set of Markov exchangeable probability measures. The general solution was posed as an unsolved problem, though they gave an urn model for a two letter alphabet. A direct solution to the general alphabet problem in Zaman (1981) implicitly involves some graph theoretic reasoning. Making an explicit identification between strings of letters and paths on graphs, a Markov exchangeable extreme point corresponds to an Eulerian walk on a graph. The original solution can then be seen as a restatement of the BEST theorem of graph theory, named after the initials of de Bruijn and Ehrenfest (1951) and Smith and Tutte (1948). Similar uses of the BEST theorem to get results for Markov chains abound in the literature, e.g. Dawson and Good (1957), Goodman (1958), as well as the survey paper by Billingsley (1961).

2. Exchangeability and Partial Exchangeability

For the assignment of subjective probabilities, exchangeability has been proposed by de Finetti (1975) as a simplifying assumption reflecting a symmetric type of ignorance. Given random variables $X_1,\ldots,X_n$, if a priori "each random variable is like every other one", then a prior should reflect this ignorance by being unchanged under a reordering of the X's. For example whatever probability is assigned to the event ($X_1 = 1$, $X_2 = 1$, $X_3 = 2$) must also be assigned to the two other rearrangements with one $X = 2$, and the other two X's $= 1$. Some enthusiasts have found exchangeability , a complete replacement for the classical i.i.d. assumption for the reasons that (1) it is an understandable assumption (2) it is the correct Bayesian counterpart to the classical concept of repeated independent identical trials (3) by using de Finetti's theorem in its infinite or finite forms, it is possible to act as if a sequence is i.i.d. with an unknown distribution, starting from the "weaker" assumption of only exchangeability.

To cover more complicated types of symmetry, the definition of exchangeability can be generalized in a number of ways. The special case of interest here is Markov exchangeability, where a sequence $X_1,\ldots,X_n$ is thought of as a time series in which the outcome of a trial may effect the outcome of it's immediate successor. In the spirit of exchangeability, if "every transition (x_i,x_{i+1}) is like every other transition (x_j,x_{j+1})", then two sequences with the same transition counts and the same initial state should be assigned the same probability. For example the probability assigned to the sequence 1,2,1,3,3,2,3,2 must also be given to the sequence 1,2,1,3,2,3,3,2 or 1,3,3,2,3,2,1,2. Any measure satisfying this property is called Markov exchangeable.

3. Extreme Points

It is rather straightforward to verify that if two measures are (Markov) exchangeable, then a mixture of the two with any mixing probability will also be (Markov) exchangeable. Thus the set of exchangeable measures and the set of Markov exchangeable measures are both convex sets.

To find the extreme points of these sets, for any sequence $x = (X_1,\ldots,X_n)$ let $[x]$ denote the set of all sequences which are required to have the same probability as x (e.g. in the case of exchangeability [x] is the set of all reorderings of the elements of x). Define $P_{[x]}$ to be the measure which picks one element equiprobably from the (finite) set [x]. That these measures are all the extreme points and their convex hull forms a simplex follows from a little introspection, or from the finite form of de Finetti's theorem in Diaconis (1980).

As an example, in the case of exchangeability, let x = (1,1,2). Then $P_{[x]}$ picks one of (1,1,2) (1,2,1) or (2,1,1) with probability 1/3. In general $P_{[x]}$ can be seen to be the measure generated by drawing a random sequence without replacement from an urn containing the n items X_1, $\dots, X_n$.

For Markov exchangeability [x] is the set of all sequences with the same initial value X_1, and the same number transitions of each type. Letting $t_{ab}(x)$ represent the number of (a,b) transitions in the sequence x, the set $[x] = \{y : y_1 = X_1, \forall_{a,b}\ t_{ab}(x) = t_{ab}(y)\}$. Though this does completely characterize the extreme points, it has very little intuitive content. The next section considers sequences generated as walks in graphs. These will be used to provide another more intuitive characterization for [x] and the extremal measures $P_{[x]}$.

4. The BEST Theorem

The word graph here will refer to what is sometimes called a multiply connected finite di-graph. Specifically, a graph is a finite set of vertices, here taken to be $\{1,\dots,m\}$ without loss of generality, and an adjacency matrix A, with entries A_{ij} counting the number of edges directed from the vertex i to the vertex j. Any other graph with the same vertex set, and an adjacency matrix B is called a subgraph if $B_{ij} \le A_{ij}$ for all i and j. The indegree of a vertex i, denoted by $D^-(i) = \sum_j A_{ji}$ is the number of edges directed into i. Similarly the outdegree $D^+(i) = \sum_j A_{ij}$. A sequence of vertices $x = x_1, x_2, \ldots, x_n$ is called a walk if each (x_k, x_{k+1}) is an edge. The walk is closed if $x_1 = x_n$. A graph is called a tree if it has no closed walks. A tree is said to be towards a vertex i if $D^+(i) = 0$ and $D^+(j) = 1$ for every other j. Graphically, this is the situation when, from every point other than i, there is exactly one edge out, eventually leading to i. A walk is Eulerian if it uses each edge of the graph exactly once. Using the language of transition counts a walk x is Eulerian if and only if $t_{ij}(x) = A_{ij}$.

The set [x] thus corresponds to all Eulerian walks on the graph with adjacency matrix $t_{ij}(x)$. A constructive method to generate all the Eulerian walks on such a graph is given below. It will be helpful to look at the graphic example in Figure 1.

Given a sequence $x = x_1, \ldots, x_n$ consider the graph with $A_{ij} = t_{ij}(x)$.

1. Let $\mathcal{G}$ be a subtree towards x_n (one always exists). The edges in $\mathcal{G}$ are called emergency exits (not to be used except as a last resort). By definition, there is exactly one emergency exit from each vertex except for x_n, which has none.
2. At each vertex pick some ordering for all the exit edges from there with the condition that the emergency exits be the last in order.
3. Construct a walk by starting at x, and taking the unused edges that are earliest in order at each point along the walk, until the walk can not continue on an unused edge.

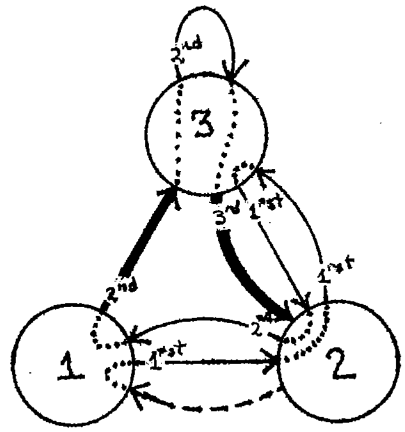


Figure 1: For x = 13233212. The solid lines are the graph with $A_{ij} = t_{ij}(x)$. The addition of the dashed line makes the graph $\tilde{A}$ referred to in the proof. The bold lines form the emergency exit tree. The Eulerian path 12321332 generated by the chosen ordering is shown by the dotted connections.

Theorem (BEST)

Every Eulerian walk starting with x_1 can be generated by an appropriate choice of the tree $\mathcal{G}$ in step 2, and the ordering in step 3. Conversely, every different choice of a tree and an ordering will generate a different Eulerian walk.

Actually the BEST theorem goes further than this, to get a method for counting all the possible trees to x_n, and from there to a formula for the total number of Eulerian paths. The above theorem is actually just a part of the proof of the BEST theorem, but it will suffice. The proof here is adapted from Kastelyn (1967), and does not do justice to that thoroughly readable and detailed proof of the entire BEST theorem.

If a further edge (x_n,x_1) is added to the original graph, then $\tilde{x} = x_n,x_1,\dots,x_n$ is a closed Eulerian walk on the new graph. For this closed walk, every enterance to a vertex x_i, namely every (x_{i-1},x_i) is associated with an exit (x_i,x_{i+1}) from that point. This is true even for (x_{n-1},x_n) , if the exit is identified to be (x_n,x_1). This implies that any other walk $y_0,y_1,\dots,y_k$ using distinct edges which starts with $y_0 = x_n$ and ends with some $y_k \neq x_n$ can always be continued, because for the last entry into y_k, there must be a corresponding unused exit from y_k. We shall now restrict our attention to this augmented graph, and consider the problem of finding all closed Eulerian walks $y = y_0,\dots,y_n$ with $y_0 = x_n$ and $y_1 = x_1$, which is equivalent to the original problem.

For the first part of the theorem, given the Eulerian walk y, order the edges at each vertex in the order that they are used. We need to show that the last edges used from each vertex other than x_n, form a tree to x_n. If any two adjacent edges (i,j) and (j,k) are the last edges used from i and j respectively, then the (j,k) step being the last exit from j msut occur after (perhaps not immediately) the (i,j) step which enters j. So the set of last edges (emergency exits) from vertices other than x_n does not contain a closed loop, because that would represent an ordered sequence of steps in the walk y, the last of which occured before the first. The remaining conditions of being a tree to x_n are easily met by the emergency exits.

To prove the converse, we need show that any walk $y = y_0, y_1, \ldots, y_k$ taken according to the prescription traverses all the edges. It has already been shown that $y_k = x_n$. Assume some edge from some vertex a_0 was not used in y. Then the emergency exit, say (a_0, a_1), must also be unused, as that was to be taken only after all other exits from a_0 were exhausted. If (a_0, a_1), an entry to a_1 is unused, so must an exit from a_1. This implies that the emergency exit from a_1, say (a_1, a_2) was also unused. Since the emergency exits form a finite tree (without cycles) to x_n, this sequence $a_0, a_1, a_2, \ldots$ must eventually lead to x_n. So there is an incoming edge to x_n which is unused. This implies that an exit from x_n remains, contradicting the assumption that the walk y cannot continue.

5. The Extremal Measures

The description of [x] given by the BEST theorem can be used to construct the measure $P_{[x]}$, thought of as the measure picking an Eulerian walk equiprobably from the set of all Eulerian walks. First consider the simple case where there exists only one subtree to x_n. This is always the case when the vertex set is {1,2}. In this case, the measure

$P_{[x]}$ is generated when all possible orderings of the non-emergency exits are equally likely, at each vertex. This will be true if the ordering at each vertex is done independently by draws from urns. With the identification of balls in urns representing the edges of a graph, the construction of a random Eulerian sequence can be rephrased as follows (Fig.2):

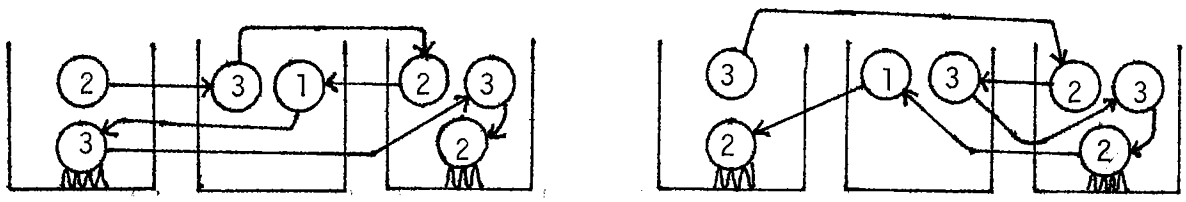
<u>The glued balls scheme</u>

Let $x = x_1,\ldots,x_n$ be a sequence taking values in $\{1,\ldots,m\}$. Let $\mathcal{G}$ be a tree to x_n, with edges (i,g_i) for $i \neq x_n$. Construct m urns, with urn $\mathcal{U}_i$ containing $t_{ij}(x)$ balls labelled j. From each urn $\mathcal{U}_i$, $i \neq x_n$, take one ball labelled g_i (one will always exist if $\mathcal{G}$ is a subtree of a graph with adjacency matrix $t_{ij}(x)$), and glue it to the bottom of $\mathcal{U}_i$. The draws from urns are considered random for the unglued balls, with the glued ball drawn with certainty as the last ball after all the unglued balls have been drawn. Now let $Y_1 = x_1$. Draw Y_2 from U_{Y_1}. Continue drawing Y_i from $U_{Y_{i-1}}$ without replacement, until a draw is forced from an empty urn.

<u>Theorem</u>

If $\mathcal{G}$ is the unique subtree to x_n, then the random sequence Y drawn by the above method has distribution $P_{[x]}$.

Notice the enormous similarity between the method described above, and Markov Chains. If the draws are made with replacement, the resulting measure on sequences would clearly be a Markov measure. Furthermore, this solution can be seen to be the same as the one given for two state Markov chains in Diaconis and Freedman (1980).



(a) (b)

Figure 2: Glued ball models for n = 13233212. The glued balls are shown at the bottom of the urns. In 2a, the sequence 12321332 is generated by the particular sequence of draws depicted by the arrows. Figure 2b shows a different choice of the glued balls tree, and the arrows there correspond to the original sequence n.

Consider now the case where no unique tree exists. This will be the case in most three or more state sequences, unless some transitions do not occur. In this case, for any specific tree $\mathcal{G}$, the previous method generates a uniform distribution $P_{[x,\mathcal{G}]}$, where $[x,\mathcal{G}]$ is the set of all the sequences which can be generated as Eulerian walks with $\mathcal{G}$ as the emergency exits. By the BEST theorem $[x] = \cup[x,\mathcal{G}]$, so $P_{[x]} = \Sigma\alpha_G P_{[x,G]}$ where the constants $\alpha_G = \#[x,g]/\#[x]$ and # denotes the cardinality of a set. Using the combinatorial formulas to compute the numbers α_G, the general urn method for any extremal Markov Exchangeable measure is given by the following theorem.

<u>Theorem</u>

For a given sequence x, choose a tree $\mathcal{G}$ from among all possible trees to x_n, with probability proportional to $\prod_{i\neq x_n} t_{i,g_i}$. For this random choice of the glued balls, a random sequence Y drawn by the glued balls scheme will have distribution $P_{[x]}$.

Notice that the probability distribution on the set of trees is as if a ball g_i is picked independently from urn U_i for each $i \neq x_n$, conditional on the event that these g_i form a tree to x_n.

6. Conclusion

The final urn model for Markov exchangeability seems like a rather contorted construction. Intuitively, the m different urns correspond to the m different probability distributions corresponding to each state. The glued balls represent only a minor modification to ensure a full length sequence. Viewed this way, the model seems a bit more natural.

In the case of exchangeability, the claim was that exchangeability was the basic intuitive concept, and the i.i.d. condition was a mathematical luxury, which is not needed. On the other hand, in this case it seems like the intuition derives from Markov chains, and is used to justify the Markov exchangeability model. Furthermore, Diaconis and Freedman (1980a) show that even an infinite Markov exchangeable sequence is not necessarily a mixture of Markov chains, and further conditions are needed before a result like de Finetti's theorem can be established in this case.

It appears that Markov chains and Markov exchangeability are both fundamental concepts, different and yet similar to each other. There may be times when one might believe a symmetry condition, and times when actual full independence of each Markov transition seems to be more natural. Carrying this analogy further, it appears that the classical i.i.d. condition should not be replaced by exchangeability, even though it probably is overused in cases where exchangeability is a more natural condition.

7. Acknowledgements

I would like to express my gratitude to Prof. Persi Diaconis for suggesting this problem, and for his insistance on understanding rather than proving results.

References

Billingsley, P. (1961). "Statistical methods in Markov chains," Ann. Math. Stat. Vol. 32, pp. 12-40.

Dawson, R. and Good, I. J. (1957). "Exact Markov probabilities from oriented linear graphs," Ann. Math. Stat., Vol. 28, pp. 946-956.

De Bruijn, N. G. and Van Aardenne-Ehrenfest, T. (1951). "Circuits and trees in oriented graphs," Simon Stevin, Vol. 28, pp. 203-217.

De Finetti, B. (1975). Theory of Probability, Vol 7. Wiley, New York.

Diaconis, P. and Freedman, D. (1980). "De Finetti's generalizations of exchangeability" in Studies in Inductive Logic and Probability, (R. Jeffrey ed.) Univ. of Calif. Press, Berkeley.

Diaconis, P. and Freedman, D. (1980a). "DeFinetti's theorem for Markov chains," Ann. Prob., Vol. 8, pp. 115-130.

Goodman, L. (1958). "Exact probabilities and asymptotic relationships for some statistics from m-th other Markov chains," Ann. Math. Stat. Vol. 29, pp. 476-490.

Kastelyn, P. W. (1967). "Graph theory and crystal physics," in Graph Theory and Theoretical Physics (f. Harary ed.), pp. 44-110. Academic Press, New York.

Smith, C. A. B. and Tutte, W. T. (1948). "On unicursal paths in a network of degree 4," Amer. Math. Monthly, Vol. 48, pp. 233-237.

Zaman, A. (1981). An Approximation Theorem for Finite Markov Exchangeability, Ph.D. Thesis, Stanford University.