The Annals of Probability
1984, Vol. 11, No. 4, 223–229

URN MODELS FOR MARKOV EXCHANGEABILITY[1]

BY ARIF ZAMAN

Florida State University

Markov exchangeability, a generalization of exchangeability that was 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.

1. Introduction. Let V be a finite set and $\{U(u)\}_{u\in V}$ be a collection of urns. Each urn $U(u)$ contains a total of $\sum_{v\in V} a_{uv}$ balls, with a_{uv} of them labeled v. Choose a fixed $X_1 \in V$ and construct a random sequence $X_1, X_2, \cdots, X_n$ by letting X_{i+1} be the label on a ball drawn from the urn $U(X_i)$. It is clear that the resulting sequence is a Markov chain if the draws are done at random with replacement. When the draws are done without replacement, after some draw X_n the urn $U(X_n)$ will be empty, so that the ball X_{n+1} can't be drawn. The probability distribution on these finite random sequences, with some modification (namely conditioning on the event that $n = 1 + \sum a_{uv}$, i.e. all balls from all urns are used), is an example of a Markov exchangeable distribution.

A probability 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 solution to the general alphabet was given in Zaman (1981) in terms of the urn models mentioned in the previous paragraph. The proof in that paper can be simplified considerably by using a well known identification between strings of letters and paths 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). The BEST theorem has been used before to get results for Markov chains, e.g. Dawson and Good (1957), Goodman (1958), as well as the survey paper by Billingsley (1961).

Sections 2 and 3 review Markov exchangeability and the finite form of de Finetti's theorem. Sections 4 and 5 review the basics of graph theory and the BEST theorem. In Section 6, these are combined so that the BEST theorem can be applied to draws from urns. The urn model for the extremal Markov exchangeable measures is given in Section 7. Section 7 may be read by itself, even though the previous sections develop the necessary background.

Received April 1982; revised July 1983.
[1] Partially supported by USARO grant DAAG29-82-K-0186
AMS 1980 *subject classifications.* Primary 60J10; secondary 62A15, 05C35.
Key words and phrases. Extreme point representation, partial exchangeability, Eulerian paths.

2. 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 $X_1, \cdots, X_n$, 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 trials and (3) by using de Finetti's theorem for finite or infinite sequences of exchangeable random variables it is possible to act as if a sequence is i.i.d. with some unknown distribution, starting from the "weaker" assumption of only exchangeability.

In a Markov chain, $X_1, \cdots, X_n$, the state X_i can only affect the probabilities of its immediate successor X_{i+1}. 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 prior probability. For example, the probability assigned to the sequence 1, 2, 1, 3, 3, 2, 3, 2 must also be given to 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.

Defining things formally, let V^n denote all the sequences of length n taking values in a finite set V and let "$\sim$" be some equivalence relation on V. Variables which are sequences are shown in bold print. A measure P on V^n is called partially exchangeable with respect to "$\sim$" iff

$$\forall\ \mathbf{x}, \mathbf{y} \in V^n \quad \mathbf{x} \sim \mathbf{y} \Rightarrow P(\mathbf{x}) = P(\mathbf{y}).$$

For example, define "$\simeq$" by $\mathbf{x} \simeq \mathbf{y}$ iff the sequence $\mathbf{x}$ is some permutation of $\mathbf{y}$. Ordinary exchangeability is easily seen to be partial exchangeability with respect to "$\simeq$". As another example, let $t_{uv}(\mathbf{x}) = \#\{i\colon x_i = u,\ x_{i+1} = v\}$ which is the number of $u \rightarrow v$ transitions in $\mathbf{x}$. Define "$\approx$" by $\mathbf{x} \approx \mathbf{y}$ iff $x_1 = y_1$ and $t_{uv}(\mathbf{x}) = t_{uv}(\mathbf{y})$ for every $u, v \in V$. A measure is Markov exchangeable iff it is partially exchangeable with respect to "$\approx$".

3. De Finetti's theorem. Given a relation "$\sim$", it is of interest to describe the set of all probability measures which are partially exchangeable with respect to it. Theorem 1, sometimes referred to as the finite form of de Finetti's theorem, provides a simple description of this set.

Let $[\mathbf{x}, \sim]$ denote the equivalence class of $\mathbf{x}$ under the relation $\sim$. For any set $A \subseteq V^n$ let P_A denote the uniform probability measure on A.

THEOREM 1 (de Finetti). *The set of all probability measures partially exchangeable with respect to $\sim$ is a simplex with extreme points of the form $P_{[\mathbf{x},\sim]}$ for $\mathbf{x} \in V^n$.*

A nice discussion of this theorem is given in Diaconis and Freedman (1980). In the form given above, the theorem follows almost directly from the definition of partial exchangeability. Its power comes when the extremal measures $P_{[\mathbf{x},\sim]}$ can be given a simple interpretation. As an example, for exchangeability let $\mathbf{x} = (1, 1, 2)$. Then $P_{[\mathbf{x},\simeq]}$ picks one of (1, 1, 2), (1, 2, 1), or (2, 1, 1) each with probability 1/3. In general any extremal measure $P_{[\mathbf{x},\simeq]}$ can be seen to be a random sequence drawn without replacement from an urn containing the n items $x_1, \cdots, x_n$.

For Markov exchangeability $[\mathbf{x}, \approx]$ is the set $\{\mathbf{y} \in V^n: y_1 = x_1, \forall u, v \in V t_{uv}(\mathbf{x}) = t_{uv}(\mathbf{y})\}$. Although this does completely characterize the equivalence classes, and hence $P_{[\mathbf{x},\approx]}$, it has very little intuitive content. The next section considers sequences generated by walks on graphs, which will be used to provide an "urn interpretation" for $[\mathbf{x}, \approx]$ and $P_{[\mathbf{x},\approx]}$.

4. Graphs. Let (V, E) be a finite directed graph. For $u, v \in V$ let $E(u, v)$ be the set of all edges in E directed from u to v. Let $E^+(u) = \cup_{v\in V}E(u, v)$; $d^+(v) = \#E^+(v)$ is known as the outdegree of the vertex v. Similarly $E^-(v) = \cup_{u\in V}E(u, v)$, and $d^-(v) = \#E^-(v)$ is known as the indegree of v. A sequence of n vertices $\mathbf{x} \in V^n$ and $n - 1$ edges $\mathbf{e} \in E^{n-1}$ is called a walk if $e_i \in E(x_i, x_{i+1})$ for $i = 1, \cdots, n - 1$. The walk is closed if it uses each edge exactly once. A graph is called a tree if it has no closed walks. A graph (V, F) is called a subtree of (V, E) towards $v \in V$ iff (V, F) is a tree, $F \subseteq E$, and for each $u \in V$, $u \neq v$ there is a unique edge $f(u) \in E^+(u)$ such that $F = \{f(u): u \neq v\}$. Graphically, this is the situation when from each vertex other than v there is exactly one edge leading out, eventually leading to v.

5. The BEST theorem. Define an exit order of a graph as a choice of a special vertex v_0 and for each vertex $v \in V$ an edge sequence $\mathbf{r}(v) \in E^{d^+}(v)$ which contains all the edges in $E^+(v)$ in some order. Note that an exit order assigns an order to the edges which lead out from any given vertex. Given an exit order $(v_0, \{\mathbf{r}(v)\}_{v\in V})$, a unique walk $(\mathbf{x}, \mathbf{e})$ can be constructed by letting $x_1 = v_0$, and continuing from any x_i by picking the first edge (in the order specified by $\mathbf{r}(x_i)$) that is unused to proceed to the next vertex. When the walk reaches some x_n where all the exit edges have been used, it will be terminated.

It is clear that only walks which do not reuse edges can be constructed by the above method. Two different exit orders may give rise to the same walk as is shown in Figure 1. On the other hand, every Eulerian walk can be constructed by an exit order, and that order is unique. For an Eulerian walk $(\mathbf{x}, \mathbf{e})$ the unique exit order corresponding to it is given by $v_0 = x_1$ and $\mathbf{r}(v)$ is the subsequence of $\mathbf{e}$ containing all the elements of $E^+(v)$. The exit orders which correspond to Eulerian walks are identified in the following theorem.

THEOREM 2 (BEST). *An exit order $(v_0, \{\mathbf{r}(v)\}_{v\in V})$ corresponds to an Eulerian walk* iff

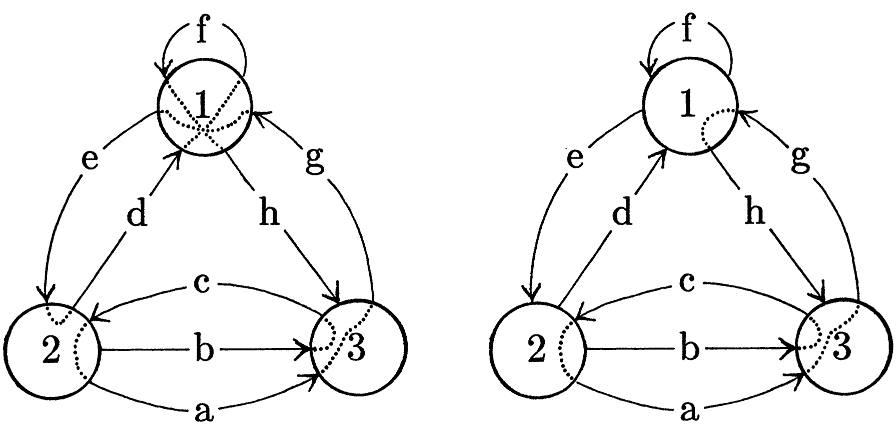
(i) $\exists v_1 \in V$ *such that* $\forall v \in V \quad d^+(v) + \delta_{v_1}(v) = d^-(v) + \delta_{v_0}(v)$.

(ii) *Let* $F = \{r_{d^+(v)}(v) \colon v \neq v_1\}$. *Then* (V, F) *is a subtree towards* v_1 *of* (V, E).

Note that condition (i) is only based on d^+, D^- and v_0, which refer only to the structure of the graph; and so if it is not satisfied, there are no Eulerian walks on that graph starting from v_0. If the graph permits Eulerian walks from v_0 (i.e. (i) holds), (ii) identifies the exit orders which produce Eulerian walks. When (i) is satisfied, the vertex v_1 is unique and all Eulerian walks from v_0 end on v_1.

A very clear proof of this theorem, as well as further results on the number of subtrees or the number of Eulerian walks possible on a graph, can be found in Kastelyn (1967).

6. Walks and urns. We shall make a correspondence between the urn model proposed in the introduction and walks on a graph. Let the vertex set of the graph be V and identify each urn $U(v)$ with the vertex v. Each of the a_{uv} balls in $U(u)$ labeled v is to be identified with an edge from u to v, so that $U(u)$ contains all the balls (edges) corresponding to $E^+(v)$. Thus the edge set corresponds to the set of all balls in all the urns. A walk corresponds to choosing an initial vertex (urn) x_1, and then at each step picking an edge (ball) e_i from the set $E^+(x_i)$(urn $U(x_i)$), and moving on to the vertex (urn) x_{i+1} given by the out direction of the edge (by the label on the ball). Notice that since each edge is considered distinguishable from every other, we shall treat by each ball as distinguishable from every other ball, even if two balls have the same label and are in the same urn.



$(\mathbf{x}, \mathbf{e})$ *with* $\mathbf{x} = (2, 3, 2, 3, 1, 2, 1, 1, 3)$ *and* $\mathbf{e} = (b, c, a, g, e, d, f, h)$ *is an Eulerian walk generated by the exit order* $\nu_0 = 2$ *and* $r(1) = (e, f, h)$, $r(2) = (b, a, d)$, $r(3) = (c, g)$.

$(\mathbf{x}, \mathbf{e})$ *with* $\mathbf{x} = (2, 3, 2, 3, 1, 3)$ *and* $\mathbf{e} = (b, c, a, g, h)$ *is generated by the exit order* $\nu_0 = 2$ *and* $r(1) = (h, e, f)$, $r(2) = (b, a, d)$, $r(3) = (c, g)$. *If* $r(1)$ *is changed to* (h, f, e) *the same walk is generated.*

FIG. 1. *Examples of exit orders and their associated walks on the graph* (V, E) *where* $V = \{1, 2, 3\}$ *and* $E = (a, b, \cdots, h)$. *The vertices are shown as circles, and the edges as arrows. The dotted lines within the circles show the walks.*

In order to get an Eulerian walk, a ball once drawn should be discarded (no edge is used more than once) and all urns should be empty at the end of the walk (each edge is used). For an Eulerian walk $(\mathbf{x}, \mathbf{e})$ with exit order $(x_1, \{\mathbf{r}(v)\})$, for any fixed $v \in V$ the sequence $\mathbf{r}(v)$ corresponds to the order in which the balls are drawn from the urn $U(v)$.

To introduce probability in this setting, completely sample each of the urns $U(v)$ at random without replacement to get sequences $\mathbf{r}(v)$ for each $v \in V$. Select a fixed $v_0 \in V$ to get a random exit order $(v_0, \{\mathbf{r}(v)\})$ with its associated probability measure P. By the BEST theorem, either there exists some unique final vertex v_1 satisfying condition (i) of the theorem, or no Eulerian walks are possible. We assume the former, and define two other random variables based on this random exit order just defined. Let $F = \{r_{d^+(v)}(v): v \neq v_1\}$ be the random set of last balls from urns other than $U(v_1)$. Let $(X, \mathbf{e})$ be the walk associated with the random exit order, if there is one. As further notation let $\varepsilon(v_0)$ denote the set of all Eulerian walks on (V, E) starting from v_0, and $\tau(v_1)$ denote all subtrees of (V, E) towards v_1. Then the BEST theorem implies that $\{(\mathbf{X}, \mathbf{e}) \in \varepsilon(v_0)\} =\{(V, F) \in \tau(v_1)\}$, and so $P(\cdot \mid (\mathbf{X}, \mathbf{e}) \in \varepsilon(v_0)) = P(\cdot \mid (V, F) \in \tau(v_1))$ when $\varepsilon(v_0)$ is not empty. We are nearly done because $P(\cdot \mid (\mathbf{X}, \mathbf{e}) \in \varepsilon(v_0))$ is very close to the elusive measure $P_{[\mathbf{x},\approx]}$ that we want, while $P(\cdot \mid (V, F) \in \tau(v_1))$ has a relatively simple interpretation.

To see the relationship between $P_{[\mathbf{x},\approx]}$ and $P(\cdot \mid (\mathbf{X}, \mathbf{e}) \in \varepsilon(v_0))$ we will show that the random variable X has the same distribution under the two probability measures. Because P is an urn measure, it is uniform on all exit orders, and hence $P(\cdot \mid (\mathbf{X}, \mathbf{e}) \in \varepsilon(v_0))$ is uniform on all exit orders corresponding to Eulerian walks. Looking only at the distribution of $\mathbf{X}$ the random vertex sequence under $P(\cdot \mid (\mathbf{X}, \mathbf{e}) \in \varepsilon(v_0))$, the number of edge sequences $\mathbf{e}$ which combine with $\mathbf{X}$ to make an Eulerian walk is a constant independent of $\mathbf{X}$, so P is a uniform probability measure on all Eulerian vertex sequences, i.e. on $\{\mathbf{x}: (\mathbf{x}, \mathbf{e}) \in \varepsilon(v_0)$ for some $\mathbf{e}\}$. Since all Eulerian walks have the same transition count, this last set $\{\mathbf{x}: (\mathbf{x}, \mathbf{e}) \in \varepsilon(\nu_0)$ for some $\mathbf{e}\} = [\mathbf{y}, \approx]$ for some $\mathbf{y} \in V$. To summarize, we simply have that there is some $y \in V^n$ such that for all $A \subseteq V^n P(\mathbf{X} \in A \mid (\mathbf{X}, \mathbf{e}) \in \varepsilon(v_0)) = P_{[\mathbf{y},\approx]}(\mathbf{X} \in A)$.

A simple urn model for $P(\cdot \mid (V, F) \in \tau(v_1))$ is obtained by first considering $P(\cdot \mid F = F_0)$ for some given $(V, F_0) \in \tau(v_1)$, with $P(F = F_0) > 0$. By the definition of a tree towards v_1 there is a set of edges $f(v)$ for $v \neq v_1$ so that $F_0 = \{f(v): v \neq v_1\}$. The random set $F = F_0$ if and only if for every urn $U(v)$ with $v \neq v_1$, the last draw from $U(v)$ is $f(v)$. A simple way to accomplish this is for each $v \neq v_1$ to take one ball labeled $f(v)$ from $U(v)$, and "glue" it to the bottom of that urn. The draws from such a "glued ball urn" will be assumed to be just like random draws from an urn, with the glued ball impossible to draw until all others have been removed.

We have now established the equivalence of

$$P_{[\mathbf{x},\approx]}(\cdot) = P(\cdot \mid \mathbf{X} \in \varepsilon) = P(\cdot \mid (V, F) \in \tau(v_1))$$

$$= \textstyle\sum_{F_0 \in \tau(v_1)} P(F_0 \mid F \in \tau(v_1)) P(\cdot \mid F = F_0).$$

It only remains to combine the various steps into one full procedure. The next section gives the complete description of an urn method to get the extremal Markov exchangeable measures.

7. The glued balls method. We first give a description of a method using urns to generate the extremal Markov exchangeable measures. This is followed up by some comments at the end, which while necessary to be precise, make the description long.

0 For a finite set V, select the parameters $v_0 \in V$, $v_1 \in V$, and $a_{uv} \geq 0$ for all u, $v \in V$, satisfying

$$\textstyle\sum_{u \in V} a_{vu} + \delta_{v_1}(v) = \sum_{u \in V} a_{uv} + \delta_{v_0}(v)$$

for all $v \in V$.

1 Construct urns $U(u)$ for $u \in V$ such that $U(u)$ contains a total of $\sum_{v \in V} a_{uv}$ balls with a_{uv} balls labeled v for each $v \in V$.

2 For each $v \neq v_1$, let $f(v)$ be the label of a ball selected at random from $U(v)$.

3 If $(V, \{f(v)\})$ is not a tree to v_1 (see comments) return to the previous step.

4 Glue the ball $f(v)$ at the bottom of $U(v)$ for each $v \neq v_1$, so that it does not interfere with random drawings of other balls, but itself can only be drawn after all others have been removed.

5 Let $X_1 = v_1$, and let X_{i+1} be the label of a ball drawn from urn $U(X_i)$, until after some X_n the urn $U(X_n)$ is empty.

The final sequence $\mathbf{X} \in V^n$ is a single sample from some extremal distribution $P_{[\mathbf{x},\approx]}$, the particular distribution determined by the choice of parameters in step 0. For a specific $\mathbf{x} \in V$, the appropriate choice of parameters to get $P_{[\mathbf{x},\approx]}$ is $v_0 = x_1$, $v_1 = x_n$ and $a_{uv} = t_{uv}(x)$ the transition counts of x.

In the third step, the notation $(V, \{f(v)\})$ refers to the graph that would be if the $f(v)$ are thought of as edges from the vertex v to the vertex $f(v)$. In fact all references to graph theory can be dropped by replacing that step with "If there is some $v \neq v_1$ and $k > 0$ for which $f^k(v) = v$ then return to the previous step," where f^k is understood to be the composition of f (considered as a function $f: V - v_1 \mapsto V$) with itself k times.

Finally, the repetition possible between steps 2 and 3 can be avoided by replacing the two steps with:

Let $\mathscr{F}$ represent all trees to v_1, or in the language of the preceeding comment, let $\mathscr{F} = \{f(\cdot): \forall v \neq v_1, \forall k > 0\, f^k(v) \neq v\}$. Select $f(\cdot) \in \mathscr{F}$ at random according to the probabilities

$$P\{f(\cdot)\} = \frac{\prod_{v \neq v_1} a_{v,f(v)}}{\sum_{g(\cdot) \in \mathscr{F}} \prod_{v \neq v_1} a_{v,g(v)}}.$$

This probability is simply $P(F = F_0 \mid F \in \tau(v_1))$ for $F_0 = \{f(v): v \neq v_1\}$ as defined in the previous section.

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

In the case of exchangeability, a claim was that exchangeability was the fundamental intuitive concept, and that the i.i.d. condition was a mathematical luxury which is not needed. On the other hand, in this case it seems as if the intuition derives from Markov chains with independent transitions; and Markov exchangeability seems to be a nonintuitive concept which can be justified or understood in the light of the more reasonable Markov chains. Furthermore, Diaconis and Freedman (1980a) show that even an infinite Markov exchangeable sequence is not necessarily a mixture of Markov chains, so that further conditions are needed before a result like de Finetti's theorem for infinite sequences can be established in this case.

It appears that Markov chains and Markov exchangeability are both fundamental and different concepts. There may be times when a prior belief may satisfy a symmetry condition, and times when the full independence of each Markov transition may more closely reflect a true prior belief. Carrying this analogy further, it appears that the classical i.i.d. condition cannot be replaced by exchangeability, even though it probably is overused in cases where exchangeability is a more natural condition.

9. Acknowledgements. I would like to thank Professor Persi Diaconis for suggesting this problem, and for his insistence on understanding rather than proving results.

REFERENCES

[1] BILLINGSLEY, P. (1961). Statistical methods in Markov chains. *Ann. Math. Statist.* **32** 12–40.
[2] DAWSON, R. and GOOD, I. J. (1957). Exact Markov probabilities from oriented linear graphs. *Ann. Math. Statist.* **28** 946–956.
[3] DE BRUIJN, N. G. and VAN AARDENNE-EHRENFEST, T. (1951). Circuits and trees in oriented graphs. *Simon Stevin* **28** 203–217.
[4] DE FINETTI, B. (1975). *Theory of Probability* Vol 1. Wiley, New York.
[5] DIACONIS, P. and FREEDMAN, D. (1980). de Finetti's generalization of exchangeability. *Studies in Inductive Logic and Probability*, (ed. Jeffrey, R.) Univ. of Calif. Press, Berkeley.
[6] DIACONIS, P. and FREEDMAN, D. (1980a). de Finetti's theorem for Markov chains. *Ann. Probab.* **8** 115–130.
[7] GOODMAN, L. (1958). Exact probabilities and asymptotic relationships for some statistics from m-th order Markov chains. *Ann. Math. Statist.* **29** 476–490.
[8] KASTELYN, P. W. (1967). Graph theory and crystal physics. *Graph Theory and Theoretical Physics*, (ed. Harary, F.) 44–110. Academic, New York.
[9] SMITH, C. A. B. and TUTTE, W. T. (1948). On unicursal paths in a network of degree 4. *Amer. Math. Monthly* **48** 233–237.
[10] ZAMAN, A. (1981). *An Approximation Theorem for Finite Markov Exchangeability.* Ph.D. Thesis, Stanford University.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306