The Point of Point Crossover: Shuffling To Randomness

Anil Menon
Whizbang! Labs (East)
4616 Henry Street,
Pittsburgh, PA 15213
anilm@acm.org

Abstract

The action of point crossover is modeled as a random walk on a group, and convergence and rate results are established for the walk. Specifically, it is shown that there is a cut-off phenomenon in the rate at which the sample get randomized. As long as the number of crossover steps is less than a certain critical number, the total variation distance (with respect to the stationary distribution) is large, and remains essentially constant. But once the critical number has been crossed, the total variation distance goes to zero (at an exponential rate). The cut-off number of steps is of order of $O(lN \ln N)$ steps, where N is the sample size, and l is the length of the chromosome. Finally, it is shown by heuristic arguments as well as by simulations, that if a statistical criterion such as Kendall's W coefficient or the average Kendall's τ coefficient is used to measure randomness (rather than total variation distance), the sample can be said to be random (upto statistical significance) in $O(\ln N)$ steps, rather than $O(lN \ln N)$ steps. The properties of such criteria are characterized.

1 Introduction

The repeated application of point crossover on a finite set of chromosomes may be viewed as a random walk on a certain graph. The aim of this paper is to show that there is a cut-off phenomenon associated with a class of such "crossover walks." Roughly, the existence of a cut-off means that if the number of times point crossover applied in the crossover phase, n, is less than a certain critical number n^*, the sample remains "far" from stationarity, but for $n > n^*$, the sample becomes very "close" to stationarity.

There have been a variety of approaches to analyzing the role of point crossover, including (to list a few) hyperplane and schema analysis[13], dynamical systems models [5], and explicit Markov modeling [11]. However, despite the strong similarities between certain random walks and the crossover operator (for example, base swapping walks on matroids), not much work has been done to explore this connection, though there are a few outstanding exceptions [12]. In particular, the relationship between crossover walks and cut-off phenomena appears to have been overlooked.

Cut-off phenomena ("phase transitions") in random walks, especially those associated with walks on groups, have been intensely studied with great success in the last two decades [2]. The basic machinery behind these results draws upon deep results from the representation theory of groups. The techniques were first applied to study the effectiveness of various card shuffling operations, such as riffle shuffles, perfect shuffles and transposition shuffles. Intuitively, there is a great deal of similarity between shuffling sets of cards and the point crossover operator. In a sense, this paper formalizes this intuition. We eschew a too-rigorous presentation of results, and focus instead on heuristic arguments and simulations that will, hopefully, inspire a much more rigorous analysis.

The structure of the paper is as follows. In Section 2 the concept of a crossover walk is introduced. The question of its convergence is resolved by using techniques from the theory of doubly stochastic matrices. An analysis of the rate of convergence of the crossover walk is taken up in Section 3. In Section 4 it is argued that the traditional criterion used to measure the degree of randomness, namely, the variation distance, may be unnecessarily strict, and two alternate criteria are introduced. Section 5 presents simulations on the

behavior of these alternate measures, and Section 6 introduces an informal model to explain them.

Notation: S_n will denote the symmetric group on n symbols (the permutation group). All logarithms are to base e. Results drawn from external sources are referred to as "Propositions."

2 Crossover Walks

A chromosome is defined to be an element in Σ^l, where Σ is some finite alphabet. Any set of chromosomes defines a *sample*. The size of a sample is the number of chromosomes in it and two samples are distinct if they contain different numbers of any given chromosome in Σ^l. Let $\mathcal{S}_l(N)$ (or simply, $\mathcal{S}_l$) denote the set of samples of size N.

The k-point crossover operator $\times_k : \mathcal{S}_l \rightarrow \mathcal{S}_l$ maps one sample to another, and is defined as follows. Select a non-empty set of indices $I \subset \{1, 2, \ldots, l\}$ with respect to the "subset" measure Pr_k (explained below). Let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_l)$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_l)$ represent two chromosomes drawn uniformly from the input sample. Then, $\times_k(\alpha, \beta) = (\alpha', \beta')$ where for all $j \in I$, $\alpha'_j = \beta_j$ and $\beta'_j = \alpha_j$. For all $j \notin I$, $\alpha'_j = \alpha_j$ and $\beta'_j = \beta_j$. The resulting chromosomes α', β' are referred to the "children" of the "parent" chromosomes α, β.

The subset measure Pr_k is used to handle the differences between 1-point, 2-point crossover etc. All point crossover operators select a subset of indices from $\{1, \ldots, l\}$, and swap the corresponding alleles from the parents at these indices. 1-point crossover (1PTX) will always select subsets of the form $\{l\}$, $\{l-1, l\}$, $\{l-2, l-2, l\}$ etc. The symmetric version of 1PTX will also select subsets of the form $\{1\}$, $\{1, 2\}$, $\{1, 2, 3\}$ etc. 2-point crossover (2PTX) will select subsets of the form $\{i, i+1, i+2, \ldots, i_k\}$ for $i \geq 1$ and $i_k \leq l$. In other words, each point crossover variant merely imposes an uniform probability measure on the set of all subsets of $\{1, 2, \ldots, l\}$. This probability measure is unique to each point crossover, and is denoted Pr_k.

The point crossover operator is usually seen as mapping chromosomal pairs to other chromosomal pairs, rather than one sample to another. While operationally there is no difference between the two views, modeling point crossover as taking samples to samples is analytically more convenient (at least for our purposes).

Define the *crossover graph* $G(S) = (V(S), E)$ where the node set $V(S) = \{v_1, \ldots, v_{|V|}\}$ is the set of all *distinct* samples that can be generated from the initial sample S by means of k-point crossover. $V(S)$ has at least one member, namely S, and $V(S)$ is a finite set because $\mathcal{S}_l$ is finite. Since k-point crossover replaces one pair with another, all samples in G have the same number of chromosomes.

Two nodes v_i and v_j in G are connected by an edge in the edge set E iff it is possible to generate v_j from v_i in *one* application of the $\times_k$ operator. In particular, every node v_i is connected to itself (self loop). The graph G is also connected (every node is reachable from every other node).

Let $\chi(S) = \{\times_k^j(S)\}_{j \geq 0}$ denote the *sequence* of samples that is generated by repeated applications of $\times_k$ on a sample S. Clearly, the sequence of samples in $\chi(S)$ then represents a random walk — the crossover walk — on the graph $G(S)$.

The two fundamental questions concerning the crossover walk on G are:

- Does the walk converge to a stationary distribution?
- If so, what is its rate of convergence?

The remainder of this section tackles the first question, and the rest of the paper considers the second.

2.1 Walk Convergence

The convergence to an stationary distribution can be established using a technique due to Feller [4, section XV.10]. As mentioned earlier, $\chi(S)$ represents the sequences of samples (nodes) encountered in walking the graph $G = (V, E)$. Define $P^t = (p_1(t), p_2(t), \ldots, p_{|V|}(t))$ where $p_i(t)$ is the probability that at time instant t, the walk finds itself at node v_i of the graph G. In other words, it is the probability that $\times_k^t(S) = v_i$. Initially, P^0 is a vector of all zeros except at one index r (say), corresponding to the fact that the walk starts at $S = v_r \in V$. From the definition of point crossover, the distributions at two successive instants of time are related by,

$$P^{t+1} = Q\,P^t \tag{1}$$

where $Q = [q_{i,j}]$ is a $|V| \times |V|$ sized matrix. The elements of Q can be assumed to be time independent, since the probability of moving to v_j given that the walk is at v_i should depend only the composition of the sample represented by v_i and v_j. Theorem 1 states the conditions under which the walk defined by Equation (1) converges.

Theorem 1 Let Q be the aperiodic, time independent transition matrix for a walk on the crossover graph $G(S) = (V(S), E)$. Then, Q is a doubly stochastic matrix, and the sequence $P^0, P^1, P^2, \ldots$, converges to the stationary distribution $P^\infty = (1/|V|, \ldots, 1/|V|)$. ∎

Proof: We first show that Q is a doubly stochastic matrix, that is, $\sum_{j=1}^{|V|} q_{i,j} = 1$ and $\sum_{i=1}^{|V|} q_{i,j} = 1$.

The matrix element $q_{i,j}$ is actually a conditional probability representing the probability of reaching v_j in one step, given that the walk is at v_i. By definition of conditional probabilities, $\sum_j q_{i,j} = 1$.

Observe that for every transition $(\alpha, \beta) \rightarrow (\alpha', \beta')$ produced by a k-point crossover operator, there corresponds a transition $(\alpha', \beta') \rightarrow (\alpha, \beta)$. In other words, if the k-point crossover operator transforms a sample $v_i \rightarrow v_j$, then it can also transform $v_j \rightarrow v_i$. Now, the expression $\sum_i q_{i,j}$ represents the probability that the vertex i can be reached from *some* vertex j. Since this can always be done, we conclude that $\sum_i q_{i,j} = 1$. Since Q is both column stochastic as well as row stochastic, Q is doubly stochastic.

From a standard result in Markov chain theory (for example, [4, section XV.7]) we know that if Q is persistent, irreducible and aperodic[1] then the sequence $P^0, P^1, P^2, \ldots$ converges to a stationary distribution. Because Q is doubly stochastic, it converges to the stationary distribution $P^\infty = (1/|V|, 1/|V|, \ldots, 1/|V|)$ [4, section XV.7, example 7(h)]. Q is persistent because it is doubly stochastic, and the construction of G guarantees that the walk on G is irreducible. From the aperiodicity, persistence and irreducibility of the walk, it follows that it converges to the stationary distribution P^∞. Q.E.D

The stationary distribution in Theorem 1 is related to but *not* the same as the linkage equilibrium distribution (which refers to the distribution of *chromsomes* in a sample randomized by crossover operations). Also, the assumption that Q is aperiodic is a trivial one, since any Markov chain can be redefined to be aperiodic [4, section XV.5].

Theorem 1 asserts that the repeated application of k-point crossover on a sample eventually randomizes it, and shows that point crossover belongs to a class of models known as quadratic dynamical systems [9]. The relationship between double stochasticity and point crossover leads to the Theorem 2. It shows that the class of Schur-convex functions are Lyapanuv functions for the crossover walk. A great deal is known about this class [8], and its functions occupy much real estate in mathematics[2]. A necessary and sufficient condition for a continuous function $\phi : R^n \rightarrow R$ to be Schur-convex is that it be symmetric ($\phi(x_1, \ldots, x_n) = \phi(x_{i_1}, \ldots, x_{i_n})$ and for any i, j, $(x_i - x_j)(\partial\phi/\partial x_i - \partial\phi/\partial x_j) \geq 0$.

Theorem 2 Let Q be the transition matrix for a walk on the crossover graph $G(S) = (V(S), E)$. Let $P^t = (p_1(t), p_2(t), \ldots, p_{|V|}(t))$ where $p_i(t)$ is the probability that at time $t \geq 0$, the walk finds itself at node v_i. If $F : R^{|V|} \rightarrow R$ is a Schur-convex function, then for all $t \geq 0$, $F(P(t+1) \leq F(P(t))$. ∎

Proof: The theorem is an immediate consequence of three facts: (1) $P(t+1) = QP(t)$, (2) Q is doubly stochastic (Theorem 1) and (3) the Hardy, Littlewood, Polya theorem (see [8, chap. 2, B.2] and [8, chap. 3, A.1]. Q.E.D.

3 Rate of Convergence

Any analysis of the convergence rate of a crossover walk depends on the "intrinsic" aspects of the walk such as the transition probabilities, and the exact k-point crossover used. In particular, it depends on the structure of the graph G. Since the structure of the graph is determined by the initial sample, the convergence rate of the walk is dependent on it.

The dependency of the walk on the initial sample complicates matters, perhaps unnecessarily so. Suppose initial sample consists of identical chromosomes. Clearly, point crossover is not going to change the composition of the sample. In this scenario, $\chi(S) = \{S, S, \ldots\}$ and $G(S) = (\{S\}, E)$ where E consists of a single self-loop. On the other hand, consider the effect of point crossover on a sample drawn randomly from Σ^l. In this case too, point crossover has no effect since the sample is already randomized, and a walk on G is essentially a walk on a random graph. It is difficult to study the general walk on G, because G can take on so many different "shapes" depending on how the initial sample was set up.

An analogy might make this idea clearer. Suppose one wished to analyze the shuffling of a pack of cards (pos-

[1] A Markov chain $Q = [q_{i,j}]$ is said to be *persistent* if it is certain that the chain starting from a state v_i will eventually return to v_i. The chain is said to be *irreducible* iff every state v_i can be reached from any other state v_j. Q is said to be *aperiodic* if $q^t_{i,i} \neq 0$ for any $t > 1$.

[2] For example, the Shannon Entropy function is a Schur-concave function (that is, negentropy is Schur-convex). So are almost all of the popular diversity metrics, such as sample variance and the Gini coefficient. It is *not* necessary that a function be continuous in order for it to be Schur-convex.

sibly incomplete), where the shuffle operation consists of transposing (with some probability) pairs of cards drawn from the pack w.r.t some probability measure. As stated, the problem is hard to study because it mixes the critical issue (the effect of random transposition) with the less important ones (possibly incomplete packs, unknown initial card distribution, transposition frequency, a measure on drawing cards etc). It is for this reason that most random transposition models assume an initially sorted, complete pack, where every shuffling step results in a transposition (unless the same card is picked twice).

Of course, the decision as to which factors are important and which ones are not, depends on the problem one is interested in. For example, if the problem is to study how point crossover "undoes" the effect of proportional selection, then the dependency on the initial sample has to be taken into account. But if the problem is (say) to prove that repeated applications of point crossover leads to linkage equilibrium, then the specifics of the initial sample is not too important (as Theorem 1 demonstrates).

What is needed is a *reference sample* against which the effectiveness (as measured by rates of convergence) of various crossover operators can be tested. In other words, a reference sample will enable the distinction between "what crossover does" from "what crossover is applied to."

The definition of the point crossover operator suggests that its action is roughly analogous to a shuffling operation on sets of decks of cards. Accordingly, a good reference sample to study its convergence rates should be an array of permutations. Specifically, consider an array of numbers (N rows and l columns), arranged as follows:

$$S = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 2 & 2 & \cdots & 2 \\ \vdots & \vdots & \cdots & \vdots \\ N & N & \cdots & N \end{pmatrix} \tag{2}$$

Each *row* in the array is interpreted as a "permutation" chromosome of length l. Upon applying k-point crossover on S, the columns of the array will tend to get randomized. For example, one such sample is,

$$S' = \begin{pmatrix} 1 & 2 & \cdots & 2 \\ 2 & 1 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ N & N & \cdots & N \end{pmatrix} \tag{3}$$

Of course, k-point crossover only disarrays a column, and does not change the allelic composition, at any step of the walk. Hence, the sample can be represented as a vector of permutations $(\eta_1, \ldots, \eta_l)$, with $\eta_j \in S_N$ where S_N is the permutation group on N symbols. Initially, $S \equiv ((1, 2, \ldots, N), \ldots, (1, 2, \ldots, N))$. The action of k-point crossover at any step consists of selecting at random a non-empty, proper subset of components from $(\eta_1, \ldots, \eta_l)$ and applying a random transposition drawn from the permutation group S_N to each of those components. For example the move from $S \rightarrow S'$ involves selecting the components 2 through l and applying the transposition $(1, 2)$ on each of those components.

In other words, the underlying crossover graph has $(N!)^l$ nodes. An edge connects two nodes $v_i = (\sigma_1, \ldots, \sigma_l)$ and $v_j = (\eta_1, \ldots, \eta_l)$ iff there exists a transposition $\tau \in S_n$ and a subset $I \subset \{1, 2, \ldots, l\}$, such that,

$$\eta_j = \begin{cases} \tau(\sigma_j) & \text{if} \quad j \in I, \\ \sigma_j & \text{otherwise.} \end{cases} \tag{4}$$

The effect of k-point crossover on the N "permutation" chromosomes is thus identical to a nearest-neighbor random walk on the nodes of the crossover graph.

Of course, real GAs do not typically operate on permutation strings, so the relevance of the above reference sample may be in question. It would appear however that the walk on $G(N, l)$ provides a good test case for analytic techniques, is related to active areas of research in probability theory, and focuses on point crossover's central feature, namely, its tendency to "shuffle" a sample's alleles. Furthermore, there are techniques to "lift" walks on symmetric groups to walks on the hypercube [2, pp. 19-20], so results may transfer as well.

3.1 Walks on $G(N, l)$

Since the structure of this graph is completely determined by the parameters N and l, it will be denoted $G(N, l)$ rather than the usual $G(S)$. It is not hard to show that $G(N, l)$ is a regular graph, where each node is connected to $d = (2^l - 1)\binom{N}{2}$ other nodes.

The analysis of the convergence rate of the walk on $G(N, l)$ depends on the criterion used to measure the "degree of randomness." One criterion popular in models of random walk on groups [2, chap. 3B] is the variation distance of two discrete distributions P and P' defined by,

$$||P - P'|| = \frac{1}{2} \sum_{i=1}^{n} |p_i - p'_i| \tag{5}$$

In particular, if for any specified $\epsilon > 0$, there exists a t_c such that for all $t > t_c$, $||P^t - P^\infty|| < \epsilon$, then

the walk is said to converge to P^{∞} w.r.t the variation distance.

A formal analysis of the rate of convergence of $||P^t - P^{\infty}||$ is complicated and involves the machinery of the representation theory of the symmetric group. Also, in this paper, the quantity of interest is not the variation distance but a different function (described in the next section). Hence, we will only present heuristic arguments for the rate of convergence of the total variation distance on $G(N,l)$.

Consider the case when $l = 1$. The problem then reduces to the random walk induced by the action of random transpositions on the symmetric group S_N. P. Diaconis and M. Shahshahani proved the following result regarding such walks [3].

Proposition 1 Let $k = \frac{1}{2}N \ln N + cN$, where $c > 0$. Then, there exists an universal constant a such that, $||P^t - P^{\infty}|| \leq a \exp(-2c)$. Correspondingly, let $k = \frac{1}{2}N \ln N - cN$, where $c > 0$. Then, there exists an universal constant b such that, $||P^t - P^{\infty}|| \geq 1 - b \exp(-2c)$. ∎

Proposition 1 shows that $k = \frac{1}{2}N \ln N + cN$ are sufficient for the variation distance to become "small". Conversely, $k = \frac{1}{2}N \ln N - cN$ are also necessary. This is the celebrated "cut-off' phenomenon, the value of $||P^t - P^{\infty}||$ is large for $k < \frac{1}{2}N \ln N - cN$, but is small after $k \geq \frac{1}{2}N \ln N + cN$.

This case provides a lower bound on the convergence rate of the random walk on $G(N,l)$. On the other hand, it can be shown using several different ways[3] that $\frac{Nl}{2} \ln N$ steps is an upper bound on the number of steps necessary to achieve stationarity

In summary, it takes at most $O(Nl \ln N)$ k-point crossover operations to randomize an array of N chromosomes of length l, where the starting sample is given by Equation (2). If the reference sample consists of N binary chromosomes each of length l, rather than permutation strings, the conclusions of the analysis of the walk on $G(N,l)$ does not fundamentally change. For example, consider the crossover walk on a reference sample of binary chromosomes where half the sample is initially "all 0"' chromosomes, and the other half are "all 1" chromosomes. This walk can be shown to be closely related to the Bernoulli-Laplace urn model [2, pp. 56-58], for which the cutoff number of steps again turns out to be $O(Nl \ln N)$.

[3]Wald's principle offers one route. Another option is to note that a walk on $G(N,l)$ can be described as a walk on the Cartesian product of the transposition graph of the symmetric group, and then use Chung's results [1, pp. 36-41].

But are such "$O(N \ln N)$" results of any practical use? In most GAs, the sample size N, and the chromosome length l, are both quite large (typically). For $N = 500$ and $l = 30$, the above result would indicate that the number of crossover steps is of the order of $100,000$ steps. Thus, the number of crossover steps required to randomize the sample is quite large, and it would appear that randomization of the sample in the point crossover phase never happens in real GA deployments.

However, in the next section it will be argued that the large number of crossover steps required to randomize the walk on $G(N,l)$ (and by association, walks on general samples), is an artifact of the variation distance criterion. The rate of convergence of two alternative criterion, the Kendall's average τ coefficient, and Kendall's W coefficient, give a very different picture on the minimal number of steps required to randomize the walk on $G(N,l)$.

4 Convergence Criteria

As far as convergence is concerned, the exact norm used to measure the distance between two distributions is not of great importance, since norms are (topologically) equivalent (so convergence w.r.t one norm implies convergence w.r.t another). But for bounds on convergence *rates*, the choice of the norm is very important [14].

The variation distance may be inappropriate in some natural context. Suppose one is given a set of decks, where each deck is arranged in some manner (not necessarily sorted). It is now required to be determined whether the cards in the decks are randomly ordered or not. The variation distance is not a very meaningful measure in this case. The statistical solution is to compute some ranking statistic on the card arrangements, and see if the null hypothesis (card are randomly arranged in each deck) can be rejected. Clearly, this idea can be also applied to each sample produced in the random walk on $G(N,l)$. Statistical tests have been evolved to test for randomness (upto specified levels of significance). It makes sense to use them to test whether the sample produced by crossover at any stage passes these tests. If it does, then we have a rigorous basis for a stopping rule.

In this case, the sample consists of permutations, and it is natural to study the change in rank-based concordance measures as a function of the stage in the random walk on the crossover graph. Two such measures will now be considered. The first, Kendall's W coefficient, is a measure of ranking concordance and

the second, Kendall's average tau coefficient, is usually interpreted as a measure of disarray of in a set of permutations.

Consider l judges ranking N objects. Each judge assigns a distinct rank to each object. The rankings can be arranged in an array of the type shown in Equation (2), where each column represents a ranking, and the i^{th} row reflects how each judge ranks the i^{th} object. Let s_{ij} denote the ranking of the i^{th} object by the j^{th} judge.

If all the columns are identical as in Equation (2), it indicates complete concordance between the judges. A standard measure of concordance is *Kendall's W coefficient* [6, chap. 6] defined as follows:

$$W = \frac{12\,D}{l^2 N(N^2-1)}, \tag{6}$$

$$\text{where,} \quad D = \sum_{i=1}^{n}(\sum_{j=1}^{l} s_{i,j} - \mu)^2, \tag{7}$$

$$\text{and,} \quad \mu = \frac{l(N+1)}{2}. \tag{8}$$

The idea is to compute for each object, the sum of the ranks assigned to it by the l judges. The sum of the squares of the deviation of each sum from the expected value μ then gives D. Kendall's W is the ratio of D with the maximum possible value. Kendall's W coefficient is widely used to measure the agreement in l rankings of a common set of objects [6]. W always lies between 0 and 1, with 1 indicating complete agreement between the rankings of the l judges.

The second ranking statistic is the *average Kendall coefficient* [2, chap. 6]. Let σ and π be any two permutations in S_N (permutation group on N symbols). Let $K(\sigma,\pi)$ $(= K(\pi,\sigma))$ be the number of adjacent transpositions required to convert the permutation σ^{-1} to π^{-1}. Kendall's τ coefficient for the pair σ and π is defined by,

$$\tau(\sigma,\pi) = 1 - \frac{4K(\sigma,\pi)}{N(N-1)}. \tag{9}$$

The τ coefficient lies between -1 and 1 (inclusive) and behaves likes a correlation coefficient. When $\tau = 1$, the permutations are identical, and when $\tau = -1$, $\sigma = \pi^{-1}$. The *average Kendall coefficient* for a set of permutations $(\sigma_1, \sigma_2, \ldots, \sigma_l)$ is given by,

$$\bar{\tau} = \frac{2\sum_{i,j=1}^{l}\tau(\sigma_i,\sigma_j)}{l(l-1)}. \tag{10}$$

Both these measures are designed to measure the degree of disarray in a sample of permutations, and their asymptotic behavior is well understood[4][6, chap. 6]. This enables their practical use in statistical significance tests.

The next section studies the change in these ranking statistics as a function of the crossover walk on $G(N,l)$ for various values of N and l. The associated graphs not only show the existence of cut-off behavior in these functions for crossover walks, but also show that, remarkably, the number of steps required to achieve randomness (upto statistical significance) is of the order of $O(\ln N)$ rather than $O(lN\ln N)$.

5 Simulations

Here the behavior of the crossover walk on the graph $G(N,l)$ for various values of N and l are studied. The basic procedure for setting up the simulations was to start with the ordered $N \times l$ array shown in Equation (2). Then, k-point crossover was repeatedly applied (usually for 100,000 steps). Each application of the operator corresponds to a step on the graph $G(N,l)$. After applying the operator, the values of the average τ and/or Kendall's W-coefficient for the sample are computed. These values are then plotted against the *logarithm* of the step number [5]. The shape of the curve, its critical points and sensitivity to the three independent variables, namely, N, l and k, are the main topics of interest. Here, only the results for fixed l and k but varying N are presented. It is worth mentioning however, that all scenarios show the existence of the cut-off phenomena, though the exact point at which cut-off happens, changes as the dependent variables are changed.

Figure 2 and Figure 3 show the plots obtained by sampling the values of Kendall's W coefficient and Kendall's average τ after every 1PTX step of the random walk on $G(N,l)$ for $N = 50, 100, 150, 200$ and $l = 15$. The curves become smoother for larger values of N, but in general the behavior is relatively insensitive to changes in values of N. Consider the point at which the Kendall's W coefficient falls below 0.5. For $N = 50, 100, 150, 200$, this happens (roughly) at $\exp(3.2) \approx 25$, $\exp(4.8) \approx 122$, $\exp(5.2) \approx 181$ and

[4] For example, the asymptotic distribution of Friedman's function $\chi_r^2 = l(N-1)W$ can be shown to be approximately χ^2 with $N-1$ degrees of freedom for large l. On the other hand, the average τ can be shown to distributed normally.

[5] The cutoff phenomenon implied by the sigmoid growth curve disappears if the ranking statistics is plotted directly against the step number. This may be one reason why the cut-off phenomena in ranking statistics for random walk models appears to have escaped the attention of probability theorists.

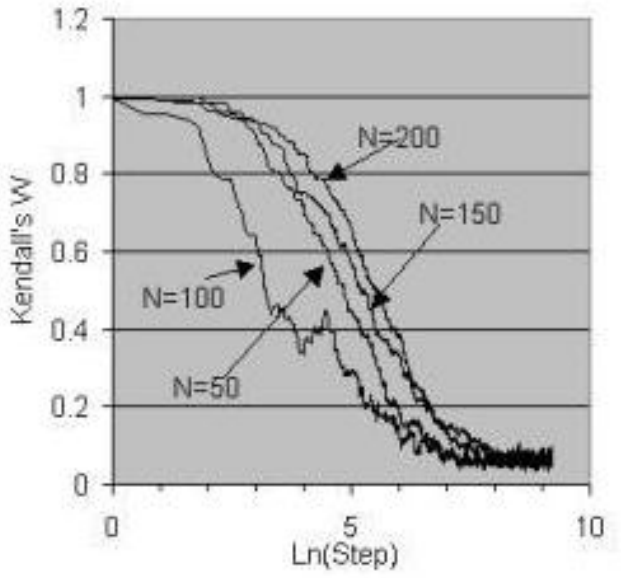

Figure 2: Kendall's W versus Ln(Crossover Step) for various population sizes

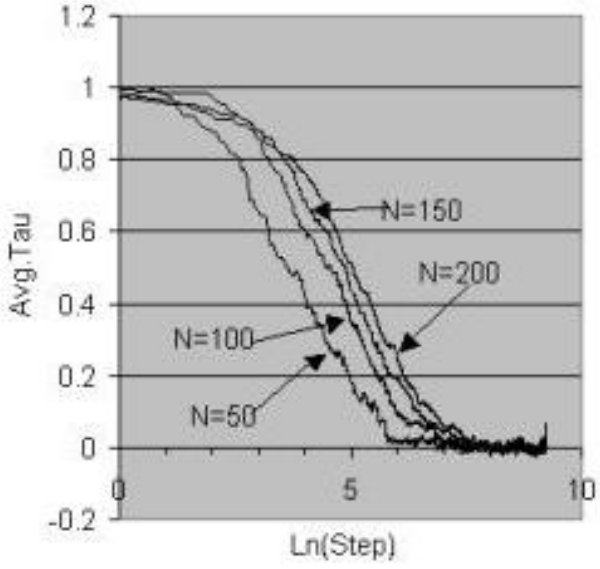

Figure 3: Kendall's Average Tau versus Ln(Crossover Step) for various population sizes

$\exp(5.57) \approx 262$ number of steps, respectively. This may seem like a significant change, but if considers the *ratio* of the number of steps (to get to the halfway point) to the sample size, the ratios are roughly constant.

The most significant aspect of these figures is how soon the cut-off point manifests itself. For $N = 200$, by about $\exp(2)$ to $\exp(3)$ number of steps, the Kendall's W starts to fall sharply, and by about $\exp(7)$ to $\exp(8)$ number of steps, it reaches its equilibrium value. Thus a few hundred applications of point crossover does have a significant impact on the value of Kendall's W coefficient for the walk on $G(N, l)$. Similar comments hold for the average τ.

6 Logistic Models

A rigorous analysis of the behavior of Kendall's W coefficient, or the average τ coefficient is likely to be very complicated. Yet, the curves are so simple in their shape that it is very tempting to believe that an equally simple explanation must be available.

In this section, an explanation based on an population growth model will be developed; it is simplistic, but the basic idea is very general and holds much promise (e.g. [7, 10]).

The problem is to model the change in a statistic Y w.r.t. the step number. Suppose it was the case that there were two kinds of events that affected the growth of Y. The "good" events cause it to increase, but the "bad" events cause it to decrease. It is also given that if only bad events happen, then the relative change in Y is inversely (directly) proportional to the relative change in r. One way to model this is,

$$\frac{\Delta Y}{Y} = \alpha(Y)\frac{\Delta t}{t}, \quad \alpha > 0, \tag{11}$$

$$\frac{\Delta Y}{Y} = -\beta(Y)\frac{\Delta t}{t}, \quad \beta > 0. \tag{12}$$

The linear relationship has been setup not between Y and t, but between the relative growths $\Delta Y/Y$ and $\Delta t/t$. The basic reason for this is that cutoff phenomena persist under scaling changes, that is, cannot be removed by ratio transformations of the dependent and independent variables. If the differential equation we are constructing is to exhibit cut-off phenomenon, then it has to be invariant under ratio transformations as well. Equations (11) and (12) have this property.

The quantities α and β have been marked as a function of Y but not time. The reason for this is that the dynamics of *any* two variables U and V can be related vacuously by a "constant" that varies with respect to U and V. To prevent this, the proportionality constants can depend at most on Y. The dependency on Y models the fact that Y, being a ranking statistic, cannot grow ceaselessly. Since it is a ranking statistic, it takes on at a finite number of values (there are only a finite number of rankings, and each ranking corresponds to one value for the statistic).

α and β are duals to each other. Assume without loss of generality that $\alpha + \beta = 1$ and $\alpha \leq 1$ (if they are not, the equations can always be rescaled to make it so). Then, for some function $g(Y)$, the functions α and β can be expressed as,

$$\alpha(Y) = 1 - \frac{g(Y)}{K}, \quad (13)$$

$$\beta(Y) = \frac{g(Y)}{K}. \quad (14)$$

where K is large enough to make $\alpha, \beta \leq 1$. Putting the above equations together,

$$\frac{\Delta Y}{Y} = \frac{\Delta t}{t}(1 - 2\frac{g(Y)}{K}). \quad (15)$$

Passing to the limit implies,

$$\frac{dy}{dt} = \frac{y}{t}(1 - 2\frac{g(y)}{K}). \quad (16)$$

To "draw" the above curve with respect to the logarithmic axis, set $t = \ln x$. Consequently,

$$\frac{dy}{dx} = y(1 - 2\frac{g(y)}{K}). \quad (17)$$

Equation 17 produces a sigmoid curve under very mild restrictions on the function $g(y)$. The case $g(y) = y$ leads it to the classic Verhulst-Pearl equation of (sigmoid) growth.

The assumptions behind this heuristic argument are minimal. All that is required is that Y be density limited, its growth has to be explainable by a two factor model (good events/bad events), and $(\Delta Y/Y) \propto (\Delta t/t)$.

Consider Kendall's average τ coefficient. Since it is a ranking statistic, it cannot grow without bounds. Every application of point crossover splits the permutations in the sample into two groups, namely, those that got affected by the crossover, and those that didn't. The τ coefficient of each pair changes only linearly with every crossover step[6]. The τ coefficients within each group do not change, but the inter-group τ-coefficients do change. The extent of that change is proportional to the product of the relative sizes of the two groups, and hence the *log* of the changes is linearly proportional to the logs of the relative sizes.

The growth in the average τ is also driven by a two-factor model, because the change in τ is driven by a two factor-model. A "good" change consist of a transposition which reduces the disarray, and hence increases the value of τ. Correspondingly, a bad transposition is one which increases the disarray and reduces the value of τ.

[6]Recall that the τ-coefficient of a pair of permutations is an affine function of the number of adjacent transpositions needed to transform one permutation to the other.

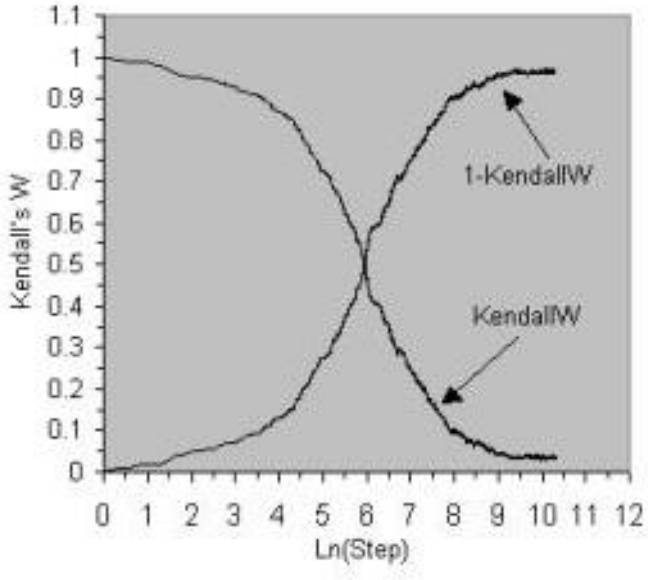

Figure 4: Kendall's W versus Ln(Crossover Step) for 1PTX walk on $G(300, 30)$ (sample of 300 chromosomes each of length 30)

Similar arguments can be made for Kendall's W coefficient, though the details are a lot more tedious. In any event, these arguments are meant to be suggestive of the possibilities of an alternative to the currently popular group-theoretic approaches.

Finally, Figure 4 shows a plot of Kendall's W for repeated applications of 1PTX on a population of 300 permutations each of length 30. Clearly, even for these small population sizes and non-trivial chromosome lengths, the sigmoid growth curve is obtained. Notice that by approximately $6\log(N)$ steps, the measure falls to its mid value (0.5).

7 Conclusion

What does point crossover do? The results of the paper formalize the intuition that repeated applications of point crossover "shuffles" a sample's alleles. The formalization was achieved by modeling the action of point crossover as a random walk on the crossover graph. Two aspects of this walk were studied. First, it was demonstrated that this walk is characterized by a homogeneous doubly stochastic Markov chain and hence may be shown to converge to a stationary distri-

bution. Second, the rate of convergence was analyzed, and it was shown that there is a cut-off phenomenon in the rate at which the original sample get randomized by the repeated action of point crossover. As long as the number of crossover steps is less than a certain critical number, the total variation distance (with respect to the stationary distribution) is large, and remains essentially constant. But once the critical number has been crossed, the total variation distance goes to zero (at an exponential rate). The cut-off number of steps is of order of $O(lN \ln N)$ steps, where N is the sample size, and l is the length of the chromosome. If different metrics are considered, say, Kendall's W or average τ coefficient, then simulation indicate that cut-off occurs at $O(N \ln N)$ rather than $O(lN \ln N)$. A heuristic explanation based on population arguments was provided for the general sigmoid nature of these curves. The existence of the cut-off suggests that point crossover is something of an all-or-nothing randomization operator. Apply it for more than the cut-off number, and the sample is rapidly randomized. Apply it for less, and as far as randomization is concerned, the sample remains far from random. Whether such phase transitions exist for other crossover operators remains an open question.

8 Acknowledgements

I would like to thank the two reviewers whose comments helped to make this a better paper. I would also like to thank Dennis Wakefield for his assistance in helping me meet the submission deadline, and Saras Sarasvathy for her general support and encouragement.

References

[1] Fan R. K. Chung. *Spectral Graph Theory*. Number 92 in Regional Conference Series in Mathematics. American Mathematical Society, Providence, Rhode Island, 1997.

[2] P. Diaconis. *Group representations in probability and statistics*. IMS Lecture Notes—Monograph Series, Hayward, CA, 1988.

[3] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, 57:159–179, 1981.

[4] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume I. Wiley Eastern, New York, Third edition, 1968.

[5] S. Forrest and G. Mayer-Kress. Genetic algorithms, nonlinear dynamical systems, and global stability models. In L. Davis, editor, *The Handbook of Genetic Algorithms*. Van Nostrand Reinhold, New York, NY., 1991.

[6] M Kendall and D. Gibbons. *Rank correlation methods*. Edward Arnold, London, fifth edition, 1990.

[7] T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability*, 7:49–58, 1970.

[8] A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and its Applications*. Academic Press, New York, 1979.

[9] A. Menon, K. Mehrotra, C. Mohan, and S. Ranka. Replicators, majorization and genetic algorithms: New models, connections and analytical tools. In R. Belew and M. Vose, editors, *Foundations of Genetic Algorithms*, volume 4, pages 155–180. Morgan Kaufman, 1997.

[10] M. Mitzenmacher. Studying balanced allocations with differential equations. Technical Report 1997-024, SRC Technical Note, Digital Systems Research Center, CA, 1997.

[11] E. A. Nix and D. M. Vose. Modeling genetic algorithms with markov chains. *Annals of Math. and Artificial Intelligence*, 5:79–88, 1992.

[12] Y. Rabani, Y. Rabinovich, and A. Sinclair. A computational view of population genetics. *Random Structures and Algorithms*, 12(4):313–334, 1998.

[13] W. M. Spears and K. DeJong. An analysis of multi-point crossover. In *Foundations of Genetic Algorithms*, volume 1, pages 301–215. Morgan Kaufman, 1991.

[14] Lloyd N. Trefethen and Lloyd M. Trefethen. How many shuffles to randomize a deck of cards? *Proc. Royal Society London, Series A*, 456:2561–2568, 2000.