Six Ways of Looking at Burtin's Lemma

S. Anoulova, J. Bennies, J. Lenhard, D. Metzler,
Y. Sung, and A. Weber

In an article in this MONTHLY in 1953 Metropolis and Ulam asked for the expected number of components of the graph induced by a purely random mapping of a set of n points into itself [7]. This problem was solved one year later by L. Kruskal [6]. In 1955, L. Kac [4] computed the probability that this random graph is connected, that is, that the number of components is 1. In 1981, S. Ross [9] treated the same questions for more general random mappings in which the function values are independent and identically distributed but not necessarily uniform. The fundamental lemma of [9] had been proved earlier by the young Russian mathematician Y. D. Burtin, several months before his death in 1977 [2, Prop. 1]:

Burtin's lemma: Let $F(1), \dots, F(n)$ be independent and identically distributed with $P\{F(i) = j\} = p_j$, $j = 0, \dots, n$. Generate a random directed graph Γ with vertices $\{0, \dots, n\}$ by drawing an edge from i to $F(i)$ for each $i = 1, \dots, n$. Then this graph is connected with probability p_0.

The proofs of Burtin's lemma given in [2] and [9] both use induction on n, lumping together all those vertices directly connected to 0. In 1984, still another inductive proof was given by Jaworski [3]. The six ways of looking at Burtin's lemma presented here arose from the attempt of each of us to understand it better.

Burtin's lemma has interesting connections with the random generation of spanning trees. Assign to each pair (i, j) $(1 \le i \le n,\ 0 \le j \le n)$ the weight p_j, and consider the problem of generating a random spanning tree on $\{0, \dots, n\}$ with root 0 such that the probability of a tree is proportional to the product of all its edge weights. Propp and Wilson [8] describe how to do this in a straightforward way: Start at an arbitrary element i of $\{1, \dots, n\}$, and consider the iterates $F(i), F^2(i), \dots$ until the first time they hit 0. Delete the cycles of the path from i to 0 and take this as the trunk of the tree. Then take an element j not on the trunk (if there are any), proceed in the way described until you hit the trunk, delete the cycles, and so on. Burtin's lemma then tells us that the probability of never producing a cycle in all these attempts is just p_0. In fact, our Proof 4 computes this probability directly. Proof 5 provides a more economical algorithm for generating a random spanning tree with the desired distribution.

Propp and Wilson [8] also treat the more general scenario in which the weights of an edge (i, j) can be of the form p_{ij} rather than p_j only. They also relate the problem of generation of random spanning trees to that of exact simulation of the equilibrium of a Markov chain by their algorithm of "coupling from the past." In [8, Sec. 1.3], they give a short history of random spanning tree generation, including references both to the algebraic method (which relies on variants of the so called matrix tree theorem [1, Ch. 2, Thm. 8] and the method using Markov chains.

And now to the proofs. Let A (for "acyclic") be the event that Γ contains no cycle. One easily sees that A can also be described as the event that

1) Γ is a (directed) tree; or
2) Γ is connected (as an undirected graph); or
3) there is a directed path from each vertex $i = 1, \dots n$ to 0. We then say "Each $i = 1, \dots n$ chooses, directly or indirectly, the vertex 0."

The first proof is a direct computation of $\Pr(A)$.

Proof 1: We enumerate all graphs from each of whose vertices $1, \dots, n$ all choose, directly or indirectly, the vertex 0. Let m_i be the number of immediate predecessors of the vertex i. Obviously $m_0 \geq 1$ and $\sum_{i=0}^{n} m_i = n$.
Construction:

a) Given $\{m_i\}$, choose successively for each vertex $i = 1, \dots, n$ its m_i immediate predecessors. For $i = 1$ there are $\binom{n-1}{m_1}$ possibilities; $F(1) = 1$ is not allowed since this would create a cycle. For $i = 2$ there are two cases:
   (i) If $F(2) = 1$, then $F(1) = 2$ is not allowed and there are $\binom{n - m_1 - 1}{m_2}$ possibilities.
   (ii) If $F(2) \neq 1$, then $F(2) = 2$ is not allowed and there are also $\binom{n - m_1 - 1}{m_2}$ possibilities.

   This argument holds for all k so in general there are $\binom{n - m_1 - \cdots - m_{k-1} - 1}{m_k}$ possibilities to choose the m_k immediate predecessors of vertex k.
b) The m_0 vertices that are left are the immediate predecessors of the vertex 0.

In all there are

$$\binom{n-1}{m_1}\binom{n - m_1 - 1}{m_2} \cdots \binom{n - m_1 - \cdots - m_{n-1} - 1}{m_n}$$
$$= \binom{n-1}{m_0 - 1, m_1, \dots, m_n}$$

possibilities to construct such a graph for given $m_0, \dots, m_n$.
Summing over all m_i we get

$$\Pr(A) = \sum_{\substack{m_0 \geq 1, \\ \sum m_i = n}} \binom{n-1}{m_0 - 1, m_1, \dots, m_n} p_0^{m_0} \cdots p_n^{m_n}$$
$$= p_0(p_0 + \cdots + p_n)^{n-1} = p_0. \qquad \blacksquare$$

Proof 2 is inductive and works by lumping together an individual n and the one it chooses.

Proof 2: Suppose the proposition is proved for $n - 1 \geq 2$ individuals; for $n = 2$ (exactly one voter) it is trivial. We partition A according to the choice of individual n.

If $F(n) = 0$, we combine the individuals 0 and n into one new individual and start again with $n-1$ individuals. By the induction hypothesis the contribution of this case to $\Pr(A)$ is $p_0(p_0 + p_n)$.

If $F(n) = j$, $j \in \{1, \ldots, n-1\}$, we combine individuals j and n and obtain probability weights $p_0, p_1, \ldots, p_j + p_n, \ldots, p_{n-1}$; hence this case contributes the probability $p_j p_0$.

Finally, if $F(n) = n$, we have a cycle, so this case contributes nothing.

Summing over all these possibilities we get:

$$\Pr(A) = p_0(p_0 + p_n) + \left(\sum_{j=1}^{n-1} p_j\right) \cdot p_0 = p_0\left(\sum_{j=0}^{n} p_j\right) = p_0. \qquad \blacksquare$$

The next proof is also by induction, but in this case we show that lumping together individuals $n-1$ and n does not change $\Pr(A)$. Iterating this argument then proves the proposition.

Proof 3: We imagine that only the individuals $1, \ldots, n-2$ have already chosen their successors.

We consider two scenarios:

(i) with individuals $n-1$ and n left to make their choice,
(ii) with the composite individual $(n-1)'$, which is obtained by lumping together individuals $n-1$ and n. This individual has weight $p_{(n-1)'} = p_{n-1} + p_n$.

Obviously the probability that there is no cycle so far is the same in both scenarios.

Let T_k be the set of all individuals who have, directly or indirectly, chosen k. At this moment, in (i), T_0, T_{n-1}, and T_n are disjoint and are trees since Γ so far contains no cycles. In (ii) the same holds for T_0 and $T_{(n-1)'}$.

In scenario (i) a cycle in Γ would be created if $F(n-1) \in T_{n-1}$ or $F(n) \in T_n$ or $\{F(n-1) \in T_n$ and $F(n) \in T_{n-1}\}$. The probability for this is $S_{n-1} + S_n$, where S_{n-1} and S_n designate the sums of the weights of all vertices (including the root) in T_{n-1} and T_n, respectively.

In scenario (ii) the choice of $(n-1)'$ produces a cycle if and only if $F((n-1)') \in T_{(n-1)'}$. The probability of this is $S_{(n-1)'} = S_n + S_{n-1}$. $\blacksquare$

In Proof 4 we consider the stochastic process of the successive choices of the individuals.

Proof 4: Let the n individuals make their choices in succession beginning with individual n and ending with individual 1. Let C_k be the event that individual k completes the first cycle. We claim:

$$\Pr(C_k) = \frac{p_k}{\sum_{i=0}^{k} p_i}.$$

If this claim holds, the proposition follows by an easy calculation:

$$\Pr(A) = \Pr\left(\bigcap_{k=n}^{1} C_k^c\right) = \prod_{k=n}^{1}\left(1 - \frac{p_k}{\sum_{i=0}^{k} p_i}\right)$$

$$= \frac{\sum_{i=0}^{n-1} p_i}{\sum_{i=0}^{n} p_i} \cdot \frac{\sum_{i=0}^{n-2} p_i}{\sum_{i=0}^{n-1} p_i} \cdot \cdots \cdot \frac{p_0}{p_0 + p_1} = p_0.$$

We check the correctness of the claim. Suppose that the individuals $\{n, \ldots, k+1\}$ have not completed a cycle: this is the case if and only if each of them has chosen (perhaps indirectly) one of $\{0, \ldots, k\}$, otherwise a directed path along the arrows would start from this individual without reaching any element of $\{0, \ldots, k\}$, and would therefore be a cycle. For example, in Figure 1, n chooses $(n-2), (n-1)$ chooses k, and $(n-2)$ chooses k (indirectly).

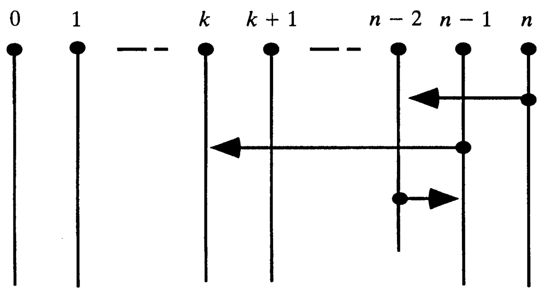


Figure 1

The event that individual k now completes the first cycle consists of two parts:

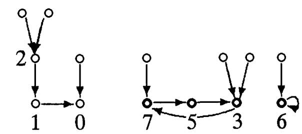
(i) $F(k) = k$, which happens with probability p_k.
(ii) $F(k) \in \{k+1, \ldots, n\}$, and $F(k)$ has chosen (perhaps indirectly) k; the first of these occurrences happens with probability $\sum_{i=k+1}^{n} p_i$. As to the second occurrence, consider the path starting at $F(k)$. The probability that it hits the set $\{0, \ldots, k\}$ in k is $p_k / \sum_{i=0}^{k} p_i$. Therefore:

$$\Pr(C_k) = p_k + \sum_{i=k+1}^{n} p_i \left(\frac{p_k}{\sum_{j=0}^{k} p_j}\right) = \frac{p_k}{\sum_{i=0}^{k} p_i}. \qquad \blacksquare$$

Proof 5 gives a bijection between cycle-free graphs of F and graphs where $F(1) = 0$. It also provides an economical algorithm for generating a random spanning tree with root 0.

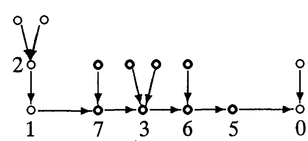
Proof 5: Consider the set B of all graphs Γ where $1 \to 0$ is an edge. B occurs with probability p_0. We want to prove $\Pr(A) = \Pr(B)$. We do this by constructing a bijective map $f: B \to A$ with the property $\Pr(f(b)) = \Pr(b)$ for all $b \in B$.

Consider a graph $b \in B$. For example:



The graph defines a partition into cycles of all the vertices that lie on a cycle. In our example: (753)(6).

Now we define $f(b) \in A$ by assigning the following graph to the corresponding permutation $(753)(6) = \binom{3567}{7365}$:



Both graphs occur with the same probability, namely $p_0^2 p_1 p_2^2 p_3^3 p_5 p_6^2 p_7^2$, because we do not change the number of edges that point toward each vertex.

The example describes how to construct a measure-preserving bijection (each permutation has a unique cycle–representation) for each selection of vertices that builds the cycles or the path between 1 and 0. Together the bijections define the desired f. ■

Proof 6 uses the fact that a cycle-free F codes a tree with root 0, and also gives a simple algorithm to generate a random spanning tree in a special case.

Proof 6: The connected graph without cycles generated by F (when A occurs) can also be considered as a labeled rooted tree. In this context it is more natural to think of $F(i)$ as the predecessor rather than the successor of i. Each vertex i has exactly one predecessor $F(i)$. The root is zero.

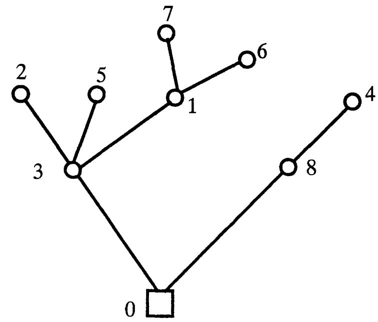
According to Knuth [**5**, p. 389] we can code each rooted $\{0,\dots,n\}$–labeled tree f by a vector $K(f) = (K_1(f),\dots,K_n(f)) \in \{0,\dots,n\}^n$, see Fig. 2.

The following proof is based on the equation

$$\Pr(A) = \sum_{x \in \{0,\dots,n\}^n} \Pr(F = K^{-1}(x))$$

and the simple structure of the summation range. Three properties of K are important for the computation of $\Pr(F = K^{-1}(x))$:

(i) $K_n(f)$ is the label of the root of f.
(ii) The number of children of a node equals the number of occurrences of its label in the multiset $[K_1(f),\dots,K_n(f)]$.
(iii) Every $x \in \{0,\dots,n\}^n$ codes a tree.



$f(7)$	$f(6)$	$f(5)$	$f(4)$	$f(8)$	$f(2)$	$f(1)$	$f(3)$
1	1	3	8	0	3	3	0
$K_1(f)$	$K_2(f)$	$K_3(f)$	$K_4(f)$	$K_5(f)$	$K_6(f)$	$K_7(f)$	$K_8(f)$

Figure 2. Knuth's coding of a rooted $\{0,\dots,n\}$-labeled tree f into $K(f) \in \{0,\dots,n\}^n$: We get the list of numbers $K_1(f),\dots,K_n(f)$ by iteratively removing the highest labeled leaf and noting the label of its predecessor. In each iteration step, a node is considered to be leaf, if all its children are already removed.

From (i) follows that $\Pr(F = K^{-1}(x)) = 0$ for $K_n(f) \neq 0$, because if F is a tree at all, it has 0 as root. For every rooted labeled tree f, (ii) implies equality of the multisets $[f(1),\dots,f(n)]$ and $[K_1(f),\dots,K_n(f)]$. Let T be the set of all $\{0,\dots,n\}$-labeled trees with root 0. We obtain:

$$\sum_{f\in T} \Pr(F=f) = \sum_{f\in T} p_{f(1)} \cdots p_{f(n)} = \sum_{f\in T} p_{K_1(f)} \cdots p_{K_n(f)}$$

$$= \sum_{x\in\{0,\dots,n\}^{n-1}\times\{0\}} p_{x_1} \cdots p_{x_n} = (p_0 + \cdots + p_n)^{n-1} \cdot p_0 = p_0.$$

Note that this also gives an algorithm for generating a random spanning tree: Just take $K^{-1}(X_1,\dots,X_n)$ with $X_1,\dots,X_n$ independent and identically distributed according to the weights of the random mapping F.

The tree can be recovered from the sequence by successively assigning to the $K_i(f)$'s their successors $K_i'(f)$. Remembering the rule of construction of the $K_i(f)$'s, one gets the arrangement illustrated in Figure 3.

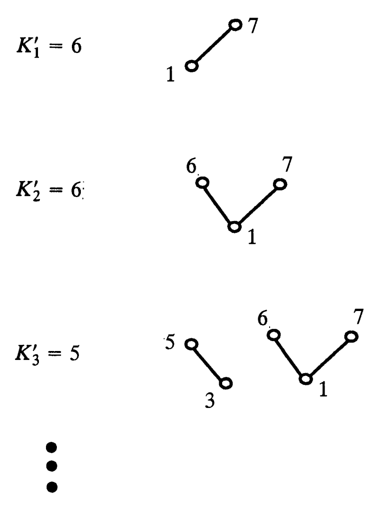


Figure 3

Formally

$$K_i' = \max\{\{1,\dots,n\} \setminus \{K_1',\dots,K_{i-1}',K_i,\dots,K_n\}\}, \quad 1 \le i \le n,$$

in our example: $K_1', K_2', K_3', \dots, K_8' = 7, 6, 5, 4, 8, 2, 1, 3.$ ■

ACKNOWLEDGMENTS. We thank Anton Wakolbinger and Brooks Ferebee for comments that have greatfully improved the presentation.

REFERENCES

1. B. Bollobas, *Graph Theory: An Introductory Course*, Springer Verlag (1979), Graduate Texts in Mathematics 63.
2. Y. D. Burtin, On a Simple Formula for Random Mappings and its Applications, *J. Appl. Prob.* **17** (1980) 403–414.
3. J. Jaworski, On a Random Mapping (T, p_j), *J. Appl. Prob.* **21** (1984), 186–191.
4. L. Kac, Probability of Indecomposability of a Random Mapping Function, *Ann. Math. Statist.* **26** (1955) 512–517.
5. D. E. Knuth, *The Art of Computer Programming*, Vol.1 (1968), Addison Wesley.
6. M. D. Kruskal, The Expected Number of Components Under a Random Mapping Function, *Amer. Math. Monthly* **61** (1954) 392–397.
7. N. Metropolis and S. Ulam, A Property of Randomness of an Arithmetical Function, *Amer. Math. Monthly* **60** (1953) 252–253.
8. J. G. Propp and D. B. Wilson, How to Get a Perfectly Random Sample from a Generic Markov Chain and Generate a Random Spanning Tree of a Directed Graph, preprint (1998), http://dimacs.rutgers.edu/~dbwilson/exact.html/
9. S. M. Ross, A Random Graph, *J. Appl. Prob.* 18 (1981), 309–315

SVETLANA ANOULOVA received her Ph.D. in 1979 at the University of Moscow. She is working on limit theorems for stochastic processes at the Institute for Control Sciences of the Russian Academy of Sciences.

JÜRGEN BENNIES is a Ph.D. student in mathematics in Frankfurt, Germany. The subjects of his research are random walks, trees, and mappings.
bennies@math.uni-frankfurt.de

JOHANNES LENHARD received his Ph.D. in mathematics from the University of Frankfurt, Germany, in 1998. In his thesis he examined a way of looking at a mathematical problem of population biology.
Lenhard@mathematik.uni-frankfurt.de

DIRK METZLER is a student of mathematics at the Johann Wolfgang Goethe-Universitat in Frankfurt, Germany. He is just about to finish his Ph.D. thesis on Poisson approximations for pattern configurations on related DNA-strands.
dmetzler@math.uni-frankfurt.de

YÜ SUNG received her Ph.D. from Frankfurt University. Her main research interests are in probability theory and finance.
y.sung@mathematik.uni-frankfurt.de

ANDREAS WEBER is an undergraduate student of mathematics at Frankfurt University (Germany).
And_Weber@compuserve.com

Johann Wofgang Goethe-Universität, Fachbereich Mathematik, AG Stochastik, Robert–Mayer–Str.10, D–60054 Frankfurt, Germany