A Mathematical Model

undefined terms: trinket, box, pack

axiom 1. The set of all trinkets, denoted by $\mathcal{T}$, has (finite) cardinality in excess of one million, and is partitioned into six equivalence classes (types) T_1, T_2, T_3, .., T_6 of equal cardinality.

defn 1. We refer to the equivalence class of a particular trinket as the type of the trinket.

axiom 2. The set of all boxes, denoted by $\mathcal{B}$, has (finite) cardinality the same as $\mathcal{T}$.

axiom 3. Pack, denoted by p, is a 1-1 onto function from $\mathcal{T}$ to $\mathcal{B}$. That is, for each trinket t there exists a unique box b such that, $p(t) = b$; conversely, for each box b there exists a unique trinket t such that $p^{-1}(b) = t$. Moreover, for each box b the unique choice of $p^{-1}(b)$ is made by random assignment from the subset of trinkets not previously mapped to a box by the packing function p. That is, each trinket is equally likely to be mapped to any particular box by the function p.

defn 2. Classify, denoted by c, is a function from B onto $\{T_1, T_2, .., T_6\}$ defined as follows: For each box b, $c(b) = T_j$ such that T_j is the type of $p^{-1}(b)$. So, classify is a function that associates each box with the type of the trinket assigned to the box by the pack function.

theorem 1. For any box b and any type T_j the probability that $c(b) = T_j$ is 1/6.

defn 3. For each subset $B = \{b_1, b_2, .., b_k\}$ of $\mathcal{B}$, of cardinality k, we define its collection to be $\{t_i : t_i = p^{-1}(b_i),\ i = 1,2,..k\}$ and its collection 6-vector as $(x_1, x_2, .., x_6)$ where x_i is the number of times T_i appears in the sequence $c(b_1)$, $c(b_2)$, ..., $c(b_k)$ for i = 1,2,..,6. So, a collection is the set of trinkets associated with a set of boxes, and a collection 6-vector identifies the number of times each type of trinket is represented in a collection.

defn 4. A collection with collection vector **v** is called complete if and only if the product of **v**'s coordinates is positive.

defn 5. Let $\mathcal{S}$ denote the set of all sequences of boxes. For any such sequence $s = \{b_i\}$ we define its minimally complete subsequence to be that subsequence consisting of the first k successive terms of s, $s_m = \{b_1, b_2, .., b_k\}$, for which $\{b_1, ..., b_k\}$ is complete but $\{b_1, ... b_{k-1}\}$ is not complete. (Note that, although s is a sequence with order important, s is also a subset of $\mathcal{B}$, so it has a collection and a collection vector.) A minimally complete sequence is identical to its minimally complete subsequence.

defn 6. Let goal, denoted by S_6, be a function from $\mathcal{S}$ into $\mathbb{Z}^+$ that associates each sequence of boxes s with the length k of its minimally complete subsequence. Note, for each sequence of boxes, S_6 represents the number of boxes up to the term when the sequence first contains a complete collection.

EXERCISE: Find the expected value of S_6.

This situation can be compared to random sampling, with replacement, from a population of size 6. See Feller, W., An Introduction to Probability Theory and Its Applications (Second Ed), New York: John Wiley & Sons, 1957. [pp 210-211]