Solution to HW 7, Problem II.11

Al-Naji, Nader

Balls switching chambers: the Ehrenfest model. Consider a system two chambers A and B (also clasically called urns). There N distinguishable balls, and, initially, chamber A contains all of them. At any instant, one ball is allowed to change from one chamber to the other. Let E_n^l be the number of possible evolutions that lead to chamber A containing l balls at instant n and $E^l(z)$ the corresponding EGF. Then:

$$E^l(z) = \binom{N}{l}(\cosh z)^l(\sinh z)^{N-l}, E^N(z) = (\cosh z)^N \equiv 2^{-N}(e^z + e^{-z})^N.$$

(Hint: the EGF E^N enumerates mappings where each preimage has an even cardinality.) In particular the probability that urn A is again full at time $2n$ is:

$$\frac{1}{2^N N^{2n}} \sum_{k=0}^{N} \binom{N}{k}(N-2k)^{2n}.$$

Solution:
First, we count the number of evolutions that lead to chamber A containing l balls and chamber B containing $N - l$ balls after n time steps. We can do this by counting strings of length n from an alphabet of cardinality N. To see this, first consider mapping each of our N labelled balls to a letter in our alphabet. Now, consider strings consisting of letters from this alphabet and let each occurrence of a particular letter correspond to the ball denoted by that letter "switching" from whatever urn it was in to the other. Then, if we have a string of length n, we can map this string *directly* to an evolution of our system and vice versa, since our system is such that one ball switches each time-step for n time-steps, thus giving us a bijection between N-alphabet strings of length n and an evolution of our system over n time-steps with N distinguishable balls. Further, it is clear that a particular ball ends up back in urn A if and only if the number of times its character occurs in the string is even, since this corresponds to an even number of switches, which puts the ball back into the urn it started in. And that a ball ends up in urn B if and only if the number of times its character occurs in the string is odd, by the same reasoning.

Thus, if we want to count the number of evolutions that give us l balls in urn A and $N - l$ balls in urn B after n time-steps, we simply need to count the number of N-alphabet strings of length n that have l of the letters occur an even number of times and $N - l$ of the letters occur an odd number of times.

We are given that $(\cosh z)^l$ gives the generating function for the number of strings from an l-character alphabet such that each letter occurs an even number of times. Further, from example $II.7$ and basic intuition from the series for $\sinh z$, we have $(\sinh z)^{N-l}$ as the generating function for the number of strings from an $(N-l)$-character alphabet such that each letter occurs an *odd* number of times. Finally, since there are $\binom{N}{l}$ possible ways to choose *which* subset of l balls remains in urn A, we have as our EGF:

$$E^l(z) = \binom{N}{l}(\cosh z)^l(\sinh z)^{N-l}.$$

We can now use this to compute the probability that urn A is full at time $2n$ by first computing the number of evolutions that lead to A being full at time $2n$, call this quantity $\Omega(2n)$, and dividing by all possible evolutions that can take place over $2n$ time steps, call this $T(2n)$. We begin by computing $\Omega(2n)$. Since we want there to be N balls in A, we set $l = N$ in the above equation; then, computing the $2n_{th}$ coefficient of the EGF and multiplying by $(2n)!$ gives us $\Omega(2n)$:

$$\begin{aligned}
E^l(z) &= \binom{N}{l}(\cosh z)^l(\sinh z)^{N-l} && \text{(From before)} \\
\Rightarrow E^N(z) &= \binom{N}{N}(\cosh z)^N(\sinh z)^{N-N} && \text{(Plugging in } l = N\text{)} \\
\Rightarrow E^N(z) &= (\cosh z)^N && \text{(Simplifying)} \\
\Rightarrow \Omega(2n) &= [z^{2n}](\cosh z)^N && \text{(By above reasoning)} \\
&= [z^{2n}]2^{-N}(e^z + e^{-z})^N && \text{(Using identity given)} \\
&= [z^{2n}]2^{-N}\sum_{i=0}^{N}\binom{N}{i}e^{-zi}e^{z(N-i)} && \text{(Binomial expansion)} \\
&= [z^{2n}]2^{-N}\sum_{i=0}^{N}\binom{N}{i}e^{z(N-2i)} && \text{(Consolidating exponents)} \\
&= [z^{2n}]2^{-N}\sum_{i=0}^{N}\binom{N}{i}\sum_{j\geq 0}(N-2i)^j z^j && \text{(Series for } e^x\text{)} \\
&= [z^{2n}]\sum_{j\geq 0}\left(2^{-N}\sum_{i=0}^{N}\binom{N}{i}(N-2i)^j\right) z^j && \text{(Swapping summations)} \\
&= 2^{-N}\sum_{i=0}^{N}\binom{N}{i}(N-2i)^{2n} && \text{(Extracting coefficient)}
\end{aligned}$$

Now, we want to divide this quantity by the total number of evolutions that can occur after $2n$ time steps. But this is equal to the total number of strings of length $2n$ using an N-character alphabet, which is N^{2n}. Thus, as our final answer we have that the probability that all of the balls are in urn A after $2n$ time steps is:

$$\begin{aligned} P &= \frac{\Omega(2n)}{T(2n)} \\ &= \frac{1}{2^N N^{2n}} \sum_{i=0}^{N} \binom{N}{i} (N-2i)^{2n} \end{aligned}$$

Solution to HW 7, Problem II.31

Al-Naji, Nader

Interpret $\tan\frac{z}{1-z}$**,** $\tan\tan z$**, and** $\tan(e^z-1)$ **as EGF's of combinatorial classes.**
Solution:
We compute the symbolic representation for each generating function below:

$$\begin{aligned}
T(z) &= \tan\frac{z}{1-z} \\
\frac{dT(z)}{dz} \equiv T_z(z) &= \frac{\sec^2\left(\frac{z}{1-z}\right)}{(z-1)^2} \\
&= \frac{1+\tan^2\left(\frac{z}{1-z}\right)}{(z-1)^2} \\
\Rightarrow T(z) &= \int_0^z \frac{1}{(u-1)^2}du + \int_0^z \frac{T(u)^2}{(u-1)^2}du \\
&= \frac{z}{(1-z)} + \int_0^z \frac{T(u)^2}{(1-u)^2}du \\
T &= Z \star SEQ(Z) + Z^\square \star T \star T \star SEQ(Z) \star SEQ(Z)
\end{aligned}$$

$$\begin{aligned}
T(z) &= \tan\tan z \\
\text{Let } J(z) &= tan(z)\text{From p. 124} \\
T'(z) &= \left(\tan^2(z)+1\right)\left(\tan^2(\tan(z))+1\right) \\
&= \left(J(z)^2+1\right)\left(T(z)^2+1\right) \\
&= J(z)^2T(z)^2 + J(z)^2 + T(z)^2 + 1 \\
\Rightarrow T(z) &= z + \int_0^z J(z)^2T(z)^2 + J(z)^2 + T(z)^2 \\
\Rightarrow T &= Z + Z^\square \star (J \star J \star T \star T + J \star J + T \star T)
\end{aligned}$$

$$\begin{aligned}
T(z) &= \tan(e^z-1) \\
T'(z) &= e^z\left(\tan^2(1-e^z)+1\right) \\
&= e^zT(z)^2 + e^z \\
\Rightarrow T(z) &= e^z - 1 + \int_0^z e^zT(z)^2 \\
\Rightarrow T &= SET(Z)_{>0} + Z^\square \star SET(Z) \star T \star T
\end{aligned}$$

From these symbolic representations, it is trivial to state interpretations in plain english. We avoid doing so here to save time and effort.

Solution to HW 7, Problem III.17

Al-Naji, Nader

Leaves and node-degree profile in Cayley trees. For Cayley trees, the bivariate EGF with u marking the number of leaves is the solution to $T(z,u) = uz + z(e^{T(u,z)} - 1)$. (By Lagrange inversion, the distribution is expressible in terms of Stirling partition numbers.) The mean number of leaves in a random Cayley tree is asymptotic to n/e. More generally, the mean number of nodes of outdegree k in a random Cayley tree of size n is asymptotic to $\frac{n}{e \cdot k!}$. Degrees are thus approximately described by a Poisson law of rate 1.

Solution:
We show that the mean number of nodes of out-degree k in a random Cayley tree of size n is asymptotic to $\frac{n}{e \cdot k!}$ by computing the cumulative number of nodes of out-degree k across all Cayley trees of size n, call this cumulative cost function $\Omega(n)$, and dividing by the number of Cayley trees of size n, calling this total number $T(n)$.

First, we use the symbolic method to get the bivariate generating function for the number of Cayley trees of size n, marking nodes with out-degree k. First, we note that a Cayley tree is simply a node attached to a set of Cayley trees, which yields:

$$\begin{aligned}
T &= Z \star \left(\{\epsilon\} + T + SET_2(T^2) + ...\right) \\
\Rightarrow T &= Z \star (SET(T)) && \text{(Def of } SET(T)\text{)} \\
\Rightarrow T(z) &= z \left(e^{T(z)}\right) && \text{(Transfer theorem)} \\
\Rightarrow T(z) &= z \left(\sum_{i \geq 0} \frac{T^i}{i!}\right) && \text{(Series for } e^x\text{)}
\end{aligned}$$

Now, we want to construct a bivariate EGF that associates a cost of 1 to each node with out-degree k. But this situation occurs exactly when a node connects to k Cayley trees, which is represented symbolically as $Z \star SET_i(T)$. Thus, marking every occurrence of such nodes with u^1, to associate a cost of 1 to each, gives us our desired bivariate EGF:

$$\begin{aligned}
T(z) &= ze^{T(z)} && \text{(From before)} \\
\Rightarrow T(u,z) &= z \left(\frac{T(u,z)^0}{1!} + \frac{T(u,z)^2}{2!} + ... + \frac{uT(u,z)^k}{k!} + ...\right) && \text{(Mark nodes with out-degree } k\text{)} \\
\Rightarrow T(u,z) &= z \left(\left(\sum_{i \geq 0} \frac{T(u,z)^i}{i!}\right) - \frac{T(u,z)^k}{k!} + \frac{uT(u,z)^k}{k!}\right) && \text{(Consolidate terms)} \\
\Rightarrow T(u,z) &= z \left(e^{T(u,z)} - \frac{T(u,z)^k}{k!} + \frac{uT(u,z)^k}{k!}\right) && \text{(Def of } e^x\text{)}
\end{aligned}$$

We can now use Lagrange inversion to get the coefficient on z^n in the equation above. In general, the Lagrange inversion theorem states that if we have $z = f(G(z))$, where $G(z) = \sum_{n \geq 0} g_n z^n$ is some generating function, we can extract coefficients of any z^n term in $G(z)$ using: $G(n) \equiv g_n = \frac{1}{n}[w^{n-1}]\left((\frac{w}{f(w)})^n\right)$. For our specific purposes, the Lagrange inversion theorem tells us that if we have $z = f(T(u, z))$, then we get: $T(n, u) = \frac{1}{n}[w^{n-1}]\left((\frac{w}{f(w)})^n\right)$, where $T(n, u)$ gives us the horizontal EGF associated with z^n, which we can use to calculate cumulative cost. Once we have $T(n, u)$, differentiating with respect to u and then setting $u = 1$ and multiplying by $n!$ (because we have an EGF) yields the desired cumulative cost function $\Omega(n)$, defined to be the total number of nodes with degree k across all trees of size

n. Having set up out computation, we now compute $\Omega(n)$ below:

$$\begin{aligned}
T(u,z) &= z\left(e^{T(u,z)} - \frac{T(u,z)^k}{k!} + \frac{uT(u,z)^k}{k!}\right) && \text{(From before)}\\
\Rightarrow z &= \frac{T(u,z)}{e^{T(u,z)} - \frac{T(u,z)^k}{k!} + \frac{uT(u,z)^k}{k!}} && \text{(Solving for } z\text{)}\\
\text{Let } f(\omega) &= \frac{\omega}{e^{\omega} - \frac{\omega^k}{k!} + \frac{u\omega^k}{k!}} && \text{(Setting up Lagrange inversion)}\\
\Rightarrow T(n,u) &= \frac{1}{n}[w^{n-1}]\left((e^{\omega} - \frac{\omega^k}{k!} + \frac{u\omega^k}{k!})^n\right) && \text{(By Lagrange inversion)}\\
\Rightarrow \frac{\delta T(n,u)}{\delta u} &= \frac{1}{n}[w^{n-1}]\left(n(e^{\omega} - \frac{\omega^k}{k!} + \frac{u\omega^k}{k!})^{n-1}\frac{\omega^k}{k!}\right) && \text{(Taking derivative)}\\
\Omega(n) &= n!\frac{\delta T(n,u)}{\delta u}(n,1) && \text{(As explained above)}\\
&= \frac{n!}{n}[w^{n-1}]\left(ne^{\omega(n-1)}\frac{\omega^k}{k!}\right) && \text{(Setting } u=1\text{)}\\
&= \frac{n!}{n}[w^{n-1}]\left(n\frac{1}{k!}\sum_{i\geq 0}\frac{(n-1)^i}{i!}\omega^{i+k}\right) && \text{(Expanding } e^x\text{)}\\
&= \frac{n!}{n}[w^{n-1}]\left(n\frac{1}{k!}\sum_{i\geq k}\frac{(n-1)^{i-k}}{(i-k)!}\omega^{i}\right) && \text{(Reindexing)}\\
&= \frac{n!}{k!}\frac{(n-1)^{n-1-k}}{(n-1-k)!} && \text{(Extracting coefficient)}\\
&= \frac{n(n-1)^{n-1-k}}{k!}\frac{(n-1)!}{(n-1-k)!} && \text{(Isolate factorials)}\\
&= \frac{n(n-1)^{n-1-k}}{k!}(n-1)\cdot \ldots \cdot((n-1)-(k-1)) && \text{(Manipulating factorials)}\\
&= \frac{n(n-1)^{n-1-k}}{k!}(n-1)^k(1)(1-\frac{1}{n-1})\ldots(1-\frac{k-1}{n-1}) && \text{(Pulling out } (n-1)^k\text{)}\\
&\sim \frac{n(n-1)^{n-1-k}}{k!}(n-1)^k && \text{(Eliminating exp small terms)}\\
&\sim \frac{n(n-1)^{n-1}}{k!} && \text{(Consolidating)}\\
&= \frac{n(n(1-\frac{1}{n}))^{n-1}}{k!} && \text{(Pulling out } n\text{)}\\
\Rightarrow \Omega(n) &\sim \frac{n\cdot n^{n-1}}{e\cdot k!} && \text{(Def of } \tfrac{1}{e}\text{)}
\end{aligned}$$

Now, all we need to do is find the total number of Cayley trees of size n and divide. We do this using Lagrange inversion below:

$$\begin{aligned}
T &= Z \times (SET(T)) \\
T(z) &= ze^{T(z)} \\
\Rightarrow z &= \frac{T(z)}{e^{T(z)}} \\
\text{Let } f(\omega) &= \frac{\omega}{e^{\omega}} \\
\Rightarrow T(n) &= \frac{n!}{n}[\omega^{n-1}]e^{wn} && \text{(Lagrange inversion)} \\
&= \frac{n!}{n}[\omega^{n-1}]\sum_{i\geq 0}\frac{n^i}{i!}w^i \\
&= n^{n-1}
\end{aligned}$$

Thus, as our final answer, we have that the probability that the mean number of nodes of outdegree k is:

$$\begin{aligned}
E_k &= \frac{\Omega(n)}{T(n)} \\
&\sim \frac{n \cdot n^{n-1}}{n^{n-1}e \cdot k!} \\
&\sim \frac{n}{e \cdot k!}
\end{aligned}$$

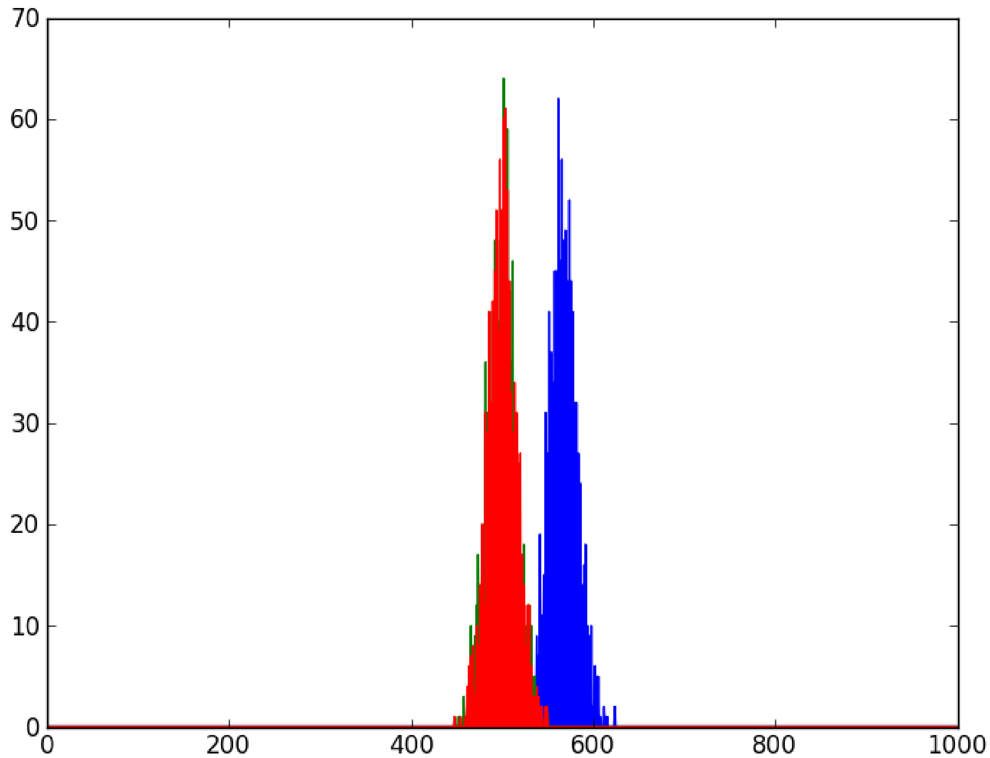
Obviously, plugging in $k = 1$ tells us the mean number of leaves is $\frac{n}{e}$.

Solution to HW 7, Problem II.1 (Programming)

Al-Naji, Nader

Write a program to simulate the Ehrenfest model and use it to plot the distribution of the number of balls in urn A after 10^3, 10^4, and 10^5 steps when starting with 10^3 balls in urn A and none in urn B.

Solution:
The following plots the distribution of balls in urn A after $10^3, 10^4$, and 10^5 steps. As can be seen, the distribution for 10^3 steps is skewed to the right, while the distributions for the higher steps converge to the center.



Solution to HW 7, Problem III.1 (Programming)

Al-Naji, Nader

Write a program that generates 1000 random permutations of size N for $N = 10^3, 10^4$, ... (going as far as you can) and plots the distribution of the number of cycles, validating the mean is concentrated at H_n

Solution:
The following is a plot of the distribution of the number of cycles for $N = 10^3, 10^4$, and 10^5 along with the respective H_n. From the plot, it is clear that the mean is concentrated along H_n.