Analytic Combinatorics

Ricardo Bittencourt

April 13, 2013

1 Week 5

1.1 Exercise 8.3

How long a string of random bits should be taken to be 50% sure that there are at least 32 consecutive 0s?

Solution

The number of bitstrings with no runs of P consecutive zeros is given by:

$$B_P(z) = \frac{1 - z^P}{1 - 2z + z^{P+1}}$$

We can approximate this using the transfer theorem for rational functions:

$$[z^n]\frac{f(z)}{g(z)} \sim C\,\beta^n\,n^{\nu-1}$$

Where $1/\beta$ is the largest root of $g(z)$, ν is its multiplicity, and C is given by:

$$C = \nu\frac{(-\beta)^\nu f(1/\beta)}{g^{(\nu)}(1/\beta)}$$

For $P = 32$, the largest root leads to $\beta = 1.999999999767169$ (found numerically), and $C = 1.0000000034924597$. The probability that a random bitstring of size n has 32 consecutive zeros is:

$$\mu = \frac{C\beta^n}{2^n}$$

The minimum value of n to reach a probability $\mu = 0.5$ therefore is:

$$n = \left\lceil \frac{\log \mu / C}{\log \beta / 2} \right\rceil$$

$$n = 5954088952$$

1.2 Exercise 8.14

Suppose that a monkey types randomly at a 32-key keyboard. What is the expected number of characters typed before the monkey hits upon the phrase THE QUICK BROWN FOX JUMPED OVER THE LAZY DOG?

Solution

First of all, we should notice that the string has length $P = 44$, and its autocorrelation polynomial is just $c_P(z) = 1$. The number of strings of size n, on a 32-key keyboard, that does not contain out target string, is given by:

$$S_P(z) = \frac{c_P(z)}{z^P + (1 - Mz)c_P(z)}$$
$$S_P(z) = \frac{1}{z^{44} - 32z + 1}$$

The expected wait time of our string is given directly by $S_P(1/32)$:

$$S_P(1/32) = 1.685 \times 10^{66}$$

1.3 Exercise 8.57

Solve the recurrence for p_N given in the proof of Theorem 8.9, to within the oscillating term: $p_N = \frac{1}{2^N}\sum_k \binom{N}{k} p_k$ for $N > 1$ with $p_0 = 0$ and $p_1 = 1$.

Solution

First we add the initial conditions, and take the exponential generating function:

$$p_n = \frac{1}{2^n}\sum_k \binom{n}{k} p_k + \frac{1}{2}[n == 1]$$
$$\sum_n p_n \frac{z^n}{n!} = \sum_n \frac{1}{2^n}\sum_k \binom{n}{k} p_k \frac{z^n}{n!} + \frac{1}{2}z$$
$$P(z) = \sum_n \sum_k \binom{n}{k} p_k \frac{(z/2)^n}{n!} + \frac{z}{2}$$

The first term is the binomial sum on $z/2$:

$$P(z) = \sum_n \sum_k \binom{n}{k} p_k \frac{(z/2)^n}{n!} + \frac{z}{2}$$
$$P(z) = e^{z/2} P\left(\frac{z}{2}\right) + \frac{z}{2}$$

Let's expand some terms:

$$P(z) = e^{z/2} P\left(\frac{z}{2}\right) + \frac{1}{2} z$$
$$P(z) = e^{3z/4} P\left(\frac{z}{4}\right) + e^{z/2} \frac{1}{4} z + \frac{1}{2} z$$
$$P(z) = e^{7z/8} P\left(\frac{z}{8}\right) + e^{3z/4} \frac{1}{8} z + e^{z/2} \frac{1}{4} z + \frac{1}{2} z$$

This suggests that $P(z)$ is:

$$P(z) = \sum_{n \geq 1} \exp\left(\frac{2^{n-1} - 1}{2^{n-1}} z\right) \frac{z}{2^n}$$

We can prove this is true by substituting:

$$P(z) = \sum_{n \geq 1} \exp\left(\frac{2^{n-1} - 1}{2^{n-1}} z\right) \frac{z}{2^n}$$
$$P(z) = e^0 \frac{z}{2} + \sum_{n \geq 2} \exp\left(\frac{2^{n-1} - 1}{2^{n-1}} z\right) \frac{z}{2^n}$$
$$P(z) = \frac{z}{2} + \sum_{k+1 \geq 2} \exp\left(\frac{2^k - 1}{2^k} z\right) \frac{z}{2^{k+1}}$$
$$P(z) = \frac{z}{2} + \sum_{k \geq 1} \exp\left(\frac{2^k - 1}{2^k} z + \frac{z}{2} - \frac{z}{2}\right) \frac{1}{2^k} \frac{z}{2}$$
$$P(z) = \frac{z}{2} + \sum_{k \geq 1} \exp\left(\frac{2^{k-1} - 1}{2^{k-1}} \frac{z}{2}\right) \exp\left(\frac{z}{2}\right) \frac{1}{2^k} \frac{z}{2}$$
$$P(z) = \frac{z}{2} + \exp\left(\frac{z}{2}\right) \sum_{k \geq 1} \exp\left(\frac{2^{k-1} - 1}{2^{k-1}} \frac{z}{2}\right) \frac{1}{2^k} \frac{z}{2}$$
$$P(z) = \frac{z}{2} + e^{z/2} P\left(\frac{z}{2}\right)$$

Back to the EGF:

$$
\begin{aligned}
P(z) &= \sum_{n\geq 1} \exp\left(\frac{2^{n-1}-1}{2^{n-1}} z\right)\frac{z}{2^n} \\
P(z) &= \sum_{n\geq 1} \exp\left(z - \frac{z}{2^{n-1}}\right)\frac{z}{2^n} \\
P(z) &= \sum_{n\geq 1} \exp\left(z\right)\exp\left(-\frac{z}{2^{n-1}}\right)\frac{z}{2^n} \\
P(z) &= ze^z \sum_{n\geq 1} \exp\left(-\frac{z}{2^{n-1}}\right)\frac{1}{2^n} \\
P(z) &= ze^z \sum_{n\geq 1}\sum_{k\geq 0} \frac{1}{k!}\frac{(-1)^k z^k}{2^{k(n-1)}}\frac{1}{2^n} \\
P(z) &= ze^z \sum_{k\geq 0} \frac{(-1)^k z^k}{k!}\sum_{n\geq 1}\frac{1}{2^{k(n-1)}}\frac{1}{2^n} \\
P(z) &= ze^z \sum_{k\geq 0} \frac{(-1)^k z^k}{k!}\frac{2^k}{2^{k+1}-1}
\end{aligned}
$$

This is a binomial convolution of two EGFs. The first one is ze^z which is the EGF of the sequence $a_n = n$, the second one is the EGF of the sequence $b_k = \frac{(-2)^k}{2^{k+1}-1}$:

$$
p_n = \sum_{0\leq k\leq n} \binom{n}{k}(n-k)\frac{(-1)^k 2^k}{2^{k+1}-1}
$$

We can use the absorption $(n-k)\binom{n}{k} = n\binom{n-1}{k}$:

$$
\begin{aligned}
p_n &= n\sum_{0\leq k\leq n} \binom{n-1}{k}\frac{(-1)^k 2^k}{2^{k+1}-1} \\
p_n &= n\sum_{0\leq k\leq n} (-1)^k\binom{n-1}{k}\frac{1}{2}\frac{2^{k+1}}{2^{k+1}-1} \\
p_n &= \frac{n}{2}\sum_{0\leq k\leq n} (-1)^k\binom{n-1}{k}\frac{1}{2}\frac{2^{k+1}+1-1}{2^{k+1}-1} \\
p_n &= \frac{n}{2}\sum_{0\leq k\leq n} (-1)^k\binom{n-1}{k}\left(1+\frac{1}{2^{k+1}-1}\right) \\
p_n &= \frac{n}{2}\sum_{0\leq k\leq n} (-1)^k\binom{n-1}{k} + \frac{n}{2}\sum_{0\leq k\leq n} (-1)^k\binom{n-1}{k}\frac{1}{2^{k+1}-1}
\end{aligned}
$$

By the binomial theorem, the first sum is equal to $(1-1)^{n-1}$, which is just 0 for $n > 2$:

$$p_n = \frac{n}{2} \sum_{0 \le k \le n} (-1)^k \binom{n-1}{k} \frac{1}{2^{k+1}-1}$$
$$p_n = \frac{n}{2} \sum_{0 \le k \le n} (-1)^k \binom{n-1}{k} \frac{1}{2^{k+1}} \frac{1}{1-\frac{1}{2^{k+1}}}$$
$$p_n = \frac{n}{2} \sum_{0 \le k \le n} (-1)^k \binom{n-1}{k} \frac{1}{2^{k+1}} \sum_{j \ge 0} \frac{1}{2^{j(k+1)}}$$
$$p_n = \frac{n}{2} \sum_{j \ge 0} \sum_{0 \le k < n} (-1)^k \binom{n-1}{k} \frac{1}{2^{k(j+1)+(j+1)}}$$
$$p_n = \frac{n}{2} \sum_{j \ge 0} \frac{1}{2^{j+1}} \sum_{0 \le k < n} (-1)^k \binom{n-1}{k} \frac{1}{2^{k(j+1)}}$$

We can use the binomial theorem again:

$$p_n = \frac{n}{2} \sum_{j \ge 0} \frac{1}{2^{j+1}} \sum_{0 \le k < n} (-1)^k \binom{n-1}{k} \frac{1}{2^{k(j+1)}}$$
$$p_n = \frac{n}{2} \sum_{j \ge 0} \frac{1}{2^{j+1}} \left(1 - \frac{1}{2^{j+1}}\right)^{n-1}$$
$$p_n = \frac{n}{2} \sum_{j \ge 1} \frac{1}{2^{j}} \left(1 - \frac{1}{2^{j}}\right)^{n-1}$$

Let's approximate it by an integral:

$$p_n = \frac{n}{2}\left(\sum_{j\geq 1}\frac{1}{2^j}\left(1-\frac{1}{2^j}\right)^{n-1} + \int_1^\infty \frac{1}{2^x}\left(1-\frac{1}{2^x}\right)^{n-1}dx - \int_1^\infty \frac{1}{2^x}\left(1-\frac{1}{2^x}\right)^{n-1}dx\right)$$

$$p_n = \frac{n}{2}\left(\frac{(1-2^{-j})^n}{n\log 2}\Big|_1^\infty + \sum_{j\geq 1}\frac{1}{2^j}\left(1-\frac{1}{2^j}\right)^{n-1} - \sum_{j\geq 1}\int_j^{j+1}\frac{1}{2^x}\left(1-\frac{1}{2^x}\right)^{n-1}dx\right)$$

$$p_n = \frac{1-2^{-n}}{2\log 2} + \frac{n}{2}\left(\sum_{j\geq 1}2^{-j}\left(1-2^{-j}\right)^{n-1} - \sum_{j\geq 1}\frac{(1-2^{-j-1})^n-(1-2^{-j})^n}{n\log 2}\right)$$

$$p_n = \frac{1-2^{-n}}{2\log 2} + \frac{1}{2\log 2}\left(\sum_{j\geq 1}2^{-j}n\log 2\left(1-2^{-j}\right)^{n-1} - (1-2^{-j-1})^n + (1-2^{-j})^n\right)$$

Now we'll use the approximation $(1-2^{-j})^{n-1} \sim \exp(-n2^{-j})$.

$$p_n \sim \frac{1-2^{-n}}{2\log 2} + \frac{1}{2\log 2}\left(\sum_{j\geq 1}2^{-j}n\log 2\exp(-n2^{-j}) - \exp(-(n+1)2^{-j-1}) + \exp(-(n+1)2^{-j})\right)$$

Let's define $k = j + \lfloor \lg n \rfloor = j + \lg n - \{\lg n\}$. Notice that $n2^{-\lg n} = 1$ and also:

$$\exp(-(n+1)2^{-j}) = \exp(-(n+1)2^{-k-\lg n+\{\lg n\}})$$
$$\exp(-(n+1)2^{-j}) = \exp(-(n+1)2^{-\lg n}\,2^{-k+\{\lg n\}})$$
$$\exp(-(n+1)2^{-j}) = \exp\left(-\frac{n+1}{n}2^{-k+\{\lg n\}}\right)$$
$$\exp(-(n+1)2^{-j}) = \exp\left(-\left(1+\frac{1}{n}\right)2^{-k+\{\lg n\}}\right)$$
$$\exp(-(n+1)2^{-j}) \sim \exp(-2^{-k+\{\lg n\}})$$

Substituting:

$$p_n \sim \frac{1-2^{-n}}{2\log 2} + \frac{1}{2\log 2}\left(\sum_{j\geq 1}(2^{-j}n\log 2+1)\exp(-n2^{-j}) - \exp(-n2^{-j-1})\right)$$

$$p_n \sim \frac{1-2^{-n}}{2\log 2} + \frac{1}{2\log 2}\left(\sum_{k\geq 1-\lfloor \lg n\rfloor}(2^{-k+\{\lg n\}}\log 2+1)\exp(-2^{-k+\{\lg n\}}) - \exp(-2^{-k-1+\{\lg n\}})\right)$$

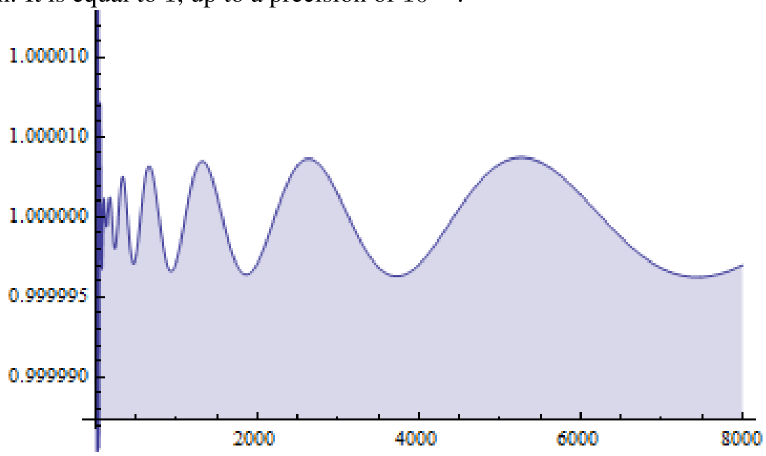
The terms for $k < 1 - \lfloor \lg n \rfloor$ are exponentially small, so we can extend the range of summation for all k:

$$p_n \sim \frac{1-2^{-n}}{2\log 2} + \frac{1}{2\log 2}\left(\sum_k (2^{-k+\{\lg n\}}\log 2 + 1)\exp(-2^{-k+\{\lg n\}}) - \exp(-2^{-k-1+\{\lg n\}})\right)$$

$$p_n \sim \frac{1-2^{-n}}{2\log 2} + f(\{\lg n\})$$

The whole summation is now a function of the fractional part of $\lg n$, so we have $f(2x) = f(x)$ and this function is therefore oscillating.

Out of curiosity, here's the graph of the original function divided by its approximation. It is equal to 1, up to a precision of 10^{-5}.



1.4 Exercise 9.5

For $M = 365$, how many people are needed to be 99% sure that two have the same birthday?

Solution

The number of persons is given by:

$$N \sim \sqrt{-2M\ln p} = \sqrt{-2 \times 365 \ln 0.01} = 57.98$$

The desired percentage can be achieved with 58 persons.

1.5 Exercise 9.37

Find $[z^n]e^{\alpha C(z)}$ where $C(z)$ is the Cayley function.

Solution

The Cayley function is defined by $C(z) = ze^{C(z)}$. We'll solve it by using this version of the Lagrange Inversion, valid when $z = f(A(z))$, $f(0) = 0$ and $f'(0) \neq 0$:

$$[z^n]g(A(z)) = \frac{1}{n}[u^{n-1}]g'(u)\left(\frac{u}{f(u)}\right)^n$$

We have:

$$\begin{aligned}
z &= \frac{C(z)}{e^{C(z)}} \\
f(u) &= \frac{u}{e^u} \\
f(0) &= \frac{0}{e^0} = 0 \\
f'(u) &= (1-u)e^u \\
f'(0) &= (1-0) \times 0 = 1 \neq 0 \\
g(u) &= e^{\alpha u} \\
g'(u) &= \alpha e^{\alpha u}
\end{aligned}$$

We need to multiply the result of the formula by $n!$, since it's an EGF.

$$\begin{aligned}
n![z^n]e^{\alpha C(z)} &= n!\frac{1}{n}[u^{n-1}]g'(u)\left(\frac{u}{f(u)}\right)^n \\
n![z^n]e^{\alpha C(z)} &= (n-1)![u^{n-1}]\alpha e^{\alpha u}\left(\frac{ue^u}{u}\right)^n \\
n![z^n]e^{\alpha C(z)} &= (n-1)![u^{n-1}]\alpha e^{(\alpha+n)u} \\
n![z^n]e^{\alpha C(z)} &= (n-1)![u^{n-1}]\sum_{k\geq 0}\alpha\frac{1}{k!}(\alpha+n)^k u^k \\
n![z^n]e^{\alpha C(z)} &= (n-1)!\alpha\frac{1}{(n-1)!}(\alpha+n)^{n-1} \\
n![z^n]e^{\alpha C(z)} &= \alpha(\alpha+n)^{n-1}
\end{aligned}$$

1.6 Exercise 9.38

(Abel's binomial theorem) Use the result of the previous exercise and the identity $e^{(\alpha+\beta)C(z)} = e^{\alpha C(z)}e^{\beta C(z)}$ to prove that

$$(\alpha+\beta)(n+\alpha+\beta)^{n-1} = \alpha\beta\sum_k \binom{n}{k}(k+\alpha)^{k-1}(n-k+\beta)^{n-k-1}.$$

Solution

We start by extracting coefficients from the three EGFs:

$$
\begin{aligned}
[z^n]e^{\alpha C(z)} &= \alpha(\alpha+n)^{n-1} \\
[z^n]e^{\beta C(z)} &= \beta(\beta+n)^{n-1} \\
[z^n]e^{(\alpha+\beta)C(z)} &= (\alpha+\beta)(\alpha+\beta+n)^{n-1}
\end{aligned}
$$

The product of EGFs if the EGF of the binomial convolution of their coefficients:

$$
\begin{aligned}
(\alpha+\beta)(\alpha+\beta+n)^{n-1} &= \sum_k \binom{n}{k}\alpha(\alpha+k)^{k-1}\beta(\beta+n-k)^{n-k-1} \\
(\alpha+\beta)(\alpha+\beta+n)^{n-1} &= \alpha\beta\sum_k \binom{n}{k}(\alpha+k)^{k-1}(\beta+n-k)^{n-k-1}
\end{aligned}
$$

1.7 Exercise 9.99

Show that the probability that a random mapping of size N has no singleton cycles is $\sim 1/e$, the same as for permutations (!).

Solution

The symbolic construction for mappings with no singletons is:

$$
\begin{aligned}
M(z) &= SET(CYC(C(z)) - CYC_1(C(z))) \\
M(z) &= \exp\left(\log\left(\frac{1}{1-C(z)}\right) - C(z)\right) \\
M(z) &= \exp\left(\log\left(\frac{1}{1-C(z)}\right)\right)\exp(-C(z)) \\
M(z) &= \frac{\exp(-C(z))}{1-C(z)}
\end{aligned}
$$

We'll use the Lagrange Inversion with the following parameters:

$$
\begin{aligned}
f(u) &= \frac{u}{e^u} \\
g(u) &= \frac{e^{-u}}{1-u} \\
g'(u) &= e^{-u}\frac{u}{(1-u)^2}
\end{aligned}
$$

Applying the formula:

$$
\begin{aligned}
n![z^n]M(z) &= \frac{n!}{n}[u^{n-1}]g'(u)\left(\frac{u}{f(u)}\right)^n \\
n![z^n]M(z) &= (n-1)![u^{n-1}]e^{-u}\frac{u}{(1-u)^2}\left(e^u\right)^n \\
n![z^n]M(z) &= (n-1)![u^{n-1}]e^{u(n-1)}\frac{u}{(1-u)^2} \\
n![z^n]M(z) &= (n-1)![u^{n-1}]\sum_{q\geq 0}\sum_{0\geq k\geq q}\frac{1}{k!}(n-1)^k(q-k)u^q \\
n![z^n]M(z) &= (n-1)!\sum_{0\geq k\geq n-1}\frac{1}{k!}(n-1)^k(n-1-k)
\end{aligned}
$$

This sum is telescoping. Let's call it $T(n)$:

$$
\begin{aligned}
T(n) &= \sum_{0\geq k\geq n-1}\frac{1}{k!}(n-1)^k(n-1-k) \\
T(n) &= \sum_{0\geq k\geq n-1}\frac{n-1}{k!}(n-1)^k - \sum_{0\geq k\geq n-1}\frac{k}{k!}(n-1)^k \\
T(n) &= \frac{(n-1)(n-1)^{n-1}}{(n-1)!} + \sum_{0\geq k\geq n-2}\frac{n-1}{k!}(n-1)^k - \sum_{1\geq k\geq n-1}\frac{k}{k!}(n-1)^k \\
T(n) &= \frac{(n-1)^n}{(n-1)!} + \sum_{0\geq k\geq n-2}\frac{n-1}{k!}(n-1)^k - \sum_{1\geq k\geq n-1}\frac{(n-1)}{(k-1)!}(n-1)^{k-1} \\
T(n) &= \frac{(n-1)^n}{(n-1)!} + \sum_{0\geq k\geq n-2}\frac{n-1}{k!}(n-1)^k - \sum_{0\geq k\geq n-2}\frac{n-1}{k!}(n-1)^k \\
T(n) &= \frac{(n-1)^n}{(n-1)!}
\end{aligned}
$$

Substituting back:

$$
\begin{aligned}
n![z^n]M(z) &= (n-1)!T(n) \\
n![z^n]M(z) &= (n-1)!\frac{(n-1)^n}{(n-1)!} \\
n![z^n]M(z) &= (n-1)^n
\end{aligned}
$$

The total number of mappings is n^n, so the probability of a random mapping having no singletons is:

$$\mu = \frac{(n-1)^n}{n^n}$$
$$\mu = \left(1 - \frac{1}{n}\right)^n$$

In the limit for large n:

$$\lim_{n\to\infty} \mu = \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n$$
$$\lim_{n\to\infty} \mu = e^{-1}$$

So the probability is $1/e$ as claimed.

2 Week 4

2.1 Exercise 6.6

What proportion of the forests with N nodes have no trees consisting of a single node? For $N = 1, 2, 3$, and 4, the answers are $0, 1/2, 2/5$, and $3/7$, respectively.

Solution

The number of general trees that are not empty is:

$$T(z) = z + zT(z) + zT(z)^2 + \cdots$$
$$T(z) = zSEQ(T(z))$$
$$T(z) = \frac{z}{1 - T(z)}$$
$$T(z)^2 - T(z) + z = 0$$
$$T(z) = \frac{1 - \sqrt{1 - 4z}}{2}$$

The number of general trees that have more than one node is:

$$
\begin{aligned}
U(z) &= zT(z) + zT(z)^2 + \cdots \\
U(z) &= zT(z)SEQ(T(z)) \\
U(z) &= \frac{zT(z)}{1 - T(z)} \\
U(z) &= \frac{z\frac{1-\sqrt{1-4z}}{2}}{1 - \frac{1-\sqrt{1-4z}}{2}} \\
U(z) &= \frac{z(1-\sqrt{1-4z})}{1+\sqrt{1-4z}} \\
U(z) &= \frac{z(1-\sqrt{1-4z})(1-\sqrt{1-4z})}{(1+\sqrt{1-4z})(1-\sqrt{1-4z})} \\
U(z) &= \frac{(1-\sqrt{1-4z})^2}{4}
\end{aligned}
$$

Therefore, the number of forests with trees not consisting of a single node is:

$$
\begin{aligned}
F(z) &= SEQ(U(z)) \\
F(z) &= \frac{1}{1 - \frac{(1-\sqrt{1-4z})^2}{4}} \\
F(z) &= \frac{4}{4 - (1-\sqrt{1-4z})^2} \\
F(z) &= \frac{4}{4 - 1 + 2\sqrt{1-4z} - (1-4z)} \\
F(z) &= \frac{2}{1 + 2z + \sqrt{1-4z}} \\
F(z) &= \frac{2(1+2z-\sqrt{1-4z})}{(1+2z+\sqrt{1-4z})(1+2z-\sqrt{1-4z})} \\
F(z) &= \frac{2(1+2z-\sqrt{1-4z})}{1+4z+4z^2-(1-4z)} \\
F(z) &= \frac{1+2z-\sqrt{1-4z}}{2z(2+z)} \\
F(z) &= \frac{1}{4}\left(\frac{1}{z} - \frac{\sqrt{1-4z}}{z} + \frac{3}{2+z} + \frac{\sqrt{1-4z}}{2+z}\right) \\
F(z) &= \frac{1}{4}\left(\frac{1}{z} - \frac{\sqrt{1-4z}}{z} + \frac{3}{2}\left(\frac{1}{1+z/2}\right) + \left(\frac{\sqrt{1-4z}}{2+z}\right)\right) \\
z[^n]F(z) &= \frac{1}{4}\left([z=-1] - [z^n]\left(\frac{\sqrt{1-4z}}{z}\right) + \frac{3}{2}(-2)^{-n} + [z^n]\left(\frac{\sqrt{1-4z}}{2+z}\right)\right)
\end{aligned}
$$

Let's change the variable on the last term, introducing $y = 4z$ so $z = y/4$:

$$
[z^n]\left(\frac{\sqrt{1-4z}}{2+z}\right) \to [y^n]\left(\frac{\sqrt{1-y}}{2+y/4}\right)
$$

We can now use the radius of convergence transfer theorem, by setting $f(y) = \frac{1}{2+y/4}$ and $\alpha = -1/2$:

$$
\begin{aligned}
[y^n]\frac{f(y)}{(1-y)^\alpha} &\sim \frac{f(1)}{\Gamma(\alpha)} n^{\alpha-1} \\
[y^n]\frac{f(y)}{(1-y)^\alpha} &\sim \left(\frac{1}{2+1/4}\right)\frac{1}{\Gamma(-1/2)} n^{-\frac{1}{2}-1} \\
[y^n]\frac{f(y)}{(1-y)^\alpha} &\sim -\left(\frac{4}{9}\right)\frac{1}{2\sqrt{\pi}} n^{-\frac{3}{2}} \\
[y^n]\frac{f(y)}{(1-y)^\alpha} &\sim -\frac{2}{9n\sqrt{n\pi}}
\end{aligned}
$$

Taking this result back to z:

$$
\begin{aligned}
\frac{f(y)}{(1-y)^\alpha} &\sim \sum_n -\frac{2}{9n\sqrt{n\pi}} y^n \\
\frac{f(4z)}{(1-4z)^\alpha} &\sim \sum_n -\frac{2}{9n\sqrt{n\pi}} (4z)^n \\
[z^n]\frac{\sqrt{1-4z}}{2+z} &\sim -\frac{2^{2n+1}}{9n\sqrt{n\pi}}
\end{aligned}
$$

We still have the second term to solve:

$$
\begin{aligned}
[z^n]\frac{\sqrt{1-4z}}{z} &= (-4)^{n+1}\binom{1/2}{n+1} \\
[z^n]\frac{\sqrt{1-4z}}{z} &= -\binom{2n+2}{n+1}\frac{1}{2n+1} \\
[z^n]\frac{\sqrt{1-4z}}{z} &\sim -\frac{2^{2n+2}}{\sqrt{\pi(n+1)}}\frac{1}{2n+1}
\end{aligned}
$$

The terms $[z=-1]$ and $\frac{3}{2}(-2)^{-n}$ are $O(1)$ so we'll ignore them. The sum of the remaining terms are:

$$
\begin{aligned}
[z^n]F(z) &\sim \frac{1}{4}\left(\frac{1}{2n+1}\frac{2^{2n+2}}{\sqrt{\pi(n+1)}} - \frac{2^{2n+1}}{9n\sqrt{n\pi}}\right) \\
[z^n]F(z) &\sim \frac{1}{4}\left(\frac{2^{2n+1}}{n\sqrt{n\pi}} - \frac{2^{2n+1}}{9n\sqrt{n\pi}}\right) \\
[z^n]F(z) &\sim \frac{1}{4}\left(1-\frac{1}{9}\right)\left(\frac{2^{2n+1}}{n\sqrt{n\pi}}\right) \\
[z^n]F(z) &\sim \frac{1}{9}\left(\frac{2^{2n+2}}{n\sqrt{n\pi}}\right)
\end{aligned}
$$

The total number of forests is equal to the total number of binary trees, which is given by the Catalan Numbers:

$$B(n) \sim \frac{2^{2n}}{n\sqrt{n\pi}}$$

Dividing the two quantities, we get the asymptotic for the desired proportion:

$$\mu \sim \frac{1}{9}\frac{2^{2n+2}}{n\sqrt{n\pi}} \Big/ \frac{2^{2n}}{n\sqrt{n\pi}}$$
$$\mu \sim \frac{4}{9}$$

2.2 Exercise 6.27

For $N = 2^n - 1$, what is the probability that a perfectly balanced tree structure (all 2^n external nodes on level n) will be built, if all $N!$ key insertion sequences are equally likely?

Solution

If the tree is perfectly balanced, then the root node is the median, and the left and right subtrees have the same size. Let $T(2^n-1)$ be the number of permutations leading to a balanced tree when the permutation has $2^n - 1$ members. The first element is the median, left and right subtrees have each $\frac{(2^n-1)-1}{2} = 2^{n-1} - 1$ members each. We can mix the elements of the subtrees in $\binom{2^n-2}{2^{n-1}-1}$ ways (you should place the $2^{n-1} - 1$ elements from the left tree into $2^n - 2$ positions, and then the right subtree positions are determined). Therefore, our recurrence is:

$$T(2^n - 1) = \binom{2^n - 2}{2^{n-1} - 1}(T(2^{n-1} - 1))^2$$

Let $k = 2^n - 1$, then:

$$\begin{aligned}
T(2^n - 1) &= \binom{2^n - 2}{2^{n-1} - 1}T(2^{n-1} - 1)^2 \\
T(k) &= \binom{k-1}{\frac{k-1}{2}}T(\frac{k-1}{2})^2 \\
\frac{T(k)}{k!} &= \frac{1}{k!}\left(\frac{(k-1)!}{\left(\frac{k-1}{2}!\right)^2}\right)T(\frac{k-1}{2})^2 \\
\frac{T(k)}{k!} &= \frac{(k-1)!}{k!}\frac{T(\frac{k-1}{2})^2}{\left(\frac{k-1}{2}!\right)^2} \\
\frac{T(k)}{k!} &= \frac{1}{k}\left(\frac{T(\frac{k-1}{2})}{\frac{k-1}{2}!}\right)^2
\end{aligned}$$

Now let $Q(k) = T(k)/k!$:

$$\frac{T(k)}{k!} = \frac{1}{k}\left(\frac{T(\frac{k-1}{2})}{\frac{k-1}{2}!}\right)^2$$
$$Q(k) = \frac{1}{k}Q(\frac{k-1}{2})^2$$
$$Q(2^n-1) = \frac{1}{2^n-1}Q(2^{n-1}-1)^2$$

Let $S(n) = Q(2^n - 1)$:

$$Q(2^n-1) = \frac{1}{2^n-1}Q(2^{n-1}-1)^2$$
$$S(n) = \frac{1}{2^n-1}S(n-1)^2$$
$$S(n) = \prod_{0\le k<n}\left(\frac{1}{2^{n-k}-1}\right)^{2^k}$$
$$S(n) = \exp\left(\log\left(\prod_{0\le k<n}\left(\frac{1}{2^{n-k}-1}\right)^{2^k}\right)\right)$$
$$S(n) = \exp\left(\sum_{0\le k<n}2^k\log\left(\frac{1}{2^{n-k}-1}\right)\right)$$
$$S(n) = \exp\left(\sum_{0\le k<n}2^k\log\left(2^{k-n}\frac{1}{1-2^{k-n}}\right)\right)$$
$$S(n) = \exp\left(\sum_{0\le k<n}2^k\left((k-n)\log 2+\log\frac{1}{1-2^{k-n}}\right)\right)$$
$$S(n) = \exp\left(\sum_{0\le k<n}2^k(k-n)\log 2+\sum_{0\le k<n}2^k\log\frac{1}{1-2^{k-n}}\right)$$

Now we have two sums to evaluate. The first one is:

$$\sum_{0\le k<n} 2^k(k-n)\log 2 = \log 2\left(\sum_{0\le k<n} k2^k + \sum_{0\le k<n} n2^k\right)$$
$$\sum_{0\le k<n} 2^k(k-n)\log 2 = \log 2\left(((n-2)2^n+2)-(n(2^n-1))\right)$$
$$\sum_{0\le k<n} 2^k(k-n)\log 2 = (n+2-2^{n+1})\log 2$$

For the second sum, we notice that $2^{k-n}<1$ and therefore we can expand the log into a series:

$$\sum_{0\le k<n} 2^k\log\frac{1}{1-2^{k-n}} = \sum_{0\le k<n} 2^k\sum_{j\ge 1}\frac{\left(2^{k-n}\right)^j}{j}$$
$$\sum_{0\le k<n} 2^k\log\frac{1}{1-2^{k-n}} = \sum_{0\le k<n}\sum_{j\ge 1} 2^k\frac{2^{kj}2^{-nj}}{j}$$
$$\sum_{0\le k<n} 2^k\log\frac{1}{1-2^{k-n}} = \sum_{j\ge 1}\sum_{0\le k<n} 2^k\frac{2^{kj}2^{-nj}}{j}$$
$$\sum_{0\le k<n} 2^k\log\frac{1}{1-2^{k-n}} = \sum_{j\ge 1}\frac{1}{j2^{nj}}\sum_{0\le k<n} 2^{k(j+1)}$$
$$\sum_{0\le k<n} 2^k\log\frac{1}{1-2^{k-n}} = \sum_{j\ge 1}\frac{1}{j2^{nj}}\sum_{0\le k<n}\left(2^{j+1}\right)^k$$
$$\sum_{0\le k<n} 2^k\log\frac{1}{1-2^{k-n}} = \sum_{j\ge 1}\frac{1}{j2^{nj}}\frac{2^{n(j+1)}-1}{2^{j+1}-1}$$
$$\sum_{0\le k<n} 2^k\log\frac{1}{1-2^{k-n}} = \sum_{j\ge 1}\frac{1}{j}\frac{2^n-2^{-nj}}{2^{j+1}-1}$$

On the last expression, 2^{-nj} is negligible next to 2^n, so we'll ignore it.

$$\sum_{0\le k<n} 2^k\log\frac{1}{1-2^{k-n}} = \sum_{j\ge 1}\frac{1}{j}\frac{2^n-2^{-nj}}{2^{j+1}-1}$$
$$\sum_{0\le k<n} 2^k\log\frac{1}{1-2^{k-n}} \sim \sum_{j\ge 1}\frac{1}{j}\frac{2^n}{2^{j+1}-1}$$
$$\sum_{0\le k<n} 2^k\log\frac{1}{1-2^{k-n}} \sim 2^n\sum_{j\ge 1}\frac{1}{j(2^{j+1}-1)}$$

This last sum is convergent, its value is 0.440539 (found numerically). Now we can rewind everything:

$$\begin{aligned}
S(n) &\sim \exp\left((n+2-2^{n+1})\log 2 + 0.440539 \times 2^n\right)\\
S(n) &\sim (\exp(\log 2))^{n+2-2^{n+1}} \exp(0.440539 \times 2^n)\\
S(n) &\sim 2^{n+2-2^{n+1}} \exp(0.440539 \times 2^n)\\
Q(2^n-1) &\sim 2^{n+2-2^{n+1}} \exp(0.440539 \times 2^n)\\
\frac{T(2^n-1)}{(2^n-1)!} &\sim 2^{n+2-2^{n+1}} \exp(0.440539 \times 2^n)\\
T(2^n-1) &\sim (2^n-1)! \left(2^{n+2-2^{n+1}} \exp(0.440539 \times 2^n)\right)
\end{aligned}$$

If we make $N = 2^n - 1$, then $n = \log_2(N+1)$ and we'll have:

$$\begin{aligned}
T(N) &\sim N! \left(2^{n+2-2^{n+1}} \exp(0.440539 \times 2^n)\right)\\
T(N) &\sim N! \left(2^{\log_2(N+1)+2-2^{\log_2(N+1)+1}} \exp(0.440539 \times 2^{\log_2(N+1)})\right)\\
T(N) &\sim N! \left(2^{\log_2(N+1)+2-2(N+1)} \exp(0.440539(N+1))\right)\\
T(N) &\sim N! \left((N+1)2^{-2N} (\exp(0.440539))^{N+1}\right)\\
T(N) &\sim N! \left(4(N+1)4^{-(N+1)} (\exp(0.440539))^{N+1}\right)\\
T(N) &\sim N! \left(4(N+1)\left(\frac{\exp(0.440539)}{4}\right)^{N+1}\right)\\
T(N) &\sim N! \left(4(N+1)0.388386^{N+1}\right)
\end{aligned}$$

To get the desired probability, we divide by the total number of permutations, which is $N!$:

$$\mu(N) \sim 4(N+1)0.388386^{N+1}$$

2.3 Exercise 6.43

Internal nodes in binary trees fall into one of three classes: they have either two, one, or zero external children. What fraction of the nodes are of each type, in a random binary search tree of N nodes?

Solution

We'll use the following theorem from the book: Consider all binary search trees of size N, and let $e(N)$ be the number of trees where the root has the desired property.

Let $E(z)$ be the exponential generating function of $e(N)$. Then, the total number of nodes with the desired property is given by the following EGF:

$$C(z) = \frac{1}{(1-z)^2}\left(E(0) + \int_0^z (1-x)^2 E'(x)dx\right)$$

Let's do the math for each case. For internal nodes with 0 external children (leaves), $e(n) = 0$ for $n > 1$, since the root will always have at least one child. Also, $e(1) = 1$, so in general $e(n) = [n = 1]$. The EGF is:

$$\begin{aligned} E(z) &= \sum_{n\geq 0} \frac{[n=1]}{n!} z^n \\ E(z) &= z \\ E'(z) &= 1 \end{aligned}$$

Substituting:

$$\begin{aligned} C(z) &= \frac{1}{(1-z)^2}\left(\int_0^z (1-x)^2 dx\right) \\ C(z) &= \frac{1}{(1-z)^2}\left(x - x^2 + \frac{x^3}{3}\Big|_0^z\right) \\ C(z) &= \frac{1}{(1-z)^2}\left(z - z^2 + \frac{z^3}{3}\right) \\ C(z) &= \frac{z}{(1-z)^2} - \frac{z^2}{(1-z)^2} + \frac{1}{3}\frac{z^3}{(1-z)^2} \\ c(n) &= n!\left(n - (n-1) + \frac{1}{3}(n-2)\right) \\ c(n) &= n!\left(\frac{n+1}{3}\right) \end{aligned}$$

This is the total number of leaves in all binary search trees of size N. If we divide by $N!$, then we get the average number of leaves in each tree. Dividing again by N, we get the ratio between leaves and nodes on a random binary search tree.

$$\begin{aligned} \mu &= \frac{n!}{n!n}\frac{n+1}{3} \\ \mu &= \frac{1}{3} + \frac{1}{3n} \\ \mu &\sim \frac{1}{3} \end{aligned}$$

Now let's do the case where a node has exactly one child. Clearly there's no such node for trees of size 1. For sizes $n > 1$, the root has one child only when the root is the greatest or the smallest element of the permutation. In each of these cases, there are $(N-1)!$ permutations starting with this element. The EGF for $n \leq 1$ is:

$$
\begin{aligned}
e(n) &= 2(n-1)! - 2[n=1] \\
E(z) &= \sum_{z\geq 0}(2(n-1)! - 2[n=1])\frac{z^n}{n!} \\
E(z) &= \sum_{z\geq 0}\frac{2(n-1)!}{n!}z^n - \sum_{z\geq 0}2[n=1]\frac{z^n}{n!} \\
E(z) &= 2\left(\sum_{z\geq 1}\frac{1}{n}z^n\right) - 2z \\
E(z) &= -2z + 2\log\frac{1}{1-z} \\
E'(z) &= -2 + \frac{2}{1-z}
\end{aligned}
$$

Substituting:

$$
\begin{aligned}
C(z) &= \frac{1}{(1-z)^2}\left(\int_0^z (1-x)^2\left(-2+\frac{1}{1-x}\right)dx\right) \\
C(z) &= \frac{2}{(1-z)^2}\left(\int_0^z -(1-x)^2 + (1-x)dx\right) \\
C(z) &= \frac{2}{(1-z)^2}\left(\frac{x^2}{2} - \frac{x^3}{3}\Big|_0^z\right) \\
C(z) &= \frac{3z^2-2z^3}{3(1-z)^2} \\
C(z) &= \frac{z^2}{(1-z)^2} - \frac{2z^3}{3(1-z)^2} \\
C(z) &= \frac{z^2}{(1-z)^2} - \frac{2z^3}{3(1-z)^2} \\
c(n) &= n!\left((n-1) - \frac{2}{3}(n-2)\right) \\
c(n) &= n!\left(\frac{n+1}{3}\right)
\end{aligned}
$$

We conclude that the number of nodes with exactly one child is the same as the number of nodes without children, and the ratio is:

$$\mu \sim \frac{1}{3}$$

On the last case, the number of trees whose root has two children is $(N-2)(N-1)!$, for a tree of size $N > 1$:

$$
\begin{aligned}
e(n) &= (n-2)(n-1)! + [n=1] \\
E(z) &= \sum_{z \geq 0} ((n-2)(n-1)! + [n=1]) \frac{z^n}{n!} \\
E(z) &= \sum_{z \geq 0} \frac{(n-2)(n-1)!}{n!} z^n - \sum_{z \geq 0} [n=1] \frac{z^n}{n!} \\
E(z) &= \left(\sum_{z \geq 1} \frac{(n-2)}{n} z^n \right) + z \\
E(z) &= z + \frac{1}{1-z} - 2 \log \frac{1}{1-z} \\
E'(z) &= 1 + \frac{1}{(1-z)^2} - \frac{2}{1-z}
\end{aligned}
$$

Substituting:

$$
\begin{aligned}
C(z) &= \frac{1}{(1-z)^2} \left(\int_0^z (1-x)^2 \left(1 + \frac{1}{(1-x)^2} - \frac{2}{1-x} \right) dx \right) \\
C(z) &= \frac{1}{(1-z)^2} \left(\int_0^z (1-x)^2 + 1 + 2(1-x) dx \right) \\
C(z) &= \frac{1}{(1-z)^2} \left(\frac{x^3}{3} \Big|_0^z \right) \\
C(z) &= \frac{z^3}{3(1-z)^2} \\
c(n) &= \frac{n-2}{3}
\end{aligned}
$$

The ratio is:

$$
\begin{aligned}
\mu &= \frac{n!}{n!n} \frac{n-2}{3} \\
\mu &\sim \frac{1}{3}
\end{aligned}
$$

2.4 Exercise 7.29

An arrangement of N elements is a sequence formed from a subset of the elements. Prove that the EGF for arrangements is $e^z/(1-z)$. Express the coefficients as a simple sum and interpret that sum combinatorially.

Solution

The set of all arrangements of size n is the same as the set composed of all permutations of size k, for all $0 \leq k \leq n$. We can model this as a cartesian par of urns and permutations: take all permutations of size k, and place all $n-k$ remaining elements on a urn. Using this construct, we will have $k!$ permutations, times 1 urn (the order of the $n-k$ elements inside the urn is not relevant). This leads directly to the symbolic construct $SEQ(z) * SET(z)$. The EGF is given by:

$$G(z) = e^z \times \frac{1}{1-z} = \frac{e^z}{1-z}$$

The coefficients can be found by the binomial convolution of EGFs:

$$G(z) = A(z)B(z) \implies g(n) = \sum_{0 \leq k \leq n} \binom{n}{k} a_k b_{n-k}$$

In our case:

$$g(n) = \sum_{0 \leq k \leq n} \binom{n}{k} k!$$

The combinatoric interpretation matches our intuition: the set of all arrangements is the same as the set composed of all permutations of size k, for all $0 \leq k \leq n$. The formula sums, for all $0 \leq k \leq n$, k elements chosen from n, and then apply $k!$ permutations on them.

2.5 Exercise 7.45

Find the CGF for the total number of inversions in all involutions of length N. Use this to find the average number of inversions in an involution.

Solution

For every inversion (a, b), with $1 \leq a < b \leq n$, let's count in how many involutions of size n it appears. There are three cases we must consider. Notation used: $I(n)$ is the total number of inversions in all involutions of size n; $I_1(n)$, $I_2(n)$ and $I_3(n)$ are the total number of inversions due to cases 1, 2 and 3; $\mathrm{invol}(n)$ is the number of involutions of size n, whose EGF is $\exp(z + z^2/2)$.

Case 1: Both a and b are in the same two-cycle. In this case there is one inversion for every $\mathrm{invol}(n-2)$ involutions where this two-cycle appears. The total number of inversions generated by this case is:

$$I_1(n) = \sum_{1 \le a < b \le n} \mathrm{invol}(n-2)$$
$$I_1(n) = \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} \mathrm{invol}(n-2)$$
$$I_1(n) = \sum_{a=1}^{n-1} (n-a)\,\mathrm{invol}(n-2)$$
$$I_1(n) = \sum_{k=1}^{n-1} k\,\mathrm{invol}(n-2)$$
$$I_1(n) = \frac{n(n-1)}{2}\mathrm{invol}(n-2)$$

Case 2: One of the elements is in a one-cycle, the other is in a two-cycle. Two subcases: this is either (a) and (b, r); or (b) and (a, r). In the first subcase, there is a inversion on invol$(n-3)$ involutions if $r < a$, on the second case if $r > b$. The total is:

$$I_2(n) = \sum_{1 \le a < b \le n} \left(\sum_{r<a} \mathrm{invol}(n-3) + \sum_{r>b} \mathrm{invol}(n-3) \right)$$
$$I_2(n) = \sum_{1 \le a < b \le n} \mathrm{invol}(n-3) \left(\sum_{r<a} 1 + \sum_{r>b} 1 \right)$$
$$I_2(n) = \sum_{1 \le a < b \le n} \mathrm{invol}(n-3) \left((a-1) + (n-b) \right)$$
$$I_2(n) = \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} (n+a-1-b)\mathrm{invol}(n-3)$$
$$I_2(n) = \sum_{a=1}^{n-1} \left((n-a)(n+a-1) - (n-a)\frac{n+a+1}{2} \right) \mathrm{invol}(n-3)$$
$$I_2(n) = \sum_{a=1}^{n-1} \left(\frac{1}{2}(n-a)(n+a-3) \right) \mathrm{invol}(n-3)$$
$$I_2(n) = \frac{1}{3}(n^3 - 3n^2 + 2n)\mathrm{invol}(n-3)$$
$$I_2(n) = \frac{1}{3}n(n-1)(n-2)\mathrm{invol}(n-3)$$

Case 3: Both elements are in different two-cycles. Let's call them (a, r) and (b, s). There is an inversion in invol$(n-4)$ involutions if $s < r$. The total is:

$$I_3(n) = \sum_{1\le a<b\le n} \sum_{1\le s<r\le n-2} \mathrm{invol}(n-4)$$
$$I_3(n) = \sum_{1\le a<b\le n} \frac{(n-2)(n-3)}{2}\mathrm{invol}(n-4)$$
$$I_3(n) = \frac{n(n-1)}{2}\frac{(n-2)(n-3)}{2}\mathrm{invol}(n-4)$$
$$I_3(n) = \frac{1}{4}n(n-1)(n-2)(n-3)\mathrm{invol}(n-4)$$

Now we'll calculate the EGF for these expressions. For case 1:

$$I_1(z) = \sum_{n\le 0} \frac{n(n-1)}{2}\mathrm{invol}(n-2)\frac{z^n}{n!}$$
$$I_1(z) = \sum_{n\le 0} \frac{1}{2}\mathrm{invol}(n-2)\frac{z^n}{(n-2)!}$$
$$I_1(z) = \frac{z^2}{2}\sum_{n\le 0} \mathrm{invol}(n-2)\frac{z^{n-2}}{(n-2)!}$$
$$I_1(z) = \frac{z^2}{2}\sum_{n\le 2} \mathrm{invol}(n)\frac{z^n}{n!}$$
$$I_1(z) = \frac{z^2}{2}\mathrm{Invol}(z)$$

This last expression is valid for $n > 2$. Using the same method we conclude:

$$I_2(z) = \frac{z^3}{3}\mathrm{Invol}(z)$$
$$I_3(z) = \frac{z^4}{4}\mathrm{Invol}(z)$$

The final EGF for all inversions in all involutions is:

$$I(z) = \left(\frac{z^2}{2} + \frac{z^3}{3} + \frac{z^4}{4}\right)\exp\left(z + \frac{z^2}{2}\right)$$

In order to find the average number of inversions in an involution, we'll consider the contributions of $I_1(n)$, $I_2(n)$ and $I_3(n)$ in separate. We'll also use the result that:

$$\mathrm{invol}(n) \sim \frac{1}{\sqrt{2\sqrt{e}}}e^{\sqrt{n}}\left(\frac{n}{e}\right)^{n/2}$$

For $I_1(n)$:

$$
\begin{aligned}
\mu_1(n) &= \frac{I_1(n)}{\mathrm{invol}(n)} \\
\mu_1(n) &= \frac{\frac{n(n-1)}{2}\mathrm{invol}(n-2)}{\mathrm{invol}(n)} \\
\mu_1(n) &\sim \frac{\frac{n(n-1)}{2} e^{\sqrt{n-2}} \left(\frac{n-2}{e}\right)^{\frac{n-2}{2}}}{e^{\sqrt{n}} \left(\frac{n}{e}\right)^{\frac{n}{2}}} \\
\mu_1(n) &\sim \frac{n(n-1)}{2} e^{\sqrt{n-2}-\sqrt{n}} \left(\frac{n-2}{e}\right)^{\frac{n}{2}-1} \left(\frac{e}{n}\right)^{\frac{n}{2}} \\
\mu_1(n) &\sim \frac{n(n-1)}{2} e^{\sqrt{n-2}-\sqrt{n}} \left(\frac{n-2}{n}\right)^{\frac{n}{2}} \left(\frac{e}{n-2}\right)
\end{aligned}
$$

Now we'll consider what happens to this expression when n is large.

$$
\lim_{n\to\infty} e^{\sqrt{n-2}-\sqrt{n}} = 1
$$

$$
\begin{aligned}
\lim_{n\to\infty} \left(\frac{n-2}{n}\right)^{\frac{n}{2}} &= \lim_{n\to\infty} \left(1-\frac{2}{n}\right)^{\frac{n}{2}} \\
\lim_{n\to\infty} \left(\frac{n-2}{n}\right)^{\frac{n}{2}} &= \lim_{k\to\infty} \left(1-\frac{1}{k}\right)^{k} \\
\lim_{n\to\infty} \left(\frac{n-2}{n}\right)^{\frac{n}{2}} &= e^{-1}
\end{aligned}
$$

Substituting:

$$
\begin{aligned}
\mu_1(n) &\sim \frac{n(n-1)}{2} e^{\sqrt{n-2}-\sqrt{n}} \left(\frac{n-2}{n}\right)^{\frac{n}{2}} \left(\frac{e}{n-2}\right) \\
\mu_1(n) &\sim \frac{n(n-1)}{2} \frac{1}{e} \left(\frac{e}{n-2}\right) \\
\mu_1(n) &\sim \frac{n(n-1)}{2(n-2)} \\
\mu_1(n) &\sim \frac{n}{2}
\end{aligned}
$$

For $I_2(n)$:

$$\mu_2(n) = \frac{I_2(n)}{\text{invol}(n)}$$

$$\mu_2(n) = \frac{\frac{n(n-1)(n-2)}{3}\text{invol}(n-3)}{\text{invol}(n)}$$

$$\mu_2(n) \sim \frac{n(n-1)(n-2)}{3} e^{\sqrt{n-3}-\sqrt{n}} \left(\frac{n-3}{e}\right)^{\frac{n}{2}-\frac{3}{2}} \left(\frac{e}{n}\right)^{\frac{n}{2}}$$

$$\mu_2(n) \sim \frac{n(n-1)(n-2)}{3} e^{\sqrt{n-3}-\sqrt{n}} \left(\frac{n-3}{n}\right)^{\frac{n}{2}} \left(\frac{e}{n-3}\right)^{\frac{3}{2}}$$

The middle expression is:

$$\lim_{n\to\infty} \left(\frac{n-3}{n}\right)^{\frac{n}{2}} = \lim_{n\to\infty} \left(1 - \frac{3}{n}\right)^{\frac{n}{2}}$$

$$\lim_{n\to\infty} \left(\frac{n-3}{n}\right)^{\frac{n}{2}} = \lim_{k\to\infty} \left(1 - \frac{3}{2}\frac{1}{k}\right)^{k}$$

$$\lim_{n\to\infty} \left(\frac{n-2}{n}\right)^{\frac{n}{2}} = e^{-\frac{3}{2}}$$

Substituting:

$$\mu_2(n) \sim \frac{n(n-1)(n-2)}{3} e^{\sqrt{n-3}-\sqrt{n}} \left(\frac{n-3}{n}\right)^{\frac{n}{2}} \left(\frac{e}{n-3}\right)^{\frac{3}{2}}$$

$$\mu_2(n) \sim \frac{n(n-1)(n-2)}{3} e^{-\frac{3}{2}} \left(\frac{e}{n-3}\right)^{\frac{3}{2}}$$

$$\mu_2(n) \sim \frac{n(n-1)(n-2)}{3\left(\sqrt{n-3}\right)^3}$$

$$\mu_2(n) \sim \frac{\left(\sqrt{n}\right)^3}{3}$$

For $I_3(n)$:

$$
\begin{aligned}
\mu_3(n) &= \frac{I_3(n)}{\text{invol}(n)} \\
\mu_3(n) &= \frac{\frac{n(n-1)(n-2)(n-3)}{4}\text{invol}(n-4)}{\text{invol}(n)} \\
\mu_3(n) &\sim \frac{n(n-1)(n-2)(n-3)}{4} e^{\sqrt{n-4}-\sqrt{n}} \left(\frac{n-4}{e}\right)^{\frac{n}{2}-2} \left(\frac{e}{n}\right)^{\frac{n}{2}} \\
\mu_3(n) &\sim \frac{n(n-1)(n-2)(n-3)}{4} \left(\frac{n-4}{n}\right)^{\frac{n}{2}} \left(\frac{e}{n-4}\right)^{2} \\
\mu_3(n) &\sim \frac{n(n-1)(n-2)(n-3)}{4} e^{-2} \left(\frac{e}{n-4}\right)^{2} \\
\mu_3(n) &\sim \frac{n(n-1)(n-2)(n-3)}{4(n-2)^2} \\
\mu_3(n) &\sim \frac{n^2}{4}
\end{aligned}
$$

Summing everything, the average number of inversions in an involution is:

$$
\begin{aligned}
\mu &\sim \frac{n}{2} + \frac{(\sqrt{n})^3}{3} + \frac{n^2}{4} \\
\mu &\sim \frac{(\sqrt{n})^2}{2} + \frac{(\sqrt{n})^3}{3} + \frac{(\sqrt{n})^4}{4}
\end{aligned}
$$

2.6 Exercise 7.61

Use asymptotics from generating functions (see Section 5.5) or a direct argument to show that the probability for a random permutation to have j cycles of length k is asymptotic to the Poisson distribution $e^{-\lambda}\lambda^j/j!$ with $\lambda = 1/k$.

Solution

The construction for random permutations may be expressed as:

$$SET(CYC(z)) = SET(\sum_k CYC_k(z))$$

We'll annotate the kth cycle, turning this EGF into a BGF:

$$R(u,z) = SET(CYC_1(z) + \cdots + CYC_{k-1}(z) + uCYC_k(z) + CYC_{k+1}(z) + \cdots)$$
$$R(u,z) = SET(CYC(z) - CYC_k(z) + uCYC_k(z))$$
$$R(u,z) = \exp\left(\ln\left(\frac{1}{1-z}\right) + (u-1)\frac{z^k}{k}\right)$$
$$R(u,z) = \frac{\exp\left((u-1)\frac{z^k}{k}\right)}{1-z}$$

Since this BGF is also an EGF, the probability of a random permutation of size n to have j cycles of length k is the coefficient of $[u^j z^n]$ on this BGF.

$$[u^j z^n]R(u,z) = [u^j z^n]\frac{\exp\left((u-1)\frac{z^k}{k}\right)}{1-z}$$
$$[u^j z^n]R(u,z) = [u^j z^n]\frac{\exp\left(u\frac{z^k}{k}\right)\exp\left(-\frac{z^k}{k}\right)}{1-z}$$
$$[u^j z^n]R(u,z) = [u^j z^n]\frac{\exp\left(-\frac{z^k}{k}\right)}{1-z}\sum_i \frac{1}{i!}u^i\frac{z^{ki}}{k^i}$$
$$[u^j z^n]R(u,z) = [z^n]\frac{\exp\left(-\frac{z^k}{k}\right)}{1-z}\frac{1}{j!}\frac{z^{kj}}{k^j}$$

Let's use the radius of convergence transfer theorem:

$$[u^j z^n]R(u,z) = [z^n]\frac{\exp\left(-\frac{z^k}{k}\right)}{1-z}\frac{1}{j!}\frac{z^{kj}}{k^j}$$
$$[u^j z^n]R(u,z) \sim \frac{\exp\left(-\frac{1^k}{k}\right)}{\Gamma(1)}\frac{1}{j!}\frac{1^{kj}}{k^j}n^{1-1}$$
$$[u^j z^n]R(u,z) \sim \exp\left(-\frac{1}{k}\right)\frac{1}{j!k^j}$$

If we take $\lambda = 1/k$, then we reach the desired Poisson expression:

$$[u^j z^n]R(u,z) \sim \exp\left(-\frac{1}{k}\right)\frac{1}{j!k^j}$$
$$[u^j z^n]R(u,z) \sim \frac{e^{-\lambda}\lambda^j}{j!}$$

3 Week 3

3.1 Exercise 4.9

If $\alpha < \beta$, show that α^N is exponentially small relative to β^N. For $\beta = 1.2$ and $\alpha = 1.1$, find the absolute and relative errors when $\alpha^N + \beta^N$ is approximated by β^N, for $N = 10$ and $N = 100$.

Solution

	$\alpha^N + \beta^N$	β^N	abs.error	rel.error
$N = 10$	8.78	6.19	2.59	29.5%
$N = 100$	82831755.13	82817974.52	13780.60	0.016%

3.2 Exercise 4.71

Show that $P(N) = \sum_{k\geq 0} \frac{(N-k)^k (N-k)!}{N!} = \sqrt{\pi N/2} + O(1)$

Solution

Let's call the summand Q(n,k):

$$Q(n,k) = \frac{(n-k)^k (n-k)!}{n!}$$
$$Q(n,k) = \exp\left(\ln\left(\frac{(n-k)^k (n-k)!}{n!}\right)\right)$$
$$Q(n,k) = \exp\left(k\ln(n-k) + \ln(n-k)! - \ln n!\right)$$

First we use Stirling's approximation:

$$\ln n! = (n+\frac{1}{2})\ln n - n + \log\sqrt{2\pi} + O(\frac{1}{n})$$

Substituting:

$$Q(n,k) = \exp\left(k\ln(n-k) + \ln(n-k)! - \ln n!\right)$$
$$Q(n,k) = \exp\left(k + (n+\frac{1}{2})\ln\left(1-\frac{k}{n}\right) + O(\frac{1}{n}) + O(\frac{1}{n-k})\right)$$

The summand is 0 when $n = k$. Otherwise, we can get rid of $O(\frac{1}{n-k})$:

$$\frac{1}{n-k} = \frac{\frac{1}{n}}{1-\frac{k}{n}}$$
$$\frac{1}{n-k} = \frac{1}{n}\left(1+O(\frac{k}{n})\right)$$
$$O(\frac{1}{n-k}) = O(\frac{1}{n})O\left(1+O(\frac{k}{n})\right)$$
$$O(\frac{1}{n-k}) = O(\frac{1}{n})+O(\frac{k}{n^2})$$

We'll use now this asymptotic to $\ln$:

$$\ln(1+k) = k - \frac{k^2}{2} + O(k^3)$$

Substituting:

$$Q(n,k) = \exp\left(k + (n+\frac{1}{2})\ln\left(1-\frac{k}{n}\right) + O(\frac{1}{n}) + O(\frac{k}{n^2})\right)$$
$$Q(n,k) = \exp\left(k + (n+\frac{1}{2})(-\frac{k}{n} - \frac{k^2}{2n^2} + O(\frac{k^3}{n^3})) + O(\frac{1}{n}) + O(\frac{k}{n^2})\right)$$
$$Q(n,k) = \exp\left(k - k - \frac{k^2}{2n} - \frac{k}{2n} - \frac{k^2}{4n^2} + O(\frac{k^3}{n^2}) + O(\frac{k^3}{n^3}) + O(\frac{1}{n}) + O(\frac{k}{n^2})\right)$$
$$Q(n,k) = \exp\left(-\frac{k}{2n}\left(k-1-\frac{1}{2n}\right) + O(\frac{k^3}{n^2}) + O(\frac{k^3}{n^3}) + O(\frac{1}{n}) + O(\frac{k}{n^2})\right)$$
$$Q(n,k) = \exp\left(-\frac{k}{2n}\left(k+O(1)\right) + O(\frac{k^3}{n^2}) + O(\frac{k^3}{n^3}) + O(\frac{1}{n}) + O(\frac{k}{n^2})\right)$$
$$Q(n,k) = \exp\left(-\frac{k^2}{2n} + O(\frac{k}{n}) + O(\frac{k^3}{n^2}) + O(\frac{k^3}{n^3}) + O(\frac{1}{n}) + O(\frac{k}{n^2})\right)$$

Some absorptions: $O(\frac{1}{n})$ is absorbed by $O(\frac{k}{n})$, and both $O(\frac{k}{n^2})$ and $O(\frac{k^3}{n^3})$ are absorbed by $O(\frac{k^3}{n^2})$:

$$Q(n,k) = \exp\left(-\frac{k^2}{2n} + O(\frac{k}{n}) + O(\frac{k^3}{n^2}) + O(\frac{k^3}{n^3}) + O(\frac{1}{n}) + O(\frac{k}{n^2})\right)$$
$$Q(n,k) = \exp\left(-\frac{k^2}{2n} + O(\frac{k}{n}) + O(\frac{k^3}{n^2})\right)$$
$$Q(n,k) = \exp\left(-\frac{k^2}{2n}\right)\left(1 + O(\frac{k}{n}) + O(\frac{k^3}{n^2})\right)$$

Let's get back to the sum. We can choose $k_0 = o(n^{2/3})$ and split the sum into two parts:

$$\sum_{k\geq 0} Q(n,k) = \sum_{0\leq k\leq k_0} Q(n,k) + \sum_{k>k_0} Q(n,k)$$

In the second part, all terms are exponentially small. We'll call them Δ.

$$\sum_{k>k_0} \exp\left(-\frac{k^2}{2n}\right)\left(1 + O(\frac{k}{n}) + O(\frac{k^3}{n^2})\right) = \Delta$$

The first part can be further split into three sums:

$$\sum_{0\leq k\leq k_0} Q(n,k) = \sum_{0\leq k\leq k_0} \exp\left(-\frac{k^2}{2n}\right)\left(1 + O(\frac{k}{n}) + O(\frac{k^3}{n^2})\right)$$
$$\sum_{0\leq k\leq k_0} Q(n,k) = \sum_{0\leq k\leq k_0} e^{-\frac{k^2}{2n}} + \sum_{0\leq k\leq k_0} e^{-\frac{k^2}{2n}} O(\frac{k}{n}) + \sum_{0\leq k\leq k_0} e^{-\frac{k^2}{2n}} O(\frac{k^3}{n^2})$$

Let's prove that the second sum is $O(1)$. First of all, notice that:

$$\frac{d}{dk}\left(e^{-\frac{k^2}{2n}}\right) = -\frac{k}{n}e^{-\frac{k^2}{2n}}$$

This implies that:

$$\int \frac{k}{n}e^{-\frac{k^2}{2n}}dk = -e^{-\frac{k^2}{2n}}$$

To prove that $\sum_{0\leq k\leq k_0} e^{-\frac{k^2}{2n}} O(\frac{k}{n}) = O(1)$ is the same as to prove that $\sum_{0\leq k\leq k_0} \frac{k}{n}e^{-\frac{k^2}{2n}} < C$ for some constant C independent of n or k. Let's find this constant.

$$\sum_{0 \leq k \leq k_0} \frac{k}{n} e^{-\frac{k^2}{2n}} < \sum_{0 \leq k \leq \infty} \frac{k}{n} e^{-\frac{k^2}{2n}}$$
$$\sum_{0 \leq k \leq k_0} \frac{k}{n} e^{-\frac{k^2}{2n}} < \int_0^\infty \frac{k}{n} e^{-\frac{k^2}{2n}}$$
$$\sum_{0 \leq k \leq k_0} \frac{k}{n} e^{-\frac{k^2}{2n}} < \left(-e^{-\frac{k}{2n}}\right)\Big|_0^\infty$$
$$\sum_{0 \leq k \leq k_0} \frac{k}{n} e^{-\frac{k^2}{2n}} < 1$$
$$\sum_{0 \leq k \leq k_0} \frac{k}{n} e^{-\frac{k^2}{2n}} = O(1)$$

The same can be done for the third sum:

$$\sum_{0 \leq k \leq k_0} \frac{k^3}{n^2} e^{-\frac{k^2}{2n}} < \sum_{0 \leq k \leq \infty} \frac{k^3}{n^2} e^{-\frac{k^2}{2n}}$$
$$\sum_{0 \leq k \leq k_0} \frac{k^3}{n^2} e^{-\frac{k^2}{2n}} < \int_0^\infty \frac{k^3}{n^2} e^{-\frac{k^2}{2n}}$$
$$\sum_{0 \leq k \leq k_0} \frac{k^3}{n^2} e^{-\frac{k^2}{2n}} < \left(-\left(\frac{k^2}{n} + 2\right) e^{-\frac{k^2}{2n}}\right)\Big|_0^\infty$$
$$\sum_{0 \leq k \leq k_0} \frac{k^3}{n^2} e^{-\frac{k^2}{2n}} < 2$$
$$\sum_{0 \leq k \leq k_0} \frac{k^3}{n^2} e^{-\frac{k^2}{2n}} = O(1)$$

Plugging back both $O(1)$s:

$$\sum_{0 \leq k \leq k_0} Q(n,k) = \sum_{0 \leq k \leq k_0} e^{-\frac{k^2}{2n}} + \sum_{0 \leq k \leq k_0} e^{-\frac{k^2}{2n}} O(\frac{k}{n}) + \sum_{0 \leq k \leq k_0} e^{-\frac{k^2}{2n}} O(\frac{k^3}{n^2})$$
$$\sum_{0 \leq k \leq k_0} Q(n,k) = \sum_{0 \leq k \leq k_0} e^{-\frac{k^2}{2n}} + O(1)$$
$$\sum_{0 \leq k \leq k_0} Q(n,k) = \sum_{k \geq 0} e^{-\frac{k^2}{2n}} - \sum_{k > k_0} e^{-\frac{k^2}{2n}} + O(1)$$

Again we have two sums. The second sum is exponentially small, we'll call it Δ_1. The first sum can be approximated by a Gaussian integral:

$$\sum_{k\geq 0} e^{-\frac{k^2}{2n}} = \int_0^\infty e^{-\frac{k^2}{2n}} dk + O(1)$$
$$\sum_{k\geq 0} e^{-\frac{k^2}{2n}} = \frac{1}{2}\sqrt{2\pi n} + O(1)$$
$$\sum_{k\geq 0} e^{-\frac{k^2}{2n}} = \sqrt{\frac{\pi n}{2}} + O(1)$$

Summing everything:

$$\sum_{k\geq 0} Q(n,k) = \sqrt{\frac{\pi n}{2}} + \Delta - \Delta_1 + O(1)$$

Using Laplace's method of bounding-and-extending the tail, we know that $\Delta - \Delta_1 = O(1)$, so the final answer is:

$$\sum_{k\geq 0} Q(n,k) = \sqrt{\frac{\pi n}{2}} + O(1)$$

3.3 Exercise 5.1

How many bitstrings of length N have no 000?

Solution

Let's start by counting all bitstrings ending in 1. These bitstrings are formed by sequences of tokens 1, 01, 001, and therefore its construction is $SEQ(A_1 + A_{01} + A_{001})$. The bitstring ending in zero are the ones ending in 1, concatenated with either 00, 0, or empty. The final construction is $SEQ(A_1 + A_{01} + A_{001}) \times (\epsilon + A_0 + A_{00})$. The generating function is:

$$A(z) = \frac{1 + z + z^2}{1 - (z + z^2 + z^3)}$$
$$A(z) = \frac{1 + z + z^2}{1 - z - z^2 - z^3}$$

The smallest pole is $1/\beta = 0.543689$ with multiplicity 1 (found numerically). We can now use this transfer theorem:

$$
\begin{aligned}
[z^n]\frac{f(z)}{g(z)} &\sim \nu\frac{(-\beta)^\nu f(1/\beta)}{g^{(\nu)}(1/\beta)}\beta^n n^{\nu-1}\\
[z^n]A(z) &\sim \frac{-\beta f(1/\beta)}{g\prime(1/\beta)}\beta^n\\
[z^n]A(z) &\sim \frac{-\beta(1+1/\beta+1/\beta^2)}{-3/\beta^2-2/\beta-1}\beta^n\\
[z^n]A(z) &\sim \beta^{n+1}\frac{\beta^2+\beta+1}{\beta^2+2\beta+3}\\
[z^n]A(z) &\sim 1.13745\times 1.83929^n
\end{aligned}
$$

3.4 Exercise 5.3

Let $\mathcal{U}$ be the set of binary trees with the size of a tree defined to be the total number of nodes (internal plus external), so that the generating function for its counting sequence is $U(z) = z + z^3 + 2z^5 + 5z^7 + 14z^9 + \ldots$. Derive an explicit expression for $U(z)$.

Solution

A tree is an external node, or an internal node connected to two trees. Its construction is $T = A_{external} + A_{internal} \times T \times T$. We're counting both internal and external nodes, so $A_{external} = z$ and $A_{internal} = z$:

$$
\begin{aligned}
T &= A_{external} + A_{internal} \times T \times T\\
U(z) &= z + zU(z)^2\\
U(z) &= \frac{1 \pm \sqrt{1-4z^2}}{2z}
\end{aligned}
$$

Only the negative square root will lead to positive solutions for $U(n)$, so the solution is:

$$
U(z) = \frac{1-\sqrt{1-4z^2}}{2z}
$$

3.5 Exercise 5.7

Derive an EGF for the number of permutations whose cycles are all of odd length.

Solution

The construction is directly $SET(\sum_{\text{odd k}} CYC_k(Z))$. The EGF is:

$$P(z) = \exp\left(\sum_{\text{odd k}} \frac{z^k}{k}\right)$$
$$P(z) = \exp\left(\sum_{k} \frac{z^k}{k}[\text{k is odd}]\right)$$
$$P(z) = \exp\left(\sum_{k} \frac{z^k}{k}\left(\frac{1-(-1)^k}{2}\right)\right)$$
$$P(z) = \exp\left(\frac{1}{2}\sum_{k} \frac{z^k}{k} - \frac{1}{2}\sum_{k} \frac{(-z)^k}{k}\right)$$
$$P(z) = \exp\left(\frac{1}{2}\ln\left(\frac{1}{1-z}\right) + \frac{1}{2}\ln(1+z)\right)$$
$$P(z) = \exp\left(\frac{1}{2}\ln\left(\frac{1+z}{1-z}\right)\right)$$
$$P(z) = \sqrt{\frac{1+z}{1-z}}$$

3.6 Exercise 5.7

Find the average number of internal nodes in a binary tree of size N with both children internal.

Solution

Let's consider only trees whose root is an internal node. There are four kinds of roots: linking to 2 external nodes (size 1, cost 0, z^1u^0), left child is an external node and right is an internal node (size 1, cost 0, z^1u^0), the reversed of the previous one (also z^1u^0), and both children internal (size 1, cost 1, z^1u^1). The construction is:

$$P(u,z) = z + 2zP(u,z) + uzP(u,z)^2$$
$$P(u,z) = \frac{1-2z-\sqrt{1-4z+4z^2-4uz^2}}{2uz}$$

By setting $u=1$ we have the total number of trees (whose root is an internal node):

$$P(1,z) = \frac{1-2z-\sqrt{1-4z}}{2z}$$
$$zP(1,z) = \frac{1-2z-\sqrt{1-4z}}{2}$$
$$[z^n]\left(zP(1,z)\right) = \frac{1}{2}\left([n=0] - 2[n=1] + \frac{1}{2n-1}\binom{2n}{n}\right)$$

The cumulative generating function of $P(u,z)$ is the partial derivative of $P(u,z)$ on u, evaluated at $u=1$.

$$
\begin{aligned}
P(u,z) &= \frac{1-2z-\sqrt{1-4z+4z^2-4uz^2}}{2uz} \\
\frac{\partial}{\partial u}P(u,z) &= \frac{z}{u\sqrt{1-4z+4z^2-4uz^2}} - \frac{1-2z-\sqrt{1-4z+4z^2-4uz^2}}{2u^2z} \\
P'(1,z) &= \frac{z}{\sqrt{1-4z}} - \frac{1-2z}{2z} - \frac{\sqrt{1-4z}}{2z} \\
[z^n]P'(1,z) &= \binom{2n-2}{n-1} + [n=0] - \frac{1}{2}[n=-1] - \frac{1}{4n+2}\binom{2n+2}{n+1}
\end{aligned}
$$

Now we just have to divide the results. For $n>1$:

$$
\begin{aligned}
\mu &= \frac{[z^n]P'(1,z)}{[z^n]P(1,z)} \\
\mu &= \frac{\binom{2n-2}{n-1} - \frac{1}{4n+2}\binom{2n+2}{n+1}}{\frac{1}{2(2(n+1)-1)}\binom{2(n+1)}{n+1}} \\
\mu &= \frac{(n-1)(n-2)}{4n-2}
\end{aligned}
$$

3.7 Exercise 5.16

Find the average number of internal nodes in a binary tree of size N with one child internal and one child external.

Solution

This is the same as the previous exercise, but now the costs are z^1u^0, $2z^1u^1$ and z^1u^0.

$$
\begin{aligned}
P(u,z) &= z + 2uzP(u,z) + zP(u,z)^2 \\
P(u,z) &= \frac{1-2uz-\sqrt{1-4uz+4u^2z^2-4z^2}}{2z}
\end{aligned}
$$

By setting $u=1$ we have the same total number of trees:

$$
\begin{aligned}
P(1,z) &= \frac{1-2z-\sqrt{1-4z}}{2z} \\
zP(1,z) &= \frac{1-2z-\sqrt{1-4z}}{2} \\
[z^n]\,(zP(1,z)) &= \frac{1}{2}\left([n=0]-2[n=1]+\frac{1}{2n-1}\binom{2n}{n}\right)
\end{aligned}
$$

The cumulative generating function of $P(u,z)$ is the partial derivative of $P(u,z)$ on u, evaluated at $u=1$.

$$
\begin{aligned}
P(u,z) &= \frac{1-2uz-\sqrt{1-4uz+4u^2z^2-4z^2}}{2z} \\
\frac{\partial}{\partial u}P(u,z) &= -1+\frac{1-2uz}{\sqrt{1-4uz+4u^2z^2-4z^2}} \\
P'(1,z) &= -1+\frac{1-2z}{\sqrt{1-4z}} \\
[z^n]P'(1,z) &= -[n=0]+\binom{2n}{n}-2\binom{2n-2}{n-1}
\end{aligned}
$$

Now we just have to divide the results. For $n>1$:

$$
\begin{aligned}
\mu &= \frac{[z^n]P'(1,z)}{[z^n]P(1,z)} \\
\mu &= \frac{\binom{2n}{n}-2\binom{2n-2}{n-1}}{\frac{1}{2(2(n+1)-1)}\binom{2(n+1)}{n+1}} \\
\mu &= \frac{(n+1)(n-1)}{2n-1}
\end{aligned}
$$

4 Week 2

4.1 Exercise 2.17

Solve the recurrence.

$A_N = A_{N-1} - \frac{2A_{N-1}}{N} + 2\left(1-\frac{2A_{N-1}}{N}\right)$ for $N>0$ with $A_0=0$.

This recurrence describes the following random process: A set of N elements collect into "2-nodes" and "3-nodes." At each step each 2-node is likely to turn into a

3-node with probability $2/N$ and each 3-node is likely to turn into two 2-nodes with probability $3/N$ What is the average number of 2-nodes after N steps?

Solution

We start by simplifying the recurrence:

$$A_N = A_{N-1} - \frac{2A_{N-1}}{n} + 2\Big(1 - \frac{2A_{N-1}}{n}\Big)$$
$$A_N = \frac{n-6}{n}A_{N-1} + 2$$

The summation factor suggested by multiplying the coefficients is:

$$S = \frac{(n-6)}{n}\frac{(n-5)}{(n-1)}\frac{(n-4)}{(n-2)}\cdots\frac{3}{9}\frac{2}{8}\frac{1}{7}$$
$$S = \frac{6\times 5\times 4\times 3\times 2\times 1}{n(n-1)(n-2)(n-3)(n-4)(n-5)}$$
$$S = \binom{n}{6}^{-1}$$

We can divide both sides by the summation factor:

$$A_N = \frac{n-6}{n}A_{N-1} + 2$$
$$\binom{n}{6}A_N = \frac{n-6}{n}\binom{n}{6}A_{N-1} + 2\binom{n}{6}$$

Using the absorption:

$$(r-k)\binom{r}{k} = r\binom{r-1}{k}$$
$$\frac{n-6}{n}\binom{n}{6} = \frac{n}{n}\binom{n-1}{6}$$
$$\frac{n-6}{n}\binom{n}{6} = \binom{n-1}{6}$$

Let's define $B_N = \binom{n}{6}A_N$:

$$
\begin{aligned}
\binom{n}{6} A_N &= \frac{n-6}{n}\binom{n}{6} A_{N-1} + 2\binom{n}{6} \\
\binom{n}{6} A_N &= \binom{n-1}{6} A_{N-1} + 2\binom{n}{6} \\
B_N &= B_{N-1} + 2\binom{n}{6}
\end{aligned}
$$

This last expression is valid for every $N > 6$:

$$
\begin{aligned}
B_N &= B_{N-1} + 2\binom{n}{6} \\
B_N &= B_6 + \sum_{7 \le k \le n} 2\binom{k}{6} \\
B_N &= B_6 + 2\sum_{7 \le k \le n} \binom{k}{6}
\end{aligned}
$$

We can use this general property of binomials:

$$
\begin{aligned}
\sum_{0 \le k \le N} \binom{k}{m} &= \binom{n+1}{m+1} \\
\sum_{7 \le k \le N} \binom{k}{6} &= -1 + \sum_{6 \le k \le N} \binom{k}{6} \\
\sum_{7 \le k \le N} \binom{k}{6} &= -1 + \sum_{0 \le k \le N} \binom{k}{6} \\
\sum_{7 \le k \le N} \binom{k}{6} &= -1 + \binom{n+1}{7}
\end{aligned}
$$

Substituting back:

$$
\begin{aligned}
B_N &= B_6 + 2\sum_{7\le k\le n}\binom{k}{6}\\
B_N &= 2 + 2\left(-1 + \binom{n+1}{7}\right)\\
B_N &= 2\binom{n+1}{7}\\
\binom{n}{6}A_N &= 2\binom{n+1}{7}\\
A_N &= 2\frac{(n+1)!}{7!(n-6)!}\frac{6!(n-6)!}{n!}\\
A_N &= \frac{2(n+1)}{7}
\end{aligned}
$$

The solution for all $N > 0$ therefore is:

$$
\begin{cases}
A_0, A_1, A_2, A_3, A_4, A_5, A_6 = 0, 2, -2, 4, 0, 2, 2 & N \le 6\\
A_N = \frac{2(n+1)}{7} & N > 6
\end{cases}
$$

4.2 Exercise 2.69

Plot the periodic part of the solution to the recurrence $a_N = 3a_{\lfloor N/3\rfloor} + N$ for $N > 3$ with $a_1 = a_2 = a_3 = 1$ for $1 \le N \le 972$.

Solution

Let's find the non-periodic part. This can be found by trying to find the exact solution when $N = 3^k$:

$$
\begin{aligned}
a_N &= 3a_{\lfloor N/3\rfloor} + n\\
a_{3^k} &= 3a_{\lfloor 3^k/3\rfloor} + 3^k\\
a_{3^k} &= 3a_{3^{k-1}} + 3^k
\end{aligned}
$$

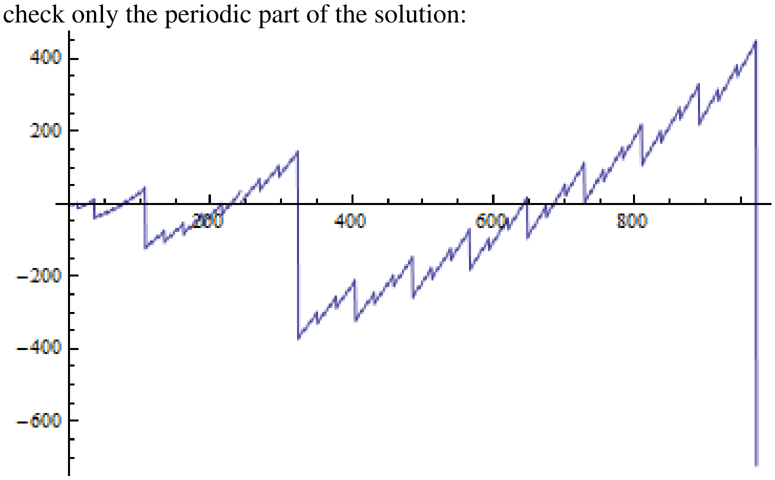
This last equation works for $k > 1$. We can now define $b_k = a_{3^k}$, where $b_2 = a_{3^2} = a_9 = 12$ and:

$$
\begin{aligned}
a_{3^k} &= 3a_{3^{k-1}} + 3^k \\
b_k &= 3b_{k-1} + 3^k \\
3^{-k} \times \big(b_k &= 3b_{k-1} + 3^k\big) \\
3^{-k}b_k &= 3^{-k} \times 3b_{k-1} + 3^{-k} \times 3^k \\
3^{-k}b_k &= 3^{-k+1}b_{k-1} + 1 \\
3^{-k}b_k &= 3^{-(k-1)}b_{k-1} + 1
\end{aligned}
$$

We further define $c_k = 3^{-k}b_k$, where $c_2 = 3^{-2}b_2 = 4/3$ and:

$$
\begin{aligned}
3^{-k}b_k &= 3^{-(k-1)}b_{k-1} + 1 \\
c_k &= c_{k-1} + 1 \\
c_k &= c_2 + \sum_{3 \le i \le k} 1 \\
c_k &= \frac{4}{3} + k - 2 \\
c_k &= k - \frac{2}{3} \\
3^{-k}b_k &= k - \frac{2}{3} \\
b_k &= (k - \frac{2}{3})3^k \\
a_{3^k} &= (k - \frac{2}{3})3^k \\
a_n &= n(-\frac{2}{3} + \log_3 n)
\end{aligned}
$$

This last equation is only exact for $n = 3^k$ and $k > 1$. We can plot the difference to check only the periodic part of the solution:



This graph cross zero when n is a power of 3.

4.3 Exercise 3.20

Solve the recurrence $a_n = 3a_{n-1} - 3a_{n-2} + a_{n-3}$ for $n > 2$ with $a_0 = a_1 = 0$ and $a_2 = 1$. Solve the same recurrence with the initial condition on a_1 changed to $a_1 = 1$.

Solution

For the first case:

$$
\begin{aligned}
a_n &= 3a_{n-1} - 3a_{n-2} + a_{n-3} + [n = 2] \\
\sum_n a_n z^n &= \sum_n z^n \left(3a_{n-1} - 3a_{n-2} + a_{n-3} + [n = 2]\right) \\
A(z) &= 3zA(z) - 3z^2 A(z) + z^3 A(z) + z^2 \\
A(z) &= \frac{z^2}{1 - 3z + 3z^2 - z^3} \\
A(z) &= \frac{z^2}{(1-z)^3} \\
z^{-2} A(z) &= \frac{1}{(1-z)^3} \\
z^{-2} A(z) &= \sum_n \binom{n+2}{2} z^n \\
a(n+2) &= \frac{(n+2)(n+1)}{2} \\
a(n) &= \frac{n(n-1)}{2}
\end{aligned}
$$

For the second case:

$$
\begin{aligned}
a_n &= 3a_{n-1} - 3a_{n-2} + a_{n-3} + [n=1] - 2[n=2] \\
\sum_n a_n z^n &= \sum_n z^n \left(3a_{n-1} - 3a_{n-2} + a_{n-3} + [n=1] - 2[n=2]\right) \\
A(z) &= 3zA(z) - 3z^2 A(z) + z^3 A(z) + z - 2z^2 \\
A(z) &= \frac{z - 2z^2}{1 - 3z + 3z^2 - z^3} \\
A(z) &= \frac{z - 2z^2}{(1-z)^3} \\
A(z) &= -\frac{1}{(1-z)^3} + \frac{3}{(1-z)^2} - \frac{2}{1-z} \\
a(n) &= -\binom{n+2}{2} + 3\binom{n+1}{1} - 2 \\
a(n) &= -\frac{(n+2)(n+1)}{2} + 3(n+1) - 2 \\
a(n) &= \frac{-n^2 - 3n - 2 + 6n + 6 - 4}{2} \\
a(n) &= \frac{-n^2 + 3n}{2} \\
a(n) &= \frac{n(3-n)}{2}
\end{aligned}
$$

4.4 Exercise 3.28

Find an expression for $[z^n]\frac{1}{\sqrt{1-z}} \ln \frac{1}{1-z}$. Hint: Expand $(1-z)^{-\alpha}$ and differentiate with respect to α.

Solution

Let's start by differentiating $(1-z)^{-\alpha}$:

$$
\begin{aligned}
f(z) &= (1-z)^{-\alpha} \\
\frac{d}{d\alpha} f(z) &= -(1-z)^{-\alpha} \ln(1-z) \\
\frac{d}{d\alpha} f(z) &= \frac{1}{(1-z)^{\alpha}} \ln \frac{1}{1-z}
\end{aligned}
$$

The function we want can be found by setting $\alpha = 1/2$. Now let's open the original function in a series, and differentiate it:

$$
\begin{aligned}
f(z) &= (1-z)^{-\alpha} \\
f(z) &= \sum_n \binom{\alpha+n-1}{n} z^n \\
\frac{d}{d\alpha} f(z) &= \sum_n \left(\frac{d}{d\alpha} \binom{\alpha+n-1}{n} \right) z^n
\end{aligned}
$$

The derivative of the binomial is:

$$
\begin{aligned}
\frac{d}{dk}\binom{k}{n} &= \binom{k}{n} \sum_{0 \le i \le n-1} \frac{1}{k-i} \\
\frac{d}{d\alpha}\binom{\alpha+n-1}{n} &= \binom{\alpha+n-1}{n} \sum_{0 \le i \le n-1} \frac{1}{\alpha+n-1-i}
\end{aligned}
$$

Substituting everything for $\alpha = 1/2$:

$$
\begin{aligned}
\frac{1}{\sqrt{1-z}} \ln \frac{1}{1-z} &= \sum_n \binom{1/2+n-1}{n} \left(\sum_{0 \le i \le n-1} \frac{1}{1/2+n-1-i} \right) z^n \\
\frac{1}{\sqrt{1-z}} \ln \frac{1}{1-z} &= 2\sum_n \binom{1/2+n-1}{n} \left(\sum_{0 \le i \le n-1} \frac{1}{2n-2i-1} \right) z^n \\
\frac{1}{\sqrt{1-z}} \ln \frac{1}{1-z} &= 2\sum_n \binom{n-1/2}{n} (H_{2n} - \frac{H_n}{2}) z^n \\
\frac{1}{\sqrt{1-z}} \ln \frac{1}{1-z} &= \sum_n \frac{1}{4^n}\binom{2n}{n} (2H_{2n} - H_n) z^n \\
[z^n]\frac{1}{\sqrt{1-z}} \ln \frac{1}{1-z} &= \frac{1}{4^n}\binom{2n}{n} (2H_{2n} - H_n)
\end{aligned}
$$

5 Week 1

5.1 Exercise 1.14

Follow through the steps above to solve the recurrence

$$
A_N = 1 + \frac{2}{N} \sum_{1 \le j \le N} A_{j-1} \quad \text{for} \quad N > 0.
$$

Solution
Let's find A_{N-1} and isolate the sum:

$$A_{N-1} = 1 + \frac{2}{N-1} \sum_{1 \le j \le N-1} A_{j-1}$$
$$\sum_{1 \le j \le N-1} A_{j-1} = \frac{N-1}{2}(A_{N-1} - 1)$$

We can substitute it on the original equation:

$$\begin{aligned}
A_N &= 1 + \frac{2}{N}\left(A_{N-1} + \frac{N-1}{2}(A_{N-1} - 1)\right) \\
A_N &= 1 + \frac{2}{N}\left(A_{N-1}\frac{N+1}{2} - \frac{N-1}{2}\right) \\
A_N &= 1 + \left(A_{N-1}\frac{N+1}{N} - \frac{N-1}{N}\right) \\
A_N &= A_{N-1}\frac{N+1}{N} + \frac{1}{N} \\
\frac{A_N}{N+1} &= \frac{A_{N-1}}{N} + \frac{1}{N(N+1)}
\end{aligned}$$

We can define $B_N = A_N/(N+1)$, then the equation becomes:

$$\begin{aligned}
B_N &= B_{N-1} + \frac{1}{N(N+1)} \\
B_N &= \sum_{1 \le j \le N} \frac{1}{j(j+1)}
\end{aligned}$$

This is a telescopic sum:

$$\begin{aligned}
B_N &= \sum_{1 \le j \le N} \frac{1}{j(j+1)} \\
B_N &= \sum_{1 \le j \le N} \frac{1}{j} - \frac{1}{(j+1)} \\
B_N &= \frac{1}{1} - \frac{1}{N+1} \\
B_N &= 1 - \frac{1}{N+1}
\end{aligned}$$

Substituting back:

$$
\begin{aligned}
B_N &= 1 - \frac{1}{N+1} \\
\frac{A_N}{N+1} &= 1 - \frac{1}{N+1} \\
A_N &= N + 1 - \frac{N+1}{N+1} \\
A_N &= N + 1 - 1 \\
A_N &= N
\end{aligned}
$$

5.2 Exercise 1.15

Show that the average number of exchanges used during the first partitioning stage (before the pointers cross) is $(N-2)/6$. (Thus, by linearity of the recurrences, the average number of exchanges used by quicksort is $1/6C_N - 1/2A_N$.)

Solution

There are $N!$ different possible vectors of size N to be sorted. For any given k, with $1 \leq k \leq N$, there's a $1/N$ probability that the first element will end up at the kth position after sorting.

When the sorting is finished, the first $k-1$ elements are smaller than the kth one, and the $N-k$ elements at the end are greater than the kth one. If j exchanges were made, then there are j elements on the first group that were originally in the second one, and vice versa. Therefore, the number of possible exchanges is $\binom{k-1}{j}\binom{N-k}{j}$.

Now we have to take order into account. Before the exchanges were made, the elements in the first group could have been arranged into $(k-1)!$ orderings, and the second group could have been in $(N-k)!$ orderings.

Since there are a total of $(N-1)!$ possible orderings of the whole vector after the first element, the probability of a vector with size N having the first element ending up as the kth position after j exchanges is:

$$
P(k,j) = \frac{1}{N}\binom{N-k}{j}\binom{k-1}{j}\frac{(N-k)!(k-1)!}{(N-1)!}
$$

To get the desired expected value, we multiply by j and sum over all pairs k, j:

$$
\begin{aligned}
E_N &= \sum_{1 \leq k \leq N} \sum_{0 \leq j \leq k-1} j \frac{1}{N}\binom{N-k}{j}\binom{k-1}{j}\frac{(N-k)!(k-1)!}{(N-1)!} \\
&= \frac{1}{N} \sum_{1 \leq k \leq N} \frac{(N-k)!(k-1)!}{(N-1)!} \sum_{0 \leq j \leq k-1} j \binom{k-1}{j}\binom{N-k}{j}
\end{aligned}
$$

Let's use the general absorption property of binomials, valid for all integers a:

$$a\binom{b}{a} = b\binom{b-1}{a-1}$$
$$j\binom{k-1}{j} = (k-1)\binom{k-2}{j-1}$$

We can also extend the domain on the second summation to all j, because the terms will be zero outside the range $0 \le j \le k-1$:

$$\begin{aligned}
E_N &= \frac{1}{N}\sum_{1\le k\le N}\frac{(N-k)!(k-1)!}{(N-1)!}\sum_{0\le j\le k-1} j\binom{k-1}{j}\binom{N-k}{j} \\
&= \frac{1}{N}\sum_{1\le k\le N}\frac{(N-k)!(k-1)!}{(N-1)!}\sum_{j}(k-1)\binom{k-2}{j-1}\binom{N-k}{j} \\
&= \frac{1}{N}\sum_{1\le k\le N}\frac{(N-k)!(k-1)!}{(N-1)!}(k-1)\sum_{j}\binom{k-2}{j-1}\binom{N-k}{j}
\end{aligned}$$

Now we can use this version of Vandermonde's convolution, valid when a, c, e are integers and $a \ge 0$:

$$\sum_{a}\binom{b}{c+a}\binom{d}{e+a} = \binom{b+d}{b-c+e}$$
$$\sum_{j}\binom{N-k}{j}\binom{k-2}{j-1} = \binom{N-2}{N-k-1}$$

Applying it to our equation:

$$
\begin{aligned}
E_N &= \frac{1}{N} \sum_{1 \le k \le N} \frac{(N-k)!(k-1)!}{(N-1)!}(k-1) \sum_j \binom{k-2}{j-1}\binom{N-k}{j} \\
&= \frac{1}{N} \sum_{1 \le k \le N} \frac{(N-k)!(k-1)!}{(N-1)!}(k-1) \binom{N-2}{N-k-1} \\
&= \frac{1}{N} \sum_{1 \le k \le N} \frac{(N-k)!(k-1)!}{(N-1)!}(k-1) \frac{(N-2)!}{(N-k-1)!(k-1)!} \\
&= \frac{1}{N} \sum_{1 \le k \le N} (k-1) \frac{(N-k)!}{(N-k-1)!} \frac{(N-2)!}{(N-1)!} \frac{(k-1)!}{(k-1)!} \\
&= \frac{1}{N} \sum_{1 \le k \le N} \frac{(k-1)(N-k)}{(N-1)} \\
&= \frac{1}{N(N-1)} \sum_{1 \le k \le N} -k^2 + (1+N)k - N \\
&= \frac{1}{N(N-1)} \left(- \left(\sum_{1 \le k \le N} k^2 \right) + (1+N) \left(\sum_{1 \le k \le N} k \right) + \left(\sum_{1 \le k \le N} N \right) \right) \\
&= \frac{1}{N(N-1)} \left(- \frac{N(N+1)(2N+1)}{6} + (1+N)\frac{N(N+1)}{2} - N^2 \right) \\
&= \frac{1}{N(N-1)} \frac{N^3 - 3N^2 + 2N}{6} \\
&= \frac{1}{N(N-1)} \frac{N(N-1)(N-2)}{6N(N-1)} \\
&= \frac{N-2}{6}
\end{aligned}
$$

5.3 Exercise 1.16

If we change the first line in the quicksort implementation above to call insertion sort when $hi - lo \le M$ then the total number of comparisons to sort N elements is described by the recurrence

$$
C_N = \begin{cases} N + 1 + \frac{1}{N} \sum_{1 \le j \le N} (C_{j-1} + C_{N-j}) & N > M \\ \frac{1}{4} N(N-1) & N \le M \end{cases}
$$

Solve this recurrence.

Solution

Let's massage the equation for $N > M$:

$$
\begin{aligned}
C_N &= N + 1 + \frac{1}{N} \sum_{1 \le j \le N} (C_{j-1} + C_{N-j}) \\
&= N + 1 + \frac{1}{N} \left(\sum_{1 \le j \le N} C_{j-1} + \sum_{1 \le j \le N} C_{N-j} \right)
\end{aligned}
$$

We can introduce $k = N - j$, so $j = N - k$ and change the limits on the second summation:

$$
\begin{aligned}
C_N &= N + 1 + \frac{1}{N} \left(\sum_{1 \le j \le N} C_{j-1} + \sum_{1 \le j \le N} C_{N-j} \right) \\
&= N + 1 + \frac{1}{N} \left(\sum_{0 \le j < N} C_j + \sum_{1 \le N-k \le N} C_k \right) \\
&= N + 1 + \frac{1}{N} \left(\sum_{0 \le j < N} C_j + \sum_{0 \le k < N} C_k \right) \\
&= N + 1 + \frac{2}{N} \sum_{0 \le j < N} C_j
\end{aligned}
$$

Let's find the equation for C_{N-1} and isolate the summation. Notice that we must assume that $N > M + 1$ in order to ensure this equation work for C_{N-1}.

$$
\begin{aligned}
C_{N-1} &= (N-1) + 1 + \frac{2}{N-1} \sum_{0 \le j < N-1} C_j \\
C_{N-1} &= N + \frac{2}{N-1} \sum_{0 \le j < N-1} C_j \\
\sum_{0 \le j < N-1} C_j &= \frac{N-1}{2} (C_{N-1} - N)
\end{aligned}
$$

Substituting back in the original equation:

$$
\begin{aligned}
C_N &= N + 1 + \frac{2}{N} \sum_{0 \le j < N} C_j \\
C_N &= N + 1 + \frac{2}{N} \left(C_{N-1} + \frac{N-1}{2} (C_{N-1} - N) \right) \\
C_N &= C_{N-1} \frac{N+1}{N} + 2 \\
\frac{C_N}{N+1} &= \frac{C_{N-1}}{N} + \frac{2}{N+1}
\end{aligned}
$$

Defining $B_N = \frac{C_N}{N+1}$:

$$
\begin{aligned}
\frac{C_N}{N+1} &= \frac{C_{N-1}}{N} + \frac{2}{N+1} \\
B_N &= B_{N-1} + \frac{2}{N+1} \\
B_N &= B_{M+1} + \sum_{M+3 \le k \le N+1} \frac{2}{k}
\end{aligned}
$$

Remember this will only work for $N > M + 1$. We need to find B_{M+1} in order to continue.

$$
\begin{aligned}
B_{M+1} &= \frac{1}{M+2} \left((M+1) + 1 + \frac{2}{M+1} \sum_{0 \le j \le M} C_j \right) \\
B_{M+1} &= \frac{M+2}{M+2} + \frac{2}{(M+1)(M+2)} \sum_{0 \le j \le M} \frac{j(j-1)}{4} \\
B_{M+1} &= 1 + \frac{2}{4(M+1)(M+2)} \sum_{0 \le j \le M} j^2 - j \\
B_{M+1} &= 1 + \frac{2}{4(M+1)(M+2)} \left(\frac{M(M+1)(2M+1)}{6} - \frac{M(M+1)}{2} \right) \\
B_{M+1} &= 1 + \frac{1}{2(M+1)(M+2)} \left(\frac{M(M+1)(2M-2)}{6} \right) \\
B_{M+1} &= 1 + \frac{M(M-1)}{6(M+2)}
\end{aligned}
$$

We can get back to C_N now:

$$
\begin{aligned}
B_N &= B_{M+1} + \sum_{M+3 \le k \le N+1} \frac{2}{k} \\
B_N &= 1 + \frac{M(M-1)}{6(M+2)} + 2 \sum_{M+3 \le k \le N+1} \frac{1}{k} \\
\frac{C_N}{N+1} &= 1 + \frac{M(M-1)}{6(M+2)} + 2\left(H_{N+1} - H_{M+2}\right)
\end{aligned}
$$

So the final solution is:

$$
C_N = \begin{cases} (N+1)\left(1 + \frac{M(M-1)}{6(M+2)} + 2(H_{N+1} - H_{M+2})\right) & N > M+1 \\ M + 2 + \frac{M(M-1)}{6} & N = M+1 \\ \frac{1}{4}N(N-1) & N \le M \end{cases}
$$

5.4 Exercise 1.17

Ignoring small terms (those significantly less than N) in the answer to the previous exercise, find a function $f(M)$ so that the number of comparisons is approximately $2N \ln N + f(M)N$. Plot the function $f(M)$, and find the value of M that minimizes the function.

Solution

Let's use the following asymptotic for the harmonic number:

$$
\begin{aligned}
H_N &= \ln N + \gamma + O(N^{-1}) \\
H_N &\sim \ln N + \gamma
\end{aligned}
$$

Let's find C_{N-1}:

$$
\begin{aligned}
C_{N-1} &= N\left(1 + \frac{M(M-1)}{6(M+2)} + 2(H_N - H_{M+2})\right) \\
&\sim N\left(1 + \frac{M(M-1)}{6(M+2)} + 2(\ln N + \gamma - \ln(M+2) - \gamma)\right) \\
&\sim 2N \ln N + N\left(1 + \frac{M(M-1)}{6(M+2)} - 2\ln(M+2)\right)
\end{aligned}
$$

Therefore:

$$
f(M) = 1 + \frac{M(M-1)}{6(M+2)} - 2\ln(M+2)
$$

Let's find its derivative:

$$
\begin{aligned}
f(M) &= 1 + \frac{M(M-1)}{6(M+2)} - 2\ln(M+2) \\
f'(M) &= \frac{((M-1)+M)(6(M+2)) - 6M(M(M-1))}{36(M+2)^2} - \frac{2}{M+2} \\
f'(M) &= \frac{M^2+4M-2}{6(M+2)^2} - \frac{2}{M+2} \\
f'(M) &= \frac{M^2-8M-26}{6(M+2)^2}
\end{aligned}
$$

The minimum for M happens when the derivative is zero:

$$
\begin{aligned}
\frac{M^2-8M-26}{6(M+2)^2} &= 0 \\
M^2-8M-26 &= 0 \\
M &= \frac{8 \pm \sqrt{(-8)\times(-8) - 4\times(-26)}}{2} \\
M &= 4 \pm \sqrt{42}
\end{aligned}
$$

The positive root is $4+\sqrt{42} \sim 10.48$ so we'll test the values for 10 and 11:

$$
\begin{aligned}
f(10) &= -2.719813 \\
f(11) &= -2.719642
\end{aligned}
$$

Therefore the best value for M is $M = 10$.