Appendix

Appendix A: Solving the master equation

We want to solve

$$f(1,k+1) = f(1,k) + \alpha - K(k)f(1,k) \tag{1}$$

$$f(i,k+1) = f(i,k) + K(k)(i-1)f(i-1,k) - K(k)if(i,k), \tag{2}$$

knowing

$$K(k) = \frac{1-\alpha}{k}. \tag{3}$$

Let us assume

$$\frac{f(i,k+1)}{f(i,k)} = \frac{k+1}{k} \tag{4}$$

for all $k > 1$ and $1 \leq i \leq k$. This last assumption induces that

$$\frac{f(i,k+1)}{f(i,k)} = \frac{k+1}{k} = \frac{f(i-1,k+1)}{f(i-1,k)}$$

so that in particular, for $k > 1$ and $2 \leq i \leq k$.

$$\frac{f(i,k+1)}{f(i-1,k+1)} = \frac{f(i,k)}{f(i-1,k)}$$

and thus this ratio does not depend on k. We write

$$\beta(i) = \frac{f(i,k+1)}{f(i-1,k+1)}. \tag{5}$$

Using the relations 3 and 4 in Equation 1 yields

$$\frac{k+1}{k}f(1,k) = f(1,k) + \alpha - \frac{1-\alpha}{k}f(1,k)$$

so that we find

$$f(1,k) = \frac{\alpha k}{2-\alpha} \tag{6}$$

Let us now turn to Equation 2. Using the relations 3, 4 and 5 yields

$$\frac{k+1}{k}f(i,k) = f(i,k) + \frac{1-\alpha}{k}(i-1)\frac{f(i,k)}{\beta(i)} - \frac{1-\alpha}{k}if(i,k).$$

Let us rework this expression after cancelling the common factor $f(i,k)$:

$$\begin{aligned}\frac{k+1}{k} &= 1 + \frac{1-\alpha}{k}(i-1)\frac{1}{\beta(i)} - \frac{1-\alpha}{k}i \\ &= 1 + \frac{1-\alpha}{k}\left[\frac{i-1-i\beta(i)}{\beta(i)}\right]\end{aligned}$$

or

$$\begin{aligned}\frac{k+1}{k} - 1 &= \frac{1-\alpha}{k}\left[\frac{i-1-i\beta(i)}{\beta(i)}\right] \\ \frac{1}{k} &= \frac{1-\alpha}{k}\left[\frac{i-1-i\beta(i)}{\beta(i)}\right] \\ 1 &= \frac{(1-\alpha)(i-1)-(1-\alpha)i\beta(i)}{\beta(i)}\end{aligned}$$

and thus we find

$$\beta(i) = \frac{(1-\alpha)(i-1)}{1+(1-\alpha)i}.$$

Introducing

$$\rho = \frac{1}{1-\alpha}$$

the last expression can be simplified to

$$\beta(i) = \frac{(1-\alpha)(i-1)}{1+(1-\alpha)i} = \frac{i-1}{\rho(1+\frac{i}{\rho})} = \frac{i-1}{\rho+i}. \tag{7}$$

Let us go back to 6, the expression can also be written as

$$\begin{aligned}f(1,k) &= \frac{\alpha k}{2-\alpha} \\ &= \frac{\alpha k}{1-\alpha+1} \\ &= \frac{\alpha k}{\frac{1}{\rho}+1} \\ &= \frac{\alpha k \rho}{\rho+1}\end{aligned}$$

and by definition of α, $\sum_{i=1}^{k} f(i,k) = \alpha k = n_k$ is the total number of different words in the k first words so we can write

$$f(1,k) = \frac{\rho n_k}{\rho+1}. \tag{8}$$

Developing 5

$$f(i,k) = \beta(i)f(i-1,k) = \beta(i)\beta(i-1)f(i-2,k) = \cdots = \beta(i)\beta(i-1)\ldots\beta(2)f(1,k)$$

and using 7 and 8 gives

$$
\begin{aligned}
f(i,k) &= \frac{i-1}{\rho+i}\frac{i-2}{\rho+i-1}\cdots\frac{1}{\rho+2}\frac{n_k\rho}{1+\rho} \\
&= \frac{(i-1)(i-2)\dots 2.1}{(\rho+i)(\rho+i-1)\dots(\rho+2)(\rho+1)}n_k\rho.
\end{aligned}
$$

This last expression can first be written using the gamma function $\Gamma(n+1) = n!$:

$$f(i,k) = \frac{\Gamma(i)\Gamma(\rho+1)}{\Gamma(i+\rho+1)}\rho n_k$$

which itself can be written using the beta function $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ to give

$$f(i,k) = B(i,\rho+1)\rho n_k. \tag{9}$$

Let us verify that it is indeed a solution of 2. Substituting 9 in 2 we find on the right hand side

$$B(i,\rho+1)\rho n_k + \frac{1-\alpha}{k}\left[(i-1)B(i-1,\rho+1)\rho n_k - iB(i,\rho+1)\rho n_k\right]. \tag{10}$$

Now note that

$$
\begin{aligned}
(i-1)B(i-1,\rho+1) &= \frac{(i-1)\Gamma(i-1)\Gamma(\rho+1)}{\Gamma(i+\rho)} \\
&= \frac{\Gamma(i)\Gamma(\rho+1)}{\Gamma(i+\rho)} \\
&= \frac{\Gamma(i)\Gamma(\rho+1)(i+\rho)}{\Gamma(i+\rho)(i+\rho)} \\
&= \frac{\Gamma(i)\Gamma(\rho+1)(i+\rho)}{\Gamma(i+\rho+1)} \\
&= (i+\rho)B(i,\rho+1),
\end{aligned}
$$

so expression 10 can be written as

$$B(i,\rho+1)\rho n_k + \frac{1-\alpha}{k}\left[(i+\rho)B(i,\rho+1)\rho n_k - iB(i,\rho+1)\rho n_k\right]$$

or

$$B(i,\rho+1)\rho\left(\alpha k + (1-\alpha)\rho\alpha\right) = B(i,\rho+1)\rho\alpha(k+1). \tag{11}$$

The last expression is exactly the left hand side of 2 since

$$f(i,k+1) = B(i,\rho+1)\rho n_{k+1} = B(i,\rho+1)\rho\alpha(k+1).$$

Now the particularity of the beta function $B(i,\rho+1)$ is that, for $i\to\infty$, it behaves as $i^{-(\rho+1)}$ and thus $f(i,k)$ follows a power law in the tail with exponent $\rho+1$: indeed, using Stirling's formula $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ we can write

$$\begin{aligned}
B(i,\rho+1) &= \frac{\Gamma(i)\Gamma(\rho+1)}{\Gamma(i+\rho+1)} \\
&\sim \frac{\sqrt{2\pi(i-1)}\left(\frac{i-1}{e}\right)^{i-1}\Gamma(\rho+1)}{\sqrt{2\pi(i+\rho)}\left(\frac{i+\rho}{e}\right)^{i+\rho}} \\
&\sim \frac{\sqrt{(i-1)}}{\sqrt{(i+\rho)}}\frac{(i-1)^{i-1}}{(i+\rho)^{i+\rho}}e^{\rho+1}\Gamma(\rho+1) \\
&\sim i^{-(\rho+1)}e^{\rho+1}\Gamma(\rho+1) \qquad \text{when } i \to \infty.
\end{aligned}$$