Consistency of `S -> S S`

JC Chappelier

20000629

Consider the SCFG $\mathcal{G} = (\{a\}, \{S\}, S, \{S \longrightarrow S\ S \quad (p), S \longrightarrow a \quad (1-p)\})$, with $0 < p < 1$.

How much is $P\left(\mathcal{L}\left(\mathcal{G}\right)\right)$?

$\mathcal{L}(\mathcal{G}) = \{a^n : n \geq 1\}$, thus $P\left(\mathcal{L}\left(\mathcal{G}\right)\right) = \sum_{n=1}^{\infty} P(a^n)$.

Each derviation of the sentence a^n has a probability $p^{n-1}(1-p)^n$: n as have been produced, thus the rule $S \longrightarrow a$ has been used n times, and from the initial S, the rule $S \longrightarrow S\ S$ had to be applied $n-1$ times (to produce the n final Ss, since each single application of that rule adds only and only one more S).

Thus $P(a^n) = \nu_n\ p^{n-1}\ (1-p)^n$, where ν_n is the number of derivations of the sentence a^n. We proove in appendix A that

$$\nu_n = \frac{1}{n}\binom{2n-2}{n-1} = C_{n-1}$$

where C_n is know as the n-th "Catalan number".

Thus:

$$\begin{aligned} P\left(\mathcal{L}\left(\mathcal{G}\right)\right) &= \sum_{n=1}^{\infty} C_{n-1}\ p^{n-1}\ (1-p)^n \\ &= (1-p)\ \sum_{n=0}^{\infty} C_n\ p^n\ (1-p)^n \\ &= (1-p)\ \mathrm{f}(p(1-p)) \end{aligned}$$

where f is the generating function of the Catalan numbres, i.e. $\mathrm{f}(z) = \sum_{n=0}^{\infty} C_n\ z^n$. We proove in appendix B that for all $0 \leq z \leq \frac{1}{4}$, $\mathrm{f}(z) = \frac{1-\sqrt{1-4z}}{2z}$.

Thus (see appendix C) :

$$\mathrm{f}(p(1-p)) = \begin{cases} \frac{1}{p} & \text{for } \frac{1}{2} \leq p < 1 \\ \frac{1}{1-p} & \text{for } 0 < p \leq \frac{1}{2} \end{cases}$$

And finally:

$$P\left(\mathcal{L}\left(\mathcal{G}\right)\right)) = \begin{cases} \frac{1-p}{p} & \text{for } \frac{1}{2} \leq p < 1 \\ 1 & \text{for } 0 < p \leq \frac{1}{2} \end{cases}$$

i.e.:

$$P\left(\mathcal{L}\left(\mathcal{G}\right)\right) = \min\left(1, \frac{1-p}{p}\right)$$

A Catalan numbers

Let's compute ν_n is the number of derivations of the sentence a^n.

Looking at all decomposition into two subsequences of string a^n, we derive that

$$\nu_n = \sum_{k=1}^{n-1} \nu_k \cdot \nu_{n-k}$$

(and $\nu_1 = 1$).

Knowing that Catalan numbers follow the following recurrence relation:

$$\begin{cases} C_0 & = & 1 \\ C_{n+1} & = & \displaystyle\sum_{k=0}^{n} C_k \cdot C_{n-k} \end{cases}$$

it's trivial to see that indeed $\nu_n = C_{n-1}$.

There are many proofs which you can easily find on the Web that $C_n = \frac{1}{n+1}\binom{2n}{n}$.

B Generating fonction for Catalan numbers

Knowing that:

$$\sum_{n=0}^{\infty} \binom{2n}{n} z^n = \frac{1}{\sqrt{1-4z}}$$

the convergence radius of which is $\frac{1}{4}$, we get (integrating):

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{1}{n+1}\binom{2n}{n} z^{n+1} &= \int_0^z \frac{\mathrm{d}u}{\sqrt{1-4u}} \\ &= -\frac{1}{2}(\sqrt{1-4z}-1) \end{aligned}$$

thus

$$z \cdot \sum_{n=0}^{\infty} C_n z^n = \frac{1-\sqrt{1-4z}}{2}$$

And finaly

$$\sum_{n=0}^{\infty} C_n z^n = \frac{1 - \sqrt{1-4z}}{2z}$$

for all $0 < |z| < \frac{1}{4}$. This definition can also be extended to $z = 0$ and $|z| = \frac{1}{4}$ by continuity (continuous extension).

C Computation of $f(p(1-p))$

For $0 < p < 1$ (thus $p(1-p) \leq \frac{1}{4}$):

$$\begin{aligned}
f(p(1-p)) &= \frac{1 - \sqrt{1-4p(1-p)}}{2p(1-p)} \\
&= \frac{1 - \sqrt{1-4p+4p^2)}}{2p(1-p)} \\
&= \frac{1 - \sqrt{(2p-1)^2}}{2p(1-p)} \\
&= \frac{1 - |2p-1|}{2p(1-p)} \\
&= \begin{cases} \frac{2p-2}{2p(1-p)} & \text{for } \frac{1}{2} \leq p < 1 \\ \frac{2p}{2p(1-p)} & \text{for } 0 < p \leq \frac{1}{2} \end{cases} \qquad = \begin{cases} \frac{1}{p} & \text{for } \frac{1}{2} \leq p < 1 \\ \frac{1}{1-p} & \text{for } 0 < p \leq \frac{1}{2} \end{cases}
\end{aligned}$$