PLAYING m PATTERNS AGAINST EACH OTHER

Generating a pattern from scratch: Visualize a sequence of random and independent TRIALS, each generating one of several possible symbols (such as A, B and C, or H and T, etc.) with the to p_a, p_b, p_c, ..(one for each symbol; they must add up to 1).

One can than consider a specific pattern of consecutive symbols (such as ABCABC) and find the distribution of the number of trials required to 'build' the first occurrence of this pattern. We know that a PGF of this distribution, say $F(s)$, is equal to

$$\frac{1}{1+(1-s)\cdot Q(s)}$$

where the denominator of $Q(s)$ is a product of the individual probabilities of all symbols in the pattern, multiplied by the corresponding power of s (for ABCABC, this would be $p_a^2 p_b^2 p_c^2 s^6$), and the numerator has as many terms as the number of perfect matches one gets when sliding the pattern past itself (one symbol at a time), each equal to a product of the individual probabilities to complete the pattern from the current match on, multiplied by the corresponding power of s. For the ABCABC this would be 1 (corresponding to the perfect match of any pattern against itself before we start sliding - this term is *always* there) plus $p_a p_b p_c s^3$ corresponding to

```
ABCABC
   ABCABC
```

(this is the only other perfect match). The PGF of the number of trials needed to generate this pattern for the first time is thus

$$F(s)=\frac{1}{1+(1-s)\cdot\dfrac{1+p_a p_b p_c s^3}{p_a^2 p_b^2 p_c^2 s^6}} \tag{1}$$

The corresponding expected value is obtained from

$$F'(1)=\frac{1+p_a p_b p_c}{p_a^2 p_b^2 p_c^2}$$

Generating a pattern assuming that the first i of its symbols are already there: To play two or more of such patterns against each other (i.e. competing to see which of them occurs first), we also need the PGF of the number of the *remaining* trials needed to generate a pattern (for the first time) given that its first symbol (first 2 symbols, first 3 symbols,....) have been already observed; the corresponding PGFs will be denoted $F_{\mathrm{A}}(s)$, $F_{\mathrm{AB}}(s)$, $F_{\mathrm{ABC}}(s)$, ... The following example indicates how to get them all (including the original $F(s)$; one should verify that the new $F(s)$ matches the old (1), as a test of correctness): All we need to do is to set up (and then solve) the following set of linear equations (we will leave out the s argument, to simplify our notation):

$$\begin{aligned}
F &= s\left(p_a F_{\mathrm{A}} + p_b F + p_c F\right) \\
F_{\mathrm{A}} &= s\left(p_a F_{\mathrm{A}} + p_b F_{\mathrm{AB}} + p_c F\right) \\
F_{\mathrm{AB}} &= s\left(p_a F_{\mathrm{A}} + p_b F + p_c F_{\mathrm{ABC}}\right) \\
F_{\mathrm{ABC}} &= s\left(p_a F_{\mathrm{ABCA}} + p_b F + p_c F\right) \\
F_{\mathrm{ABCA}} &= s\left(p_a F_{\mathrm{A}} + p_b F_{\mathrm{ABCAB}} + p_c F\right) \\
F_{\mathrm{ABCAB}} &= s\left(p_a F_{\mathrm{A}} + p_b F + p_c\right)
\end{aligned}$$

To understand the logic of these, let's have a closer look at one of them, say $F_{\mathrm{AB}} = s\left(p_a F_{\mathrm{A}} + p_b F + p_c F_{\mathrm{ABC}}\right)$. Here we are assuming that AB have been just generated, and the next potential symbol is either A, B or C, according to their respective probabilities. Getting A results in ABA - we can utilize only the last symbol of this string, so we multiply p_a by F_{A}; getting B (i.e. ABB) forces us to start from scratch (thus p_b gets multiplied by F); getting C (i.e. ABC) yields the first 3 correct symbols of the whole pattern, so we add $p_c F_{\mathrm{ABC}}$.

We quote the result for only the last of the resulting PGFs, namely

$$F_{\mathrm{ABCAB}}(s) = \frac{p_c s\left(1 - s + p_a p_b p_c s^3 - p_a p_b p_c s^4 + p_a^2 p_b^2 p_c s^5\right)}{1 - s + p_a p_b p_c s^3 - p_a p_b p_c s^4 + p_a^2 p_b^2 p_c^2 s^6}$$

One can see that $F_{\mathrm{ABCAB}}(1) = 1$, as it must, and that the expected value of the number of trials to generate ABCABC for the fist time (assuming we start from ABCAB) is given by

$$F'_{\mathrm{ABCAB}}(1) = \frac{1 - p_c + p_a p_b p_c - p_a p_b p_c^2}{p_a^2 p_b^2 p_c^2}$$

Another example: Similarly, to generate the HTTHH pattern, the corresponding PGF is

$$F(s) = \frac{1}{1 + (1-s)\cdot\dfrac{1 + p^2 q^2 s^4}{p^3 q^2 s^5}}$$

having the expected value of

$$\mu = \frac{1 + p^2 q^2}{p^3 q^2}$$

(it is now more convenient to denote the probabilities of H and T by p and q respectively). The set of equations for the five Fs (from F to F_{HTTH}) now reads

$$\begin{aligned}
F &= s\left(p F_{\mathrm{H}} + q F\right) \\
F_{\mathrm{H}} &= s\left(p F_{\mathrm{H}} + q F_{\mathrm{HT}}\right) \\
F_{\mathrm{HT}} &= s\left(p F_{\mathrm{H}} + q F_{\mathrm{HTT}}\right) \\
F_{\mathrm{HTT}} &= s\left(p F_{\mathrm{HTTH}} + q F\right) \\
F_{\mathrm{HTTH}} &= s\left(p + q F_{\mathrm{HT}}\right)
\end{aligned}$$

resulting in (we quote only one of them, as an example):

$$F_{\mathrm{HT}}(s) = \frac{p^2qs^3\left(1-s+pqs^2\right)}{1-s+p^2q^2s^4-p^2q^3s^5}$$

with the expected value of

$$\mu_{\mathrm{HT}} = \frac{1-p^3q}{p^3q^2}$$

Two patterns competing: Suppose now that we want to play the above pattern (calling it Pattern 1) against another one, say THHHT (Pattern 2). Note that these don't need to be of the same length; the only requirement is the one is not a substring of the other. First, we have to find the corresponding set of PGFs for the second pattern (using $\hat{F}$ instead of F to differentiate between the two), getting

$$\begin{aligned}
\hat{F} &= s\left(p\hat{F}+q\hat{F}_{\mathrm{T}}\right) \\
\hat{F}_{\mathrm{T}} &= s\left(p\hat{F}_{\mathrm{TH}}+q\hat{F}_{\mathrm{T}}\right) \\
\hat{F}_{\mathrm{TH}} &= s\left(p\hat{F}_{\mathrm{THH}}+q\hat{F}_{\mathrm{T}}\right) \\
\hat{F}_{\mathrm{THH}} &= s\left(p\hat{F}_{\mathrm{THHH}}+q\hat{F}_{\mathrm{T}}\right) \\
\hat{F}_{\mathrm{THHH}} &= s\left(p\hat{F}+q\right)
\end{aligned}$$

This time we quote $\hat{F}$ and $\hat{F}_{\mathrm{THH}}$, each with its expected value:

$$\begin{aligned}
\hat{F}(s) &= \frac{1}{1+(1-s)\cdot\dfrac{1+p^3qs^4}{p^3q^2s^5}} \\
\hat{\mu} &= \hat{F}'(1) = \frac{1+p^3q}{p^2q^2} \\
\hat{F}_{\mathrm{THH}}(s) &= \frac{pqs^2\left(1-s+p^2qs^3\right)}{1-s+p^3qs^4-p^4qs^5} \\
\hat{\mu}_{\mathrm{THH}} &= \hat{F}'_{\mathrm{THH}}(1) = \frac{1-pq+p^3q}{p^3q^2}
\end{aligned}$$

At this point, we introduce a new notation: $F_{i()}$ is the PGF to generate Pattern i from scratch (equal to F and $\hat{F}$ for our Pattern 1 and Pattern 2 respectively), while $F_{i()j}$ indicates the PGF to generate Pattern i given that Pattern j has just been completed, and as many of its symbols can be utilized to help generate Pattern i as possible (note that in our case $F_{1()2} = F_{\mathrm{HT}}$ and $F_{2()1} = \hat{F}_{\mathrm{THH}}$). For the corresponding expected values we use $\mu_i \equiv F'_{i()}(1)$ and $\mu_{i|j} \equiv F'_{i()j}(1)$..

It is well known (we state this without proof) that the generating function of the sequence of probabilities that Pattern i wins over Pattern j at the n^{th} trial is given by

$$F_{i(j)} = \frac{F_{i()} - F_{i()j} \cdot F_{j()}}{1 - F_{i()j} \cdot F_{j()i}} \tag{2}$$

and reverse (i.e. its $i \leftrightarrow j$ analog).

Probability of winning; game's expected duration: Evaluating the RHS at $s = 1$ yields the probability that Pattern i wins (is generated first) over Pattern j (in any number of trials); unfortunately, a simple substitution yields $\frac{0}{0}$ and we have to use L'Hospital rule instead, getting

$$P_{i(j)} = \frac{\mu_j - \mu_i + \mu_{i|j}}{\mu_{i|j} + \mu_{j|i}}$$

which involves only the four expected values computed earlier. In terms of our example, this yields

$$P_{1(2)} = \frac{\dfrac{1+p^3q}{p^2q^2} - \dfrac{1+p^2q^2}{p^3q^2} + \dfrac{1-p^3q}{p^3q^2}}{\dfrac{1-p^3q}{p^3q^2} + \dfrac{1-pq+p^3q}{p^3q^2}} = \frac{1-p^2q^2}{2-pq}$$

Note that when $p = q = \frac{1}{2}$, Pattern 1 wins with the probability of $\frac{15}{28} = 53.57\%$ (being far from a fair game).

The expected number of trials to finish such a game is equal to

$$M_{(ij)} = F'_{i(j)}(1) + F'_{j(i)}(1) = \frac{\mu_i \mu_{j|i} + \mu_j \mu_{i|j} - \mu_{i|j}\mu_{j|i}}{\mu_{i|j} + \mu_{j|i}}$$

In our case, this equals

$$M_{(12)} = \frac{1 + p^2q - p^3q^2 + p^5q^3}{p^3q^2(2-pq)}$$

(when $p = q = \frac{1}{2}$, this equals to $\frac{281}{14}$ trials - slightly over 20 flips of a coin).

Three patterns competing: Now, if we want to bring yet another pattern (say k) into this game. we first need to modify (2) to find the generating function of Pattern i winning over Pattern j under the assumption that Pattern k has just been generated (rather than starting from scratch). This modification requires putting k after each set of parentheses which are not followed by any index yet, namely

$$F_{i(j)k} = \frac{F_{i()k} - F_{i()j} \cdot F_{j()k}}{1 - F_{i()j} \cdot F_{j()i}}$$

Note that this implies

$$P_{i(j)k} = \frac{\mu_{j|k} - \mu_{i|k} + \mu_{i|j}}{\mu_{i|j} + \mu_{j|i}}$$

and

$$M_{(ij)k} = \frac{\mu_{i|k}\mu_{j|i} + \mu_{j|k}\mu_{i|j} - \mu_{i|j}\mu_{j|i}}{\mu_{i|j} + \mu_{j|i}}$$

for the expected duration of the game.

Having done this, we have to modify (2) again, this time inserting an extra k into each set of parentheses, thus:

$$F_{i(jk)} = \frac{F_{i(k)} - F_{i(k)j} \cdot F_{j(k)}}{1 - F_{i(k)j} \cdot F_{j(k)i}} = \frac{F_{i(j)} - F_{i(j)k} \cdot F_{k(j)}}{1 - F_{i(j)k} \cdot F_{k(j)i}}$$

(the last two expressions yield identical results), where $F_{i(jk)}$ stands for the generating function of Pattern i winning over *both* Patterns j and k at Trial n.

This implies

$$\begin{aligned} &P_{i(jk)} = \lim_{s \to 1} F_{i(jk)}(s) = \\ &\frac{P_{i(k)} - P_{i(k)j}P_{j(k)}}{1 - P_{i(k)j}P_{j(k)i}} = \frac{P_{i(j)} - P_{i(j)k}P_{k(j)}}{1 - P_{i(j)k}P_{k(j)i}} = \\ &\frac{\left(\begin{array}{c} \mu_j(\mu_{i|k} + \mu_{k|j} - \mu_{i|j}) + \mu_k(\mu_{i|j} + \mu_{j|k} - \mu_{i|k}) - \mu_i(\mu_{j|k} + \mu_{k|j}) \\ +\mu_{i|j}\mu_{j|k} + \mu_{i|k}\mu_{k|j} - \mu_{j|k}\mu_{k|j} \end{array} \right)}{\left(\begin{array}{c} \mu_{i|j}\mu_{k|i} + \mu_{i|k}\mu_{j|i} + \mu_{j|i}\mu_{k|j} + \mu_{j|k}\mu_{i|j} + \mu_{k|i}\mu_{j|k} + \mu_{k|j}\mu_{i|k} \\ -\mu_{i|j}\mu_{j|i} - \mu_{i|k}\mu_{k|i} - \mu_{j|k}\mu_{k|j} \end{array} \right)} \end{aligned}$$

(which can be easily extended to $P_{i(jk)\ell}$; just replace, in the last expression, μ_i, μ_j and μ_k by $\mu_{i|\ell}$, $\mu_{j|\ell}$ and $\mu_{k|\ell}$ respectively), and

$$\begin{aligned} &M_{(ijk)} = \lim_{s \to 1} \left(F'_{i(jk)}(s) + F'_{j(ik)}(s) + F'_{k(ij)}(s) \right) \\ &\frac{\left(\begin{array}{c} \mu_i(\mu_{j|i}\mu_{k|j} + \mu_{k|i}\mu_{j|k} - \mu_{j|k}\mu_{k|j}) + \mu_j(\mu_{i|j}\mu_{k|i} + \mu_{k|j}\mu_{i|k} - \mu_{i|k}\mu_{k|i}) \\ +\mu_k(\mu_{i|k}\mu_{j|i} + \mu_{j|k}\mu_{i|j} - \mu_{i|j}\mu_{j|i}) - \mu_{i|j}\mu_{j|k}\mu_{k|i} - \mu_{j|i}\mu_{k|j}\mu_{i|k} \end{array} \right)}{\left(\begin{array}{c} \mu_{i|j}\mu_{k|i} + \mu_{i|k}\mu_{j|i} + \mu_{j|i}\mu_{k|j} + \mu_{j|k}\mu_{i|j} + \mu_{k|i}\mu_{j|k} + \mu_{k|j}\mu_{i|k} \\ -\mu_{i|j}\mu_{j|i} - \mu_{i|k}\mu_{k|i} - \mu_{j|k}\mu_{k|j} \end{array} \right)} \end{aligned}$$

We can then define

$$F_{i(jk)\ell} = \frac{F_{i(k)\ell} - F_{i(k)j} \cdot F_{j(k)\ell}}{1 - F_{i(k)j} \cdot F_{j(k)i}}$$

and extend the game to four patterns, getting

$$F_{i(jk\ell)} = \frac{F_{i(k\ell)} - F_{i(k\ell)j} \cdot F_{j(k\ell)}}{1 - F_{i(k\ell)j} \cdot F_{j(k\ell)i}}$$

etc.

m **patterns competing:** One can show that, in the fully general case of m patterns competing, we get

$$F_{j(12\ldots j-1,j+1\ldots m)} = \frac{\left(\sum\limits_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1,i_2,\ldots i_m) \cdot \prod\limits_{k=1}^{m} F_{i_k()k} \right)_{F_{i_j()j} \to F_{i_j()}}}{\sum\limits_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1,i_2,\ldots i_m) \cdot \prod\limits_{k=1}^{m} F_{i_k()k}} \tag{3}$$

$$\frac{\sum\limits_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1,i_2,\ldots i_m) \cdot F_{i_j()} \prod\limits_{k\neq j}^{m} F_{i_k()k}}{\sum\limits_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1,i_2,\ldots i_m) \cdot \prod\limits_{k=1}^{m} F_{i_k()k}}$$

for the PGF of the j^{th} pattern winning over the rest, at Trial n, where $F_{j()i_j} \to F_{j()}$ implies the corresponding replacement throughout the expression in big parentheses (*including* $F_{j()j} \to F_{j()}$). Note that the summation is over all $m!$ *permutations* of the $i_1, i_2, \ldots i_m$ indices, sgn indicates the permutation's signature (1 or -1 for even or odd permutation, respectively). and that each $F_{i()i}$ (for any value of i) is equal to 1.

This implies (after applying the L'Hospital rule to (3), differentiating its numerator and denominator, individually, $m-1$ times, and then substituting 1 for the argument of each F - note that only the producs of first derivatives survive) that

$$P_{j(12\ldots j-1,j+1\ldots m)} = \frac{\left(\sum\limits_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1,i_2,\ldots i_m) \cdot \sum\limits_{\ell=1}^{m} \prod\limits_{k\neq \ell}^{m} \mu_{i_k|k} \right)_{\mu_{i_j|j} \to \mu_j}}{\sum\limits_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1,i_2,\ldots i_m) \cdot \sum\limits_{\ell=1}^{m} \prod\limits_{k\neq \ell}^{m} \mu_{i_k|k}}$$

where (*after* the $\mu_{j|i_j} \to \mu_j$ replacement, *including* $\mu_{j|j} \to \mu_j$), we set $\mu_{i|i} = 0$ for every other value of i. Alternately (and more explicitely), we can write

$$P_{j(12\ldots j-1,j+1\ldots m)} = \frac{\sum\limits_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1,i_2,\ldots i_m) \cdot \left(\prod\limits_{k\neq j}^{m} \mu_{i_k|k} + \mu_{i_j} \sum\limits_{\ell\neq j}^{m} \prod\limits_{k\neq \ell\neq j}^{m} \mu_{i_k|k} \right)}{\sum\limits_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1,i_2,\ldots i_m) \cdot \sum\limits_{\ell=1}^{m} \prod\limits_{k\neq \ell}^{m} \mu_{i_k|k}} \tag{4}$$

with the understanding that $\mu_{i|i} = 0$ (this formula no longer contains $\mu_{j|j}$). The $k \neq j$ notation indicates that the corresponding product is over all values of k, with the exception of j; similarly $k \neq \ell \neq j$ means that k will be skipping two values, ℓ and j (which are, due to the $\ell \neq j$ summation, different from each other). Note that the sum of these probabilities over all j values (from 1 to m) equals to 1, as it should be.

Finally

$$M_{(12\ldots m)} = \frac{\displaystyle\sum_{j=1}^{m} \mu_j \left(\sum_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1, i_2, \ldots i_m) \cdot \sum_{\ell=1}^{m} \prod_{k\neq\ell}^{m} \mu_{k|i_k} \right)_{\mu_{j|i_j}\to 0} - \sum_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1, i_2, \ldots i_m) \cdot \prod_{k=1}^{m} \mu_{k|i_k}}{\displaystyle\sum_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1, i_2, \ldots i_m) \cdot \sum_{\ell=1}^{m} \prod_{k\neq\ell}^{m} \mu_{k|i_k}}$$

or, more explicitely

$$M_{(12\ldots m)} = \frac{\displaystyle\sum_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1, i_2, \ldots i_m) \cdot \left(\sum_{j=1}^{m} \mu_j \prod_{k\neq j}^{m} \mu_{k|i_k} - \prod_{k=1}^{m} \mu_{k|i_k} \right)}{\displaystyle\sum_{(i_1,i_2,\ldots i_m)} \mathsf{sgn}(i_1, i_2, \ldots i_m) \cdot \sum_{\ell=1}^{m} \prod_{k\neq\ell}^{m} \mu_{k|i_k}}$$

To derive this result, one has to differentiate $\sum_{j=1}^{m} F_{j(12\ldots j-1,j+1\ldots m)}$ - see Formula (3) - with respect to s (the implicit argumet of all F's) and evaluate the answer at $s = 1$. The latter requires the L'Hospital rule again; one must also recall (and utilize) the fact that $\sum_{j=1}^{m} F_{j(12\ldots j-1,j+1\ldots m)}$ itself evaluates to 1.

Three pattern example: Returning to the three pattern version, we now introduce HTHTH (Pattern 3) to be played against the previous two. What we need now is

$$
\begin{aligned}
F_{3()} &= \frac{1}{1+(1-s)\cdot\dfrac{1+pqs^2+p^2q^2s^4}{p^3q^2s^5}} \\
\mu_3 &= \frac{1+pq+p^2q^2}{p^3q^2} \\
F_{3()1} &= \tilde{F}_{\mathrm{H}} = \frac{p^2q^2s^4(1-qs)}{1-s+pqs^2-pqs^3+p^2q^2s^4-p^2q^3s^5} \\
\mu_{3|1} &= \frac{1+pq}{p^3q^2} \\
F_{3()2} &= \tilde{F}_{\mathrm{HT}} = \frac{p^2qs^3(1-s+pqs^2)}{1-s+pqs^2-pqs^3+p^2q^2s^4-p^2q^3s^5} \\
\mu_{3|2} &= \frac{1+pq-p^3q}{p^3q^2} \\
F_{1()3} &= F_{\mathrm{HT}} = F_{1()2} = \frac{p^2qs^3\left(1-s+pqs^2\right)}{1-s+p^2q^2s^4-p^2q^3s^5} \\
\mu_{1|3} &= \mu_{\mathrm{HT}} = \frac{1-p^3q}{p^3q^2} \\
F_{2()3} &= \hat{F}_{\mathrm{TH}} = \frac{p^2qs^3\left(1-s+pqs^2\right)}{1-s+p^3qs^4-p^4qs^5} \\
\mu_{2|3} &= \frac{1-p^2q^2}{p^3q^2}
\end{aligned}
$$

where $\tilde{F}$, $\tilde{F}_{\mathrm{H}}$,... represent the PGFs of Pattern 3, found from

$$
\begin{aligned}
\tilde{F} &= s\left(p\tilde{F}_{\mathrm{H}}+q\tilde{F}\right) \\
\tilde{F}_{\mathrm{H}} &= s\left(p\tilde{F}_{\mathrm{H}}+q\tilde{F}_{\mathrm{HT}}\right) \\
\tilde{F}_{\mathrm{HT}} &= s\left(p\tilde{F}_{\mathrm{HTH}}+q\tilde{F}\right) \\
\tilde{F}_{\mathrm{HTH}} &= s\left(p\tilde{F}_{\mathrm{H}}+q\tilde{F}_{\mathrm{HTHT}}\right) \\
\tilde{F}_{\mathrm{HTHT}} &= s\left(p+q\tilde{F}\right)
\end{aligned}
$$

This results in

$$
\begin{aligned}
P_{1(23)} &= \frac{1+pq^2-p^3q^3+p^5q^3}{3+pq-2p^2q+p^4q^2} \\
P_{2(13)} &= \frac{1-p^4q-p^5q^3}{3+pq-2p^2q+p^4q^2} \\
P_{3(12)} &= \frac{1-p^2q^2}{3+pq-2p^2q+p^4q^2}
\end{aligned}
$$

or (when $p = q = \frac{1}{2}$) 36.92%, 31.99% and 31.09% respectively (Pattern 1 is clearly the big winner). Similarly

$$M_{(123)} = \frac{1 + 2pq - pq^3 - p^5q^2 - 2p^4q^4 - p^5q^5 - p^8q^4}{p^3q^2(3 - pq + 2pq^2 + p^4q^2)}$$

or (when $p = q = \frac{1}{2}$), 15.08 flips per game (a long run average).