A MARKOVIAN GENERALIZATION OF FELLER'S COIN TOSSING CONSTANTS

CYRIL BANDERIER & MATHIAS VANDENBOGAERT.
PERSONAL NOTE. 27 NOVEMBER 2000
ALGORITHMS PROJECT, INRIA.
HTTP://ALGO.INRIA.FR/

This note answers to Steve Finch's question at the end of his webpage [1] on "Feller's Coin Tossing Constants". The question is how to adapt known techniques (the classical Bernoulli model) in order to study motifs in a random text with respect to a Markovian model?

1. Feller's Coin Tossing Constants. Bernoulli Model.

We are interested in the probability $w(n)$ that in n independent tosses of an coin, no run of k (or more) consecutive heads appears.

This means that we only accept sequences of the shape

$$\mathcal{L}_k = (\epsilon + H + HH + \cdots + H^{k-1})(T^+H + T^+HH + \cdots + T^+H^{k-1})^*T^*$$

where $T^+ = T + T^2 + T^3 + \ldots$ stands for a positive number of consecutive T, and $A^* = A^0 + A + A^2 + \ldots$ stands for a nonnegative number of consecutive occurrences of A, and ϵ stands for the empty word.

This is a *regular expression* which define a language, that is, a set of words. Note that the above expression is non-ambiguous, that is, any sequence without k consecutive heads is decomposable in a unique way via the above expression. This implies that the number of valid sequences of length n is exactly the number of words of length n which can generated by the above expression.

Let a_n be the number of valid sequences (=words belonging to $\mathcal{L}_k$) of length n and consider now the generating function $L_k(z)$ associated to $\mathcal{L}_k$, defined by

$$L_k(z) = \sum_{n\geq 0} a_n z^n\,.$$

Since Chomsky–Schützenberger (1963), it is well known that such a generating function (associated to a language which is given by a regular expression) is rational.

The *symbolic method* shows us how to convert immediately a regular expression into its generating function. Replace the letters H and T by z and then $(\epsilon + H + HH + \cdots + H^{k-1})$ becomes $\frac{z^k-1}{z-1}$, T^* becomes $\frac{1}{1-z}$, T^+ becomes $\frac{z}{1-z}$ and so on. Thus

$$L_k(z) = \frac{z^k-1}{z-1}\frac{1}{1-\frac{z}{1-z}\frac{z^k-z}{z-1}}\frac{1}{1-z}$$

If you want to take of the fact that the probability to have H or T (that is, you want to consider a Bernoulli probabilistic model with parameter (p,q)), then replace H by pz and T by qz.

This leads to

$$L_k(z) = \frac{(pz)^k - 1}{pz - 1} \frac{1}{1 - \frac{qz}{1-qz}\frac{(pz)^k - pz}{pz-1}} \frac{1}{1-qz} = \frac{1-(pz)^k}{1 - z + qz(pz)^k}$$

Evaluation with $k = 3$ and $p = q = \frac{1}{2}$ gives

$$L_k(z) = 1 + z + z^2 + \frac{7}{8}z^3 + \frac{13}{16}z^4 + \frac{3}{4}z^5 + O(z^6).$$

Remind that asymptotics of coefficients of a rational function depends directly on the poles of this function :

Proposition 1. *If F is a rational function with poles $\rho_1, \dots, \rho_m \neq 0$, then it exists m polynomials $P_1, \dots, P_m$ (where P_j has degree order of the pole ρ_j minus 1) such that*

$$F_n = [z^n]F(z) = \sum_{j=1}^{m} P_j(n)\rho_j^{-n}.$$

Proof. Use the decomposition into partial fraction form $F(z)$ and

$$[z^n](z-\rho)^{-k} = \frac{(-1)^k}{\rho^{n+k}}\binom{n+k-1}{k-1}$$

this gives a polynomial of degree $k-1$ in n. □

Considering only the dominant term into the asymptotics leads to the following result

Theorem 1. *The probability to get a sequence of length n without k consecutive heads is asymptotically*

$$w(n) = a_n = \frac{1 - p\alpha}{(k + 1 - k\alpha)q\alpha}\alpha^{-n} + o(\alpha^{-n})$$

where α is the smallest (with respect to the modulus) root of the denominator

$$1 - z + qz(pz)^k$$

of the generating function $L_k(z)$.

Proof. Consider $L_k(z) = \frac{A(z)}{B(z)}$. For $k > 1$, the smallest root α of B is a simple root, thus

$$L_k(z) \sim_{x \sim \alpha} \frac{A(\alpha)}{B'(\alpha)} \frac{1}{z - \alpha}.$$

One gets $\beta = \frac{-A(\alpha)}{\alpha B'(\alpha)}$ and $w(n) = \beta\alpha^{-n}$. □

2. Feller's Coin Tossing Constants. Markov Model.

For the Markov model (that is, the probability to toss H or T depends now on the last toss), the same idea applies.

Simply rewrite $\mathcal{L}_k = (\epsilon + S)(T^+S)^*T^*$ (where $S = H + \cdots + H^{k-1}$) in such a way that it is easy to see when a H follows H or T (idem for T). Firstly expand the previous expression for $\mathcal{L}_k$:

$$\mathcal{L}_k = \epsilon + TT^* + (TT^*S)(TT^*S)^*(\epsilon + TT^*) + S(1 + TT^*) + S(TT^*S)^+(\epsilon + TT^*)$$

And then rewrite T_i for any T beginning a word, rewrite T_0 for any T following a H and rewrite $S_i = H_i(H + \cdots + H^{k-2})$ for a run of heads beginning a word (whereas $S_i = H_1(H + \cdots + H^{k-2})$ stands for intermediate sequences of heads)

$$\mathcal{L}_k = \epsilon + T_iT^* + (T_iT^*S)(T_0T^*S)^*(\epsilon + T_0T^*) + +S_i\left(\epsilon + (T_0T^*S)^+\right)(\epsilon + T_0T^*)$$

Finally, replace T by $p_{11}z$, T_0 by $p_{01}z$ and T_i by i_1z. Replace also H by $p_{00}z$, H_1 by $p_{10}z$ and H_i by i_0z. This leads to

$$L_k(z) = 1 + \frac{i_0z}{1-p_{00}z} + \frac{i_0z}{1-p_{00}z}p_{01}z\,\frac{(p_{11}z)^{k-1}-1}{p_{11}z-1}\,\frac{1+\frac{p_{10}z}{1-p_{00}z}}{1-\frac{p_{10}z}{1-p_{00}z}p_{01}z\frac{(p_{11}z)^{k-1}-1}{p_{11}z-1}}$$

$$+i_1z\frac{(p_{11}z)^{k-1}-1}{p_{11}z-1}\left(1+\frac{\frac{p_{10}z}{1-p_{00}z}p_{01}z\frac{(p_{11}z)^{k-1}-1}{p_{11}z-1}}{1-\frac{p_{10}z}{1-p_{00}z}p_{01}z\frac{(p_{11}z)^{k-1}-1}{p_{11}z-1}}\right)\left(1+\frac{p_{10}z}{1-p_{00}z}\right).$$

This simplifies to

$$L_k(z) = \frac{A(z)}{B(z)} = \frac{p_{11} + (p_{01}p_{11} - p_{11}^2)z + (p_{11}^k(i_0-1))z^k + (p_{11}^{k+1}(1-p_{01}-i_0))z^{k+1}}{p_{11} + (p_{11}p_{01} - p_{11}^2 - p_{11})z + (p_{11}^2 - p_{01}p_{11})z^2 + (p_{01}p_{11}^k(1-p_{11}))z^{k+1}}$$

For sure, setting $k = 3$ and $p_{00} = p_{10} = p_{11} = p_{01} = \frac{1}{2}$ gives the same result as in the previous exemple for the Bernoulli case:

$$L_k(z) = \frac{16-2z^3}{16-16z+z^4} = 1 + z + z^2 + \frac{7}{8}z^3 + \frac{13}{16}z^4 + \frac{3}{4}z^5 + O(z^6).$$

As for the Bernoulli model, one deals now with the asymptotics of a rational generating function and one gets:

Theorem 2. *Under a Markov model* $\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$, *with an initial probability* i_0 *to begin with a tail, and* i_1 *to begin with a head, the probability to get a sequence of length* n *without* k *consecutive heads is asymptocally*

$$w(n) = \beta\alpha^{-n} + o(\alpha^{-n})$$

where α *is the smallest (with respect to the modulus) root of the denominator* $B(z)$ *of the generating function* $L_k(z)$ *and where*

$$\beta = \frac{-A(\alpha)}{\alpha B'(\alpha)} = \frac{\left(-1 + (2p_{11} - p_{01})\alpha + (p_{01}p_{11} - p_{11}^2)\alpha^2\right)(i_1 + (p_{01} - i_1)\alpha)}{p_{01}p_{10}\alpha\left(-1 - k + k(p_{11} + p_{00})\alpha + (p_{11} - p_{01})(1-k)\alpha^2\right)}.$$

3. General patterns, Bernoulli or Markov Model.

For general patterns, it is easier to adapt the automaton point of view. An automaton is graph with the edges (the "transitions") are oriented and where the vertices (the "states") can be of three types "initial", "final" or "intermediary" (a state can be both intial, final, and intermediary). What is more, each transition is marked by a letter of the alphabet (say H and T in our coin-tossing problem).

A word is said to be recognized by the automaton if it is possible to start in a initial state, then to read each letter of the word following some transition in the automaton and finally ending in a final state when one reads the last letter of the word. A classical resultasserts that the sets of words recognized by an automata corresponds to a regular expressions (and reciprocally). Thus, an automaton can be associated to the regular expression for $\mathcal{L}_k$. A Markov chain is simply can be considered as an automaton for which the transitions are probabilized. For the Bernoulli model, the automata recognizing the pattern (= theregular expression) under study can be easily built, then the generating function can be made explicit and asymptotics follow. For the Markov model, making a kind of tensor product between the automata and of the Markov matrix gives a new automata and then proceed as in the classical Bernoulli model (note that we used a slightly different trick in the previous section).

That's what Nicodème, Salvy & Flajolet explain in their article [2]. They give an algorithm for making explicit the (bivariate) generating function for the number of occurrences of any regular expression (thus for every pattern, or even an infinite set of patterns). The bivariate generating function $F_E(z,u) = \sum_{n,k\geq 0} f_{n,k} u^k z^n$ gives the probablity f_n, k to get k occurrences of the pattern E in a random text of length n.

The text can be random under two models : Bernoulli's model or Markov's model.

The bivariate generating function can be used to perform asymptotics (average, higher moments...). Thus, they proved that a Gaussian law holds.

Naturally, setting $k = 0$ in the bivariate generating function gives the probability that there is *zero* occurrence of the given pattern in a text of length n.

Note that most of these computations can be done automatically, e.g. in Maple, with the help of the packages [3] `Combstruct` and `Regexpcount`. A worksheet will be available "soon". :-)

References

[1] Steve Finch. Feller's coin tossing constants. 1998. http://www.mathsoft.com/asolve/constant/feller/feller.html.

[2] Pierre Nicodème, Bruno Salvy, and Philippe Flajolet. Motif statistics. In J. Nešetril, editor, *Algorithms*, volume 1643 of *Lecture Notes in Computer Science*, pages 194–211, 1999. Proceedings of 7th Annual European Symposium on Algorithms ESA'99, Prague, July 1999.

[3] Algorithms Project. AlgoLib (Libraries for Maple), 2000. http://algo.inria.fr.