In memory of my academic great-grandfather, Willy FELLER and my biological great-grandfather, Adolf PINNER

Two (Truly) GREAT Grandparents: Adolf PINNER and Willy FELLER

No wonder I am such a good writer! It runs in the biological family. My (biological) great-grandfather, Adolf Pinner (1842-1909), in addition to being a brilliant chemist—he discovered the structure of Nictoine and coined the name Pyrimidine—was also a great expositor, and his chemistry books[P1][P2] were the standard textbooks in that subject between 1872 and 1910 (and beyond) in Germany and elsewhere. He also wrote a charming popular book[P3], The Laws of the Phenomena of Nature, that my second-cousin, Sister Dr. Katherine Wolff (who is also his great-grandchild, and also a great writer!) is currently translating from German to English, and her translation is available from http://www.math.rutgers.edu/~zeilberg/family/gesetze.html.

No wonder I am such a good writer! It runs in the academic family. My (academic) great-grandfather, Willy Feller (1906-1970), in addition to being a brilliant probabilist—he discovered the Feller process and contributed significantly to the relation between Markov processes and differential equations—was also a great expositor, and his probability books[F1][F2] were the standard textbooks in that subject between 1950 and 1980 (and beyond) in the USA and elsewhere, until students got dumber and needed more watered-down texts.

Patience

So if I am such a good writer, how come I don’t write textbooks? Well, it takes more than that. Both my great-grandfathers had in abundance something that I don’t have: patience. Both Feller and Pinner had lots and lots of drafts, and kept editing and polishing them.

The Legendary Chapter III

My favorite chapter in Feller’s classic[F1] is Chapter III. It is chuckfull of clever reflective combinatorics, and waxes eloquent about the amazing discrete arcsine law that is so counter-intuitive: Most people are either winners most of the time or losers most of the time. However, if they break even when they die, then it is equally likely that they were winners half of the time as that they were winners all the time or no time, or quarter of the time, etc.
Generating Functions

The very first approach, abandoned later, was to use generating functions. Feller (and, even many people today) considered it as an analytical method. In the old days, one first used combinatorial or probabilistic arguments to derive recurrences for the sequence. Then they took the (ordinary or exponential) generating function of the sequence, and showed that it satisfied a (differential and/or algebraic and/or functional) equation, then one solved the equation for the generating function, and if in luck was able to deduce closed-form expressions for the coefficients.

Weight-Enumerators

But Pólya, Tutte and Schützenbecker showed us that the above outline is stupid. First, generating functions are better viewed as neither generating nor functions. They are rather formal power series that are weight-enumerators of combinatorial sets, and one can operate with them directly. Also, with computers, we no longer care how “messy” these generating functions turn out to be, and computers can guess, and then a posteriori prove, the closed-form expressions.

Coin Tosses

Feller considered sequences of coin tosses \((H = 1, T = -1)\), let’s call them \(w\) of length \(n\) and was interested in the following parameters.

\[ a_1(w) := \text{the number of losing times, in symbols:} \]
\[ a_1(w) := |\{i | \sum_{j=1}^{i} w_j < 0 \text{ or } \sum_{j=1}^{i} w_j = 0 \text{ and } \sum_{j=1}^{i-1} w_j < 0\}|, \]

\[ a_2(w) := \text{the number of times it was breaking-even, in symbols:} \]
\[ a_2(w) := |\{i | \sum_{j=1}^{i} w_j = 0\}|, \]

\[ a_3(w) := \text{the last break-even time, in symbols:} \]
\[ a_3(w) := \max\{i | \sum_{j=1}^{i} w_j = 0, \sum_{j=1}^{r} w_j > 0 \text{ for } r > i\}, \]

\[ a_4(w) := \text{number of sign-changes, in symbols:} \]
\[ a_4(w) := |\{i | \sum_{j=1}^{i-1} w_j < 0 \text{ and } \sum_{j=1}^{i+1} w_j > 0 \text{ or } \sum_{j=1}^{i-1} w_j > 0 \text{ and } \sum_{j=1}^{i+1} w_j < 0\}|. \]
Feller [F1] (Ch. III) only considered each of these parameters one at a time, but we can consider the grand generating function

\[ F(z, t_1, t_2, t_3, t_4) := \sum_{n=0}^{\infty} \sum_{i_1=0}^{n} \sum_{i_2=0}^{n} \sum_{i_3=0}^{n} \sum_{i_4=0}^{n} A_n(i_1, i_2, i_3, i_4) t_1^{i_1} t_2^{i_2} t_3^{i_3} t_4^{i_4} z^n, \]

where \( A_n(i_1, i_2, i_3, i_4) \) is the number of sequences \( w \) of length \( n \) with \( a_1(w) = i_1, a_2(w) = i_2, a_3(w) = i_3, a_4(w) = i_4 \). Equivalently, (and much better!)

\[ F(z, t_1, t_2, t_3, t_4) := \sum_{w \in \{-1, 1\}^*} \text{weight}(w), \]

where the weight of an arbitrary word in the alphabet \( \{-1, 1\} \) is

\[ \text{weight}(w) := t_1^{a_1(w)} t_2^{a_2(w)} t_3^{a_3(w)} t_4^{a_4(w)} z^{\text{length}(w)}. \]

Also let \( C(z, t_1, t_2, t_3, t_4) \) be the analogous weight-enumerator for those sequences that add to 0 (i.e. broke-even at the end).

**Deriving the Five-Variable Generating Function ONCE and FOR ALL**

The beginning is as in [Z]. Let \( \phi(z) \) be the generating function, according to length, of words in \( \{-1, 1\} \) that broke-even at the end but was never losing. Then of course

\[ \phi = 1 + z^2 \phi^2, \]

since such a sequence is either empty (weight 1) or is a catenation of an irreducible sequence (one that only breaks-even at the beginning and the end) [weight \( z\phi(z)z \)] and a shorter sequence [weight \( \phi(z) \)]. Using the 3000-year-old formula for the quadratic equation, one can get \( \phi(z) \) (and hence \( \psi(z) := z^2 \phi(z) \)) explicitly. I am resisting the temptation to write it down, since why bother? the computer can do it for itself.

Let \( \alpha(z) \) be the generating function (weight-enumerator), according to length, of the set of sequences that except at the beginning always had a positive amount (partial sum).

Let \( \beta(z) \) be the generating function (weight-enumerator), according to length, of the set of sequences that except at the beginning always had a negative amount (partial sum).

By symmetry \( \alpha = \beta \) and by factoring such a winning sequence according to each time where it had the next largest amount for the last time we immediately get

\[ \alpha(z) = \beta(z) = \frac{z}{1 - z - \psi(z)}. \]

Now given a sequence, let’s call
• a segment between two consecutive break-even times, where it was *not losing* in-between, a $P$-segment.

• a segment between two consecutive break-even times, where it was *not winning* in-between, an $N$-segment.

• a segment that started with 0 dollars, but was *winning* for ever after, a $P'$-segment.

• a segment that started with 0 dollars, but was *losing* for ever after, an $N'$-segment.

If you factor a break-even sequence (i.e. with sum 0) according to its sign-changes, and look at it as a sequence of $P$’s and $N$’s, we have the following *regular expression*

$$\text{empty} \lor P(NP)^* \lor P(NP)^*N \lor N(PN)^* \lor N(PN)^*P,$$

whose generating function (alias weight-enumerator) (in $P,N,t_4$) is:

$$\mathcal{C} = 1 + P \frac{1}{1 - t_4 N t_4 P} + P \frac{1}{1 - t_4 N t_4 P} t_4 N + N \frac{1}{1 - t_4 P t_4 N} + N \frac{1}{1 - t_4 P t_4 N} t_4 P.$$

If you factor an *arbitrary* sequence according to its sign-changes, and look at it as a sequence of $P$’s and $N$’s, possibly followed by $N'$ or $P'$ we have the following *regular expression*

$$\text{empty} \lor P(NP)^* \lor P(NP)^*N \lor N(PN)^* \lor N(PN)^*P$$

$$\lor P' \lor P(NP)^*P' \lor P(NP)^*NP' \lor N(PN)^*P' \lor N(PN)^*PP'$$

$$\lor N' \lor P(NP)^*N' \lor P(NP)^*N'N' \lor N(PN)^*N' \lor N(PN)^*PN'.$$

whose generating function (in $P,N,P',N',t_4$) is:

$$\mathcal{F} = 1 + P \frac{1}{1 - t_4 N t_4 P} + P \frac{1}{1 - t_4 N t_4 P} t_4 N + N \frac{1}{1 - t_4 P t_4 N} + N \frac{1}{1 - t_4 P t_4 N} t_4 P$$

$$+ P' \frac{1}{1 - t_4 N t_4 P} P' + P \frac{1}{1 - t_4 N t_4 P} t_4 N t_4 P' + N \frac{1}{1 - t_4 P t_4 N} t_4 P' + N \frac{1}{1 - t_4 P t_4 N} t_4 PP'$$

$$+ N' \frac{1}{1 - t_4 N t_4 P} t_4 N' + P \frac{1}{1 - t_4 N t_4 P} t_4 N N' + N \frac{1}{1 - t_4 P t_4 N} N' + N \frac{1}{1 - t_4 P t_4 N} t_4 P t_4 N'.$$

Now substituting in $\mathcal{F}$,

$$P = t_2 \psi(t_3 z)/(1 - t_2 \psi(t_3 z))$$

$$N = t_2 \psi(t_1 t_3 z)/(1 - t_2 \psi(t_1 t_3 z))$$

$$P' = \alpha(z)$$

$$N' = \alpha(t_1 z).$$

you would get an *explicit* (but “large”) expression for $F(z,t_1,t_2,t_3,t_4)$.

Similarly, plugging-in $\mathcal{C}$ the $P$ and $N$ above, would give $C(z,t_1,t_2,t_3,t_4)$, the analog for those sequences that add to 0.
This is all done internally by the Maple package FELLER that easily proves Theorem III.4.1 (that the number of sequences of \{-1,1\} of length 2n where the longest prefix that sums to 0 is 2k equals \(\binom{2k}{k}\binom{2n-2k}{n-k}\)), Theorem III.4.2 (that the number of sequences of \{-1,1\} of length 2n that was winning exactly 2k times equals \(\binom{2k}{k}\binom{2n-2k}{n-k}\) [the discrete arc-sine law]), and Theorem III.9 (that the number of sequences of \{-1,1\} of length 2n and that sum to 0, that was winning exactly 2k times equals \(\binom{2n}{n}/(n+1)\) [the Chung-Feller theorem]).

These are proved in the Maple package FELLER by typing ThIII41();, ThIII42();, and ThIII9();, respectively.

It is possible to milk \(F(z,t_1,t_2,t_3,t_4)\) and \(C(z,t_1,t_2,t_3,t_4)\) much more. You can easily find more refined enumerations, averages, variances, higher moments, covariances etc. etc. You are welcome to explore FELLER on your own!

References


