Automated Derivation of Limiting Distributions Of Combinatorial Random Variables Whose Generating Functions are Rational

By Doron ZEILBERGER

Abstract: The author's methodology of automated derivation of explicit expressions for the average, variance, and higher moments of combinatorial random-variables that possess explicit generating functions is extended to those whose bi-variate generating function is an arbitrary rational function.

Maple Package and Sample Output

This article is accompanied by a Maple package, `BiVariateMoms.txt`, that is available, along with some sample input and output files, from

`http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/crv.html` .

Ancient History

If you toss a fair coin n times, and are interested in the random variable 'number of Heads', let's call it H_n, and plot the graph of

$$k \to Prob(H_n = k) \quad ,$$

and n is large enough, you **very famously**, get something close to a *bell-shaped curve*, and as n gets larger and larger, it would get closer and closer to it. The official names of the bell-shaped curve are '*standard normal distribution*' and '*Gaussian distribution*'. More precisely, letting μ_n denote the *expectation* (aka *mean*, aka *average*), and σ_n the standard deviation (that happen to be equal to $\frac{n}{2}$ and $\frac{\sqrt{n}}{2}$ respectively), and defining the *centralized and scaled* version,

$$X_n := \frac{H_n - \mu_n}{\sigma_n} \quad ,$$

then the sequence of discrete random variables, X_n, 'tend to', in a certain precise sense, to the (**continuous distribution**) called *standard normal distribution*, whose *probability density function* is *famously*

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad .$$

This means That

$$\lim_{n\to\infty} Pr(X_n \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \quad .$$

This was first proved, by de Moivre and Laplace, using fairly *ad hoc* methods, but has since been greatly generalized, and the collective name is *Central Limit Theorems.*

From Enumeration to Statistics

Suppose that we have a finite set S, on which a certain numerical attribute, called *random variable*, X, (using the probability/statistics lingo), is defined.

For any non-negative integer i, let's define

$$N_i := \sum_{s \in S} X(s)^i \quad .$$

In particular, $N_0(X)$ is the number of elements of S.

The expectation of X, $E[X]$, denoted by μ, is, of course,

$$\mu = \frac{N_1}{N_0} \quad .$$

For $i > 1$, the i-th straight moment is

$$E[X^i] = \frac{N_i}{N_0} \quad .$$

The i-th *moment about the mean* is

$$m_i := E[(X-\mu)^i] = E[\sum_{r=0}^{i} \binom{i}{r} (-1)^r \mu^r X^{i-r}] = \sum_{r=0}^{i} (-1)^r \binom{i}{r} \mu^r E[X^{i-r}]$$

$$= \sum_{r=0}^{i} (-1)^r \binom{i}{r} \left(\frac{N_1}{N_0}\right)^r \frac{N_{i-r}}{N_0}$$

$$= \frac{1}{N_0^i} \sum_{r=0}^{i} (-1)^r \binom{i}{r} N_1^r N_0^{i-r-1} N_{i-r} \quad .$$

Finally, the most interesting quantity, statistically speaking, apart from the mean μ and variance m_2 are the **scaled-moments**, also known as, *alpha coefficients*.

$$\alpha_i := \frac{m_i}{m_2^{i/2}} \quad .$$

Getting the Moments via Generating Functions

The *generating function* of our random variable, X, also called *weight-enumerator*, with respect to the variable t, is defined by

$$f(t) = f_{S,X}(t) \; := \; \sum_{s \in S} t^{X(s)} \quad .$$

Note that when $t = 1$, we get the *naive enumeration*, $|S|$.

More generally we have, with the above notation,

$$N_i = \left(t \frac{t}{dt} \right)^i f(t) \Big|_{t=1} \quad .$$

In many enumeration scenarios, we don't have just *one* finite set, S, but an *infinite sequence* of sets S_n, and the generating function $f_n(t)$ can be expressed *explicitly* in terms of **both** the 'continuous' variable, t, and discrete variable n.

For example, if the set S_n is the set $\{H, T\}^n$ and $X_n(s)$ is the "number of heads", then by 'independence', we have a **closed form formula** for $f_n(t)$:

$$f_n(t) = (1 + t)^n \quad .$$

More generally, if S_n is the set of sequences of outcomes of rolling, n times, a k-faced die whose faces are marked with $a_1, \ldots, a_k$ dots, in other words the set $\{a_1, \ldots, a_k\}^n$, and X_n is the random variable 'total number of dots' (i.e. the sum of the sequence of outcomes), then

$$f_n(t) = \left(\sum_{j=1}^{k} t^{a_j} \right)^n \quad .$$

So, for each non-negative integer i, we have a *sequence of numbers*, $N_i(n)$, obtained by applying $(t\frac{d}{dt})^i$ to $f_n(t)$, and then plugging-in $t = 1$. We always get *polynomials* for the $N_i(n)$. From them we can get polynomial expressions for the moments-about-the-mean, $m_i(n)$. This is better delegated to computers. In many cases, one can also get the leading terms for **symbolic** i, and prove completely automatically, *asymptotic normality* up to any desired moment (i.e. that the scaled moments, $\alpha_i(n)$, tend, as n goes to infinity, to those of the standard normal distribution $1, 0, 3, 0, 5, 0, 15, 0, 105, \ldots$. In fact, it is even possible to prove it for an arbitrary moment (symbolic i), hence giving fully automated proofs of central limit theorems. This was described in [Z1] and [Z2].

The Grand Generating Function

Since we have an *infinite* sequence of polynomials, $f_n(t)$, we can form the **grand generating function**

$$F(t, z) := \sum_{n=0}^{\infty} f_n(t) z^n \quad .$$

For example, for the coin-tossing example, we have

$$F(t,z) = \frac{1}{1-(1+t)z} \quad ,$$

and for the more general die-rolling example, we have

$$F(t,z) = \frac{1}{1-z\left(\sum_{j=1}^{k} t^{a_j}\right)} \quad .$$

Notice that these are *rational functions* in **both** t and z.

But there are **many** examples in enumerative combinatorics (and statistical physics!), where there is no 'closed-form' expressions for the *individual* $f_n(t)$, but the 'grand-generating function', $F(t,z)$ is rational. These occur, for example, in *tiling* problems, see the beautiful article [Z3] with its accompanying Maple package `TILINGS`.

Note that in such cases, the enumerating sequence, $\{N_0(n)\}$ is C-finite, i.e. satisfies a (homog.) linear recurrence equation with **constant** coefficients (see [Z4] for a brief overview of C-finite sequences, and the modern classic [KP] for an in-depth account.) Obviously,

$$\sum_{n=0}^{\infty} N_0(n) z^n = F(1,z) \quad .$$

More generally,

$$\sum_{n=0}^{\infty} N_i(n) z^n = \left(t\frac{d}{dt}\right)^i F(t,z)\Big|_{t=1} \quad ,$$

are always rational functions of z, and hence the sequences $\{N_i(n)\}_{n=0}^{\infty}$ are always C-finite.

Since $F(t,z)$ is a rational function, we can write it as a quotient of polynomials

$$F(t,z) = \frac{P(t,z)}{Q(t,z)} \quad .$$

It follows from the *quotient rule* from 'calc1', that, for each i,

$$\sum_{n=0}^{\infty} N_i(n) z^n = \left(t\frac{d}{dt}\right)^i F(t,z)\Big|_{t=1} = \frac{P_i(z)}{Q(1,z)^{i+1}} \quad ,$$

for *some* polynomial $P_i(z)$. Hence, by partial fraction decomposition, it follows that, denoting by L the order of the recurrence for $N_0(n)$ (alias the degree, in z, of $Q(1,z)$), we are *guaranteed* that for each i, there exist polynomials $A_{i,j}(n)$, $0 \leq j \leq L-1$, of degree i in n, such that

$$N_i(n) = \sum_{j=0}^{L-1} A_{i,j}(n) \cdot N_0(n-j) \quad .$$

Since we are guaranteed, *a priori*, for each i, that the L polynomials $A_{i,0}(n), \ldots, A_{i,L-1}(n)$ (of degree i in n), **exist**, we can ask our computers to find them empirically, by cranking-out sufficiently many terms of the sequence $\{N_0(n)\}$ and $\{N_i(n)\}$, and using *undetermined coefficients*.

Once we have 'explicit' expressions (in the above sense) for $N_i(n)$, we can get 'explicit' expressions (in the same sense), for the (straight) moments, and from them, the more informative *moments about the mean*, that, in turn, lead to the scaled moments $\alpha_i(n)$. Denoting by β the smallest positive root of $Q(1,z)$, (a certain *algebraic* number) and noting that

$$\lim_{n\to\infty} \frac{N_0(n-1)}{N_0(n)} = \beta \quad ,$$

and hence, more generally, for $0 \leq j < L$,

$$\lim_{n\to\infty} \frac{N_0(n-j)}{N_0(n)} = \beta^j \quad ,$$

we can get asymptotic expressions (with exponential 'error' in n) for all the quantities, in particular, for the expectation $\mu(n)$ and variance $m_2(n)$ in terms of n and β. Also we can get asymptotic expressions (done automatically!) for the scaled moments $\alpha_i(n)$, and verify, each time, that as n goes to infinity, they converge to 0 if i is odd, and $i!/(2^{i/2}(i/2)!)$ if i is even (i.e. the moments of the standard normal distribution). Even better, we can get more refined asymptotics, all done automatically, to any order in n.

It is possible to show, under mild hypothesis, that the sequence of random variable X_n whose grand-generating function is rational in t and z is always asymptotically normal, hence all we need is to find, in each case, the expectation, $\mu(n)$, and the variance $m_2(n)$, in terms of $N_0(n)$ and its shifts, and if all we care about is the limiting distribution, then it suffices to ask the computer to find the asymptotic expressions of these in terms of n and β.

The Maple Package `BiVariateMoms.txt`

Everything is implemented in the Maple package `BiVariateMoms.txt`. See the webpage of this article, where there are numerous input and output files.

The simplest non-trivial example

Let the set S_n be the set of sequences in $\{1,2\}$ that add-up to n, and let $X_n(s)$ be the random variable 'number of 2s' in s. It is readily seen that

$$F(t,z) = \frac{1}{1 - z - t\,z^2} \quad .$$

Here $N_0(n) = F_{n+1}$.Then the program says that the expectation, $\mu(n)$ equals **exactly**

$$\frac{2\,n\,F_{n+1} - (n+1)\,F_n}{5F_{n+1}} \quad ,$$

while the variance, $m_2(n)$, equals

$$m_2(n) = \frac{n(n+3)\,F_{n+1}^2 - (n^2-1)\,F_{n+1}F_n - (n+1)^2\,F_n^2}{25\,F_{n+1}^2}.$$

Let ϕ be the golden ratio, then $N_0(n) = a(n)$ is asymptotic (with exponentially small error) to $\frac{\phi+1}{\phi+2} \cdot \phi^n$. The expectation is, asymptotically,

$$\frac{(2\phi-1)n-1}{5\phi} \quad ,$$

and the variance, $m_2(n)$, is, asymptotically,

$$\frac{(3\phi+1)n+(\phi-1)}{25(\phi+1)} \quad .$$

For all the moments up to the sixth, see the output file

`http://www.math.rutgers.edu/~zeilberg/tokhniot/oBiVariateMoms1.txt` .

For numerous other examples, see the front of this article, mentioned above:

`http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/crv.html` .

Acknowledgment: Many thanks to Tamar Zeilberger for insightful comments.

References

[KP] Manuel Kauers and Peter Paule, *"The Concrete Tetrahedron"*, Springer, 2011.

[Z1] Doron Zeilberger, *The Automatic Central Limit Theorems Generator (and Much More!)*, `http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/georgy.html` ; Also appeared in: "Advances in Combinatorial Mathematics: Proceedings of the Waterloo Workshop in Computer Algebra 2008 in honor of Georgy P. Egorychev", chapter 8, pp. 165-174, (I.Kotsireas, E.Zima, eds. Springer Verlag, 2009.)

[Z2] Doron Zeilberger, *HISTABRUT: A Maple Package for Symbol-Crunching in Probability theory*, The Personal Journal of Shalosh B. Ekhad, Aug. 25, 2010, `http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/histabrut.html` . Also published in `https://arxiv.org/abs/1009.2984`.

[Z3] Doron Zeilberger, *Automatic CounTilings*, The Personal Journal of Shalosh B. Ekhad and Doron Zeilberger, Jan. 20, 2006.
`http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/tilings.html` .

[Z4] Doron Zeilberger, *The C-finite Ansatz*, `http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/c`
also published in the Ramanujan Journal **31** (2013), 23-32.

Doron Zeilberger, Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, USA;
DoronZeil at gmail dot com ; `http://www.math.rutgers.edu/~zeilberg/` .

Exclusively published in The Personal Journal of Shalosh B. Ekhad and Doron Zeilberger (`http://www.math.rutgers.edu/~zeilberg/pj.html`) .

Dec. 24, 2016