The Automatic Central Limit Theorems Generator (and Much More!)

Doron ZEILBERGER[1]

Dedicated to Georgy Petrovich EGORYCHEV on his 70th birthday

Why I hate the Continuous and Love the Discrete

I have always loved the discrete and hated the continuous. Perhaps it was the trauma of having to go through the usual curriculum of "rigorous" , Cauchy-Weierstrass-style, real calculus, where one has all those tedious, pedantic and *utterly boring*, $\epsilon - \delta$ proofs. The meager (obvious) conclusions hardly justify the huge mental efforts!

Complex Analysis was a different story. Even though officially "continuous", it has the feel of discrete math, and one can "cheat" and consider power series as formal power series, and I really loved it.

Georgy P. Egorychev: A Bridge-Builder between the Discrete and the Continuous

Eight years after I finished my doctorate, I came across Egorychev's fascinating modern classic[E], about using the methods of complex analysis to evaluate (discrete) combinatorial sums. That was a pioneering ecumenical work, that influenced me greatly. Its content, of course, but especially its *spirit* and *philosophy*.

The Discrete vs. The Continuous: A Two-Way Street

Egorychev went from the discrete to the continuous. But the *bridge* that he helped build can be transversed **both** ways. With the advent of so-called **Wilf-Zeilberger** (WZ) theory[WZ] one can indeed go both ways. Sometimes the discrete is easier to handle, and sometimes the continuous. But nothing is really *continuous*. There is only the *discrete* and the "*continuous*", the quotation-marks indicating that it is really discrete in disguise, and, on a fundamental level, continuous mathematics is just a degenerate case of the discrete, as I have already preached in [Z1].

Initially, I was hoping to write something about interfacing Egorychev's brilliant approach with WZ theory, but meanwhile I got distracted by another project, that also has the **discrete-continuous** theme, namely for automatically deriving *limit laws* in probability theory, and decided to make this my *tribute* to Georgy Egorychev's 70th birthday.

[1] Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, USA. `zeilberg at math dot rutgers dot edu` ,
`http://www.math.rutgers.edu/~zeilberg` . First version: Dec. 1, 2008. Accompanied by Maple packages `CLT` and `AsymptoticMoments` downloadable from the webapge of this article
`http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/clt.html`, where one can also find some sample input and output files. Supported in part by the NSF.

Probability Limit Laws

One of the central themes of modern probability theory are *limit laws*, the most celebrated one being the Central Limit Theorem, that roughly says that if you repeat the same experiment many times, and the "atomic" experiment can have an arbitrary probability distribution (with finite variance), then in the limit, after one "centralizes" and "normalizes" (divides by the so-called standard deviation) one gets the (continuous) *Standard Normal Distribution*:

$$Pr(a \leq X \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx \quad .$$

The *iconic* example of a discrete probability distribution is the random variable "number of Heads" upon tossing a (loaded) coin n times, whose probability distribution is given by the **Binomial Distribution**, usually denoted by $\mathbf{B}(n,p)$. It describes the experiment of tossing a coin n times with the probability of Heads being p. The "sample space" is the set of all 2^n outcomes $\{H,T\}^n$, and the probability of an "atomic" event is $p^{NumberOfHeads}(1-p)^{NumberOfTails}$, and hence the probability of the "compound event", *NumberOfHeads*$=k$, is $\binom{n}{k}p^k(1-p)^{n-k}$. If we call this random variable X_n, then its mean (see below) is np and its variance (also see below) is $\sigma^2 := np(1-p)$. Introducing the *centralized and normalized* random variable

$$Z_n := \frac{X_n - np}{\sqrt{np(1-p)}} \quad ,$$

The "original" (De Moivre-Laplace) Central Limit Theorem asserts that

$$Z_n \to \mathcal{N} \quad ,$$

where $\mathcal{N}$ is the Standard Normal Distribution.

More generally, quoting from Feller[F], p.244:

Central Limit Theorem. Let $\{X_k\}$ be a sequence of mutually independent random variables with a common distribution. Suppose that $\mu := \mathbf{E}[X_k]$ and $\sigma^2 := Var[X_k]$ exist and let $\mathbf{S}_n = X_1 + \ldots + X_n$. Then for every fixed β,

$$\mathbf{P}\left\{ \frac{\mathbf{S}_n - n\mu}{\sigma\sqrt{n}} < \beta \right\} \to \mathcal{N}(\beta) \quad ,$$

where $\mathcal{N}(x)$ is the normal distribution defined above.

There are many extensions and generalizations. In this article we will present yet another extension, but in a completely different direction, and because of the heavy use of computers, we are pretty sure that these are new results.

A Quick Review of Discrete Probability Distributions

The most basic scenario is that we have a *finite* set S, called the *sample space*, consisting of *atomic events*, and each $s \in S$ has a certain probability (a number in $[0,1]$) attached to it, where, of course, $\sum_{s \in S} p_s = 1$.

We also have a *random variable* $X : S \to R$, where R is a finite set of real numbers (often, but not always, of integers), and one is interested in its *probability* distribution $Pr(\{s \in S | X(s) = r\})$. A convenient way to encode it is via its, *probability generating function* ,

$$f(t) := \sum_{r \in R} Pr(X(s) = r)\, t^r \quad ,$$

that is easily seen to be equal to the *weighted counting* of the set S

$$\sum_{s \in S} p_s t^{X(s)} \quad .$$

The most important number associated to a random variable is its *expectation*

$$\mu = E[X] := \sum_{s \in S} p_s X(s) \quad .$$

This is also called the *first* moment. Analogously, the *higher moments* (about the mean) are defined by

$$m_r(X) := \sum_{s \in S} p_s (X(s) - \mu)^r \quad .$$

It follows from "general nonsense" that, under some mild conditions (that are always satisfied for finite sets), the moments completely determine the probability distribution (even in the general, "infinite", case), and the probability distribution can be gotten by inverse-Fourier-Transforming the **moment (exponential) generating function** $\sum_r m_r (it)^r / r! = E[exp(itX)]$.

Another set of moments, easier to work with, are the *factorial moments*

$$f_r(X) := \sum_{s \in S} p_s (X(s) - \mu)^{(r)} \quad ,$$

where $X^{(r)}$ is the *falling factorial*:

$$X^{(r)} := X(X-1)(X-2)\dots(X-r+1) \quad .$$

It turns out to be easier (see below) to compute the factorial moments, but once these are known, one can get the ordinary moments, thanks fo the **connection formula** (e.g. [GKP], p. 250):

$$X^r = \sum_{k=1}^{r} S(r,k) X^{(k)} \quad ,$$

where $S(r,k)$ are the *Stirling Numbers of the Second kind*, that may be defined by the recurrence ([GKP], p. 250):

$$S(r,k) = kS(r-1,k) + S(r-1,k-1) \quad , \qquad (StirlingRecurrence)$$

subject to the initial condition $S(1,k) = 1$ if $k = 1$ and $S(1,k) = 0$ otherwise.

If follows that the moments can be computed in terms of the factorial moments:

$$m_r = \sum_{k=1}^{r} S(r,k) f_r \quad .$$

Computing Moments

Suppose that we have the probability generating function $f(t)$. We can find its *mean*, μ, by differentiating with respect to t, and plugging-in $t = 1$:

$$\mu = f'(1) \quad .$$

Immediately we can find the probability generating function of the centralized random variable $X_C(s) := X(s) - \mu$. It is simply

$$\frac{f(t)}{t^\mu} \quad .$$

From now, let's assume that all our random variables have mean 0, in other words, assume that we have already done this centralization, and let's rename it $f(t)$. Using the new, adjusted, $f(t)$, we can easily find the factorial moments, by taking successive derivatives, and substituting $t = 1$ at the end:

$$f_r = \left. \frac{d^r f(t)}{dt^r} \right|_{t=1} \quad .$$

Alternatively, we can consider $f(1+z)$ and do a Maclaurin expansion around $z = 0$:

$$f(1+z) = \sum_{r=0}^{\infty} f_r \frac{z^r}{r!} \quad .$$

Repeating It n times

So far what we said is true in general. A frequently occurring situation is when we repeat something n times, like tossing a coin, or rolling a die, and we are interested in the sum of the outcomes. In that case, we have a sequence of random variables whose probability generating function is

$$F(t)^n \quad ,$$

where $F(t)$ is the probability generating function for the single event. For example, for tossing a single coin, where the random variable is "number of Heads", and the probability of a Head is p, we have

$$F(t) = \frac{pt + (1-p)}{t^p} = p\,t^{1-p} + (1-p)\,t^{-p} \quad ,$$

and for rolling a loaded (cubic) die, with its probabilities of landing on $1, 2, 3, 4, 5, 6$ being $p_1, p_2, p_3, p_4, p_5, p_6$ respectively, (where of course $p_1 + \ldots + p_6 = 1$), is

$$F(t) = \frac{\sum_{i=1}^{6} p_i t^i}{t^\mu} \quad , \quad where \quad \mu := \sum_{i=1}^{6} i p_i \quad ,$$

etc.

To get the first R factorial moments, for any specific, desired R, we simply find the first $R+1$ terms in the Taylor expansion of $F(t)^n$, at $t=1$, that Maple can easily do symbolically, getting *explicit* polynomial expressions, in n, for the r-th factorial moment, for each specific, numeric, r. What it *can't* do is find the general expression for *symbolic* r (as well as n, of course).

An even more efficient way to crank-out explicit polynomial expressions for the factorial moments, is to, *once and for all*, crank out sufficiently many coefficients of $F(1+z)$ itself (equivalently find sufficiently many factorial moments of the "atomic" experiment), let's call them F_i, where, of course, $F_0 = 1$ and $F_1 = 0$.

$$F(1+z) = 1 + \sum_{r=2}^{\infty} \frac{F_r}{r!} z^r \quad ,$$

and then use the obvious fact that

$$F(1+z)^{n+1} = F(1+z)^n \cdot F(1+z)$$

that entails:

$$1 + \sum_{r=2}^{\infty} \frac{f_r(n+1)}{r!} z^r = \left(1 + \sum_{r=2}^{\infty} \frac{f_r(n)}{r!} z^r\right) \left(1 + \sum_{r=2}^{\infty} \frac{F_r}{r!} z^r\right) \quad .$$

Rearranging, and comparing coefficient of z^r, we have the following *recurrence*

$$f_r(n+1) - f_r(n) = \sum_{s=2}^{r} \binom{r}{s} F_s f_{r-s}(n) \quad , \qquad (Recurrence)$$

Since obviously $f_r(0) = 0$, this uniquely determines $f_r(n)$ as the indefinite sum of the right side, and it immediately follows by induction that the even factorial moments $f_{2r}(n)$ are polynomials of degree r, and the odd factorial moments $f_{2r+1}(n)$ are also polynomials of degree r. (Of course $f_1(n) = 0$).

Asymptotic Factorial Moments

There is no way that we can get an *explicit*, symbolic, expression, in **both** n and r for the general factorial moments $f_{2r}(n), f_{2r+1}(n)$. But, thanks to the miracle of computers, we can get **explicit** expressions for their s-leading terms *for any desired s*.

Either "cheating" and using our knowledge that the normalized even factorial moments $f_{2r}(n)/f_2(n)^r$ should tend to the even moments $(2r)!/(2^r r!)$ of the Standard Normal Distribution, and the normalized odd factorial moments $f_{2r}(n)/f_2(n)^{r+1/2}$ should tend to the odd moments (0) of the Standard Normal Distribution, but better still, doing it *ab initio*, by staring at the leading terms and making the obvious conjectures, we can write:

$$f_{2r}(n) = f_2(n)^r \frac{(2r)!}{2^r r!} \left[\left(1 + \sum_{i=1}^{s} \frac{A_i(r)}{n^i}\right) + O(\frac{1}{n^{s+1}}) \right] \quad ,$$

and, analogusly

$$f_{2r+1}(n) = f_2(n)^r \frac{(2r)!}{2^r r!} \left[\left(\sum_{i=0}^{s} \frac{B_i(r)}{n^i} \right) + O(\frac{1}{n^{s+1}}) \right] \quad .$$

(Note that $f_2 = nF_2$).

Substituting this *ansatz* into $(Recurrence)$, it emerges that the $A_i(r)$'s and $B_i(r)$'s are certain polynomials in r. Rather than untangle the complicated implied recurrences for them, we empirically, in turn, for $i = 0, 1, 2, \ldots$, crank-out $A_i(r), B_i(r)$ for sufficiently many numeric r and then "fit" appropriate polynomials, using *undetermined coefficients* in the context of the *polynomial ansatz* (see [Z2]). Once we have the *conjectured* explicit expressions, for the asymptotic expansion up to our desired order $(1/n^s)$, we can, *a posteriori*, prove them *rigorously* by verifying $(Recurrence)$ to that desired order.

The Central Limit Theorem only asserts that the normalized r-th moments converge to the moments of the Standard Normal Distributions, i.e. the case $s = 0$. So in particular, our computer reproved the Central Limit Theorem, but with a *vengeance*, it gave us the first s terms in the asymptotics, where s is as big as we wish (of course the higher the s, the longer that it would take).

What about the ordinary moments?

From

$$m_r = \sum_{k=1}^{r} S(r,k) f_r \quad ,$$

we get:

$$m_r(n) = \sum_{k=0}^{s} S(r, r-k) f_{r-k}(n) + O(\frac{1}{n^{s+1}}) \quad .$$

Define

$$S_k(r) := S(r, r-k) \quad .$$

It is easy to see that $S_k(r)$ are polynomials in r of degree $2k$. Indeed the defining recurrence $(StirlingRecurrence)$ transcribes to:

$$S_k(r) - S_k(r-1) = (r-k) S_{k-1}(r-1) \quad ,$$

from which Maple can easily compute, recursively, as many of the $S_k(r)$ as needed, starting at the obvious initial condition $S_0(r) = 1$, and taking the indefinite sum, with respect to r, of the already known right hand side.

So to get the up-to-order-s asymptotics for the ordinary moment $m_r(n)$, Maple simply computes, *all by itself*,

$$m_r(n) = \sum_{k=0}^{s} S_k(r) f_{r-k}(n) + O(\frac{1}{n^{s+1}}) \quad ,$$

using the already computed expressions (in *symbolic* r and n) for f_{2r} and f_{2r+1} obtained above (up to the desired order s). Of course, we would have to treat the even moments, m_{2r}, and the odd moments m_{2r+1} separately, and they obviously have different expressions, but the computer does not mind.

Repeating n times a Generic Probability Distribution

The above discussion applies equally to repeating a general probability distribution, given by its ordinary moments $M_1 = 0$, $M_2 = 1$, $M_3, M_4 \dots$. One first finds the factorial moments (now using the Stirling numbers of the *first* kind), and using the above formula, one can get the asymptotics of the moments of the "repeated" n-times random variable, to any desired order s, of the $2r$-th and $(2r+1)$-th moments for the normalized sum of n repetitions. For example, the first term is:

$$1 + \frac{(-1+r)\, r\, \left(2\, r M_3{}^2 + 3\, M_4 - 9 - 4\, M_3{}^2\right)}{18n} + O(\frac{1}{n^2}) \quad .$$

More terms are available at the webpage of this article.

This leads us to the following interesting observation, that, once made, could be easily proved using moment generating functions.

Refined Central Limit Theorem. Let $\{X_k\}$ be a sequence of mutually independent random variables with a common distribution. Suppose that $\mu := \mathbf{E}[X_k] = 0$ and $\sigma^2 := E[X^2] = 1$, and all the first $2s$ moments, $M_1 = 0, M_2 = 1, M_3, M_4, \dots, M_{2s}$, are finite. Let $\mathbf{S}_n = X_1 + \dots + X_n$, and let $m_{2r}(n)$ be the $2r$-th moment of $\mathbf{S}_n$. Then for even s,

$$m_{2r}(n) = (2r)!/(2^r r!)(1 + O(1/n^s))$$

if and only of the first $2s$ moments of X are the same as the first $2s$ moments of the Standard Normal Distribution (namely: 0,1,0,3,0,15,0,105, ...) .

Limit Laws for Sequences of Discrete Probability Distributions

The Central Limit Theorem talks about the limit of a family of discrete probability distributions, whose probability generating functions are given by the extremely simple

$$P_n(t) := F(t)^n \quad ,$$

that satisfy a first-order recurrence with *constant*, (in n) coefficients

$$P_{n+1}(t) = F(t) P_n(t) \quad .$$

Many natural families of discrete probability distributions, especially those arising from generating functions in combinatorial enumeration ("q-counting"), satisfy a more general kind of first-order recurrence:

$$P_{n+1}(t) = F(n, t, t^n) P_n(t) \quad ,$$

where $F(n, t, t^n)$ is a certain explicit rational function of n, t, t^n.

For example (switching to the letter q to respect combinatorial tradition), consider the set of permutations on n elements, under the "mahonian" statistics, whose *counting* generating function is:

$$\prod_{i=1}^{n} \frac{1-q^i}{1-q} \quad .$$

The expectation is, of course, $n(n-1)/4$, so dividing by $n!q^{n(n-1)/4}$, we get that the probability generating function for the random variable "number of inversions" is:

$$P_n(q) = \prod_{i=1}^{n} \frac{q^{-i/2} - q^{i/2}}{i(q^{-1/2} - q^{1/2})} \quad .$$

So, in this case,

$$F(n, q, q^n) = \frac{q^{-(n+1)/2} - q^{(n+1)/2}}{(n+1)(q^{-1/2} - q^{1/2})} \quad .$$

See [F](sec. X.6)[LP][M][S] for other approaches for proving Asymptotic Normality.

Another example is the q-Catalan distribution, whose asymptotic normality has been recently proved by Chen, Wang, and Wang [CWW], who also proved more general results.

The discussion in the previous section goes almost verbatim to such more general families of discrete probability distributions, except that now we can no longer (always) find the first factorial moments directly. Now we *must* use the generalization of (*Recurrence*).

Instead of

$$F(1+z) = 1 + \sum_{r=2}^{\infty} \frac{F_r}{r!} z^r \quad ,$$

we now have:

$$F(n, 1+z, (1+z)^n) = 1 + \sum_{r=2}^{\infty} \frac{F_r(n)}{r!} z^r \quad ,$$

where now the $F_r(n)$ are polynomials of n, and no longer have an interpretation as factorial moments for an "atomic event". They are just the Maclaurin coefficients of this more general object. The analog of (*Recurrence*) reads:

$$f_r(n+1) - f_r(n) = \sum_{s=2}^{r} \binom{r}{s} F_s(n) f_{r-s}(n) \quad . \qquad (General Recurrence)$$

The same empirical approach as before still applies. We normalize, and guess explicit expressions for the coefficients $A_i(r), B_i(r)$ in the normalized factorial moments, that are then rigorously proved *a posteriori*. Once explicit asymptotic expressions for the even and odd factorial moments have been derived and proved, one uses the Stirling polynomials in order to deduce explicit expressions for the even and odd (usual) moments (about the mean), in particular proving *asymptotic normality*, but with precise asymptotic expansion, to any desired order, of the general moments.

Accompanying Maple Packages

This article is accompanied by two Maple packages: `CLT`, and `AsymptoticMoments`. Most of the procedures in `CLT` are subsumed by the more general procedures of `AsymptoticMoments`, but the former has some extra features.

These packages can be downloaded from the webpage of this article:

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

where there is ample sample input and output.

In particular, `AsymptoticMoments` is applied to the above-mentioned cases of the mahonian and q-Catalan distribution, thereby sharpening them, by not only proving asymptotic normality, but presenting a more detailed asymptotics for the moments. We also post numerous other examples, for example, plane partitions whose 3D Ferrers diagrams are bounded in a box of any given, (numeric) height.

Let us cite the simplest output. If you toss a fair coin n times, then the $2r$-th moment is $(n/4)^r(2r)!/(2^r r!)$ times

$$1-1/3\,\frac{(r-1)\,r}{n}+\frac{1}{90}\,\frac{r\,(r-1)\,(r-2)\,(5\,r+1)}{n^2}-\frac{1}{5670}\,\frac{(r-1)\,(r-2)\,(r-3)\,\left(35\,r^2+21\,r-32\right)r}{n^3}+O(\frac{1}{n^4}) \quad .$$

Conclusion: Why is this interesting?

Locally, it is interesting for its own sake, but *globally* it is interesting since it presents a beautiful example how probability theory would have been very different, had the computer been available three hundred years ago. Using *symbol-crunching* the computer can derive deep theorems, and largely obviates all the human attempts at a "rigorous" foundation of continuous probability, using measure theory and Kolmogorov's "axiomatic" approach. The passage from the discrete to the continuous becomes much more concrete and down-to-earth, and it is apparent that Discrete Math rules, and Continuous Math is indeed a degenerate case. For other examples of probability computerized-redux, see [Z3].

Future Work

This is just the tip of an iceberg. One should be able to consider much larger families of discrete probability distributions, not just those given by first-order recurrences. Also *joint distributions*, and *multivariate* limit laws should be amenable to the present approach. For example, proving the joint asymptotic normality of the number of inversions and the *major index* on the set of permutations on $\{1,2,\ldots,n\}$, using the more complicated recurrences derived in [Z4].

References

[CWW] W. Y.C. Chen, C. J. Wang, and L.X.W. Wang, *The Limiting Distribution of the coefficients of the q-Catalan Numbers*, Proc. Amer. Math. Soc. **136** (2008), 3759-3767.

[E] G.P. Egoryechev, "*Integral Representation and the Computation of Combinatorial Sums*", (translated from the Russian), Amer. Math. Soc., Providence, 1984.

[F] William Feller, "*An Introduction to Probability Theory and Its Application*", volume 1, three editions. John Wiley and sons. First edition: 1950. Second edition: 1957. Third edition: 1968.

[GKP], R. Graham, O. Patashnik, and D.E. Knuth, *Concrete Mathematics: A Foundation for Computer Science*, Addison Wesley, Reading, 1989.

[LP] G. Louchard and H. Prodinger, *The number of inversions in permutations: a saddle point approach*, J. Integer Seq. **6** (2003), A03.2.8.

[M] B.H. Margolius, *Permutations with inversions*, J. Integer Seq. **4** (2001), A01.2.4.

[S] V.N. Sachkov, "*Probabilistic Methods in Combinatorial Analysis*", Cambridge University Press, New York, 1997.

[WZ] H. Wilf and D. Zeilberger, *An algorithmic proof theory for hypergeometric (ordinary and "q") multisum/integral identities*, Invent. Math. **108** (1992), 575-633.

[Z1] Doron Zeilberger, *"Real" Analysis is a Degenerate Case of Discrete Analysis.* In: "New Progress in Difference Equations"(Proc. ICDEA 2001), edited by Bernd Aulbach, Saber Elaydi, and Gerry Ladas, Taylor & Francis, London, 2004. Available on-line from:
`http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/real.html` .

[Z2] Doron Zeilberger, *An Enquiry Concerning Human (and Computer!) [Mathematical] Understanding.* In: C.S. Calude ,ed., "Randomness & Complexity, from Leibniz to Chaitin", World Scientific, Singapore, Oct. 2007. Available on-line from:
`http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/enquiry.html` .

[Z3] Doron Zeilberger, *Fully AUTOMATED Computerized Redux of Feller's (v.1) Ch. III (and Much More!)*, Personal Jouranl of Ekhad and Zeilberger (Nov. 14, 2006),
`http://www.math.rutgers.edu/~zeilberg/pj.html` .

[Z4] Doron Zeilberger, *A Lattice Walk Approach to the "inv" and "maj" q-counting of Multiset Permutations*, J. Math. Anal. Applications **74** (1980), 192-199.