HISTABRUT: A Maple Package for Symbol-Crunching in Probability theory

Doron ZEILBERGER[1]

Abstract: A Maple package `HISTABRUT` (available from `http://www.math.rutgers.edu/~zeilberg/tokhniot/HISTABRUT`) is presented and briefly described. It uses the polynomial ansatz to discover (often fully rigorously, but in some cases only semi-rigorously (yet rigorizably!)) explicit asymptotic formulas for the moments of uni-variate and, more impressively, bi-variate, discrete probability random variables. It would be hopefully extended, in the future, to *multi-variate* random variables.

Many sequences of discrete random variables (e.g. tossing a (fair or loaded) coin n times, and keeping track of the number of Heads minus the number of Tails) are *asymptotically normal.* In [Z1], I introduced and described Maple packages, `CLT` and `AsymptoticMoments`, that empirically-yet-rigorously prove asymptotic normality for a wide class of sequences of discrete random variables. They used the method of moments. Furthermore, they are able to prove much stronger theorems than mere "asymptotic normality" by finding the asymptotics (to any desired order!) of the (normalized) moments, rather than only the leading terms (that should be those of the normal distribution $e^{-x^2/2}/\sqrt{2\pi}$, namely $1 \cdot 3 \cdot 5 \cdots (2r-1)$ for the even $2r$-th moment, and 0 for the odd moments).

But not *all* discrete probability random variables are asymptotically normal! For example, the number of times that your current capital is positive, upon tossing a fair coin n times and winning a dollar if it is Heads and losing a dollar it is Tails, that converges to Paul Lévy's *arcsine distribution* (see [Z2]), and the other random variables considered by Feller (see [Z3]). Another intriguing random variable is the *duration* of a *gambler's ruin* considered in [Z4].

The much larger Maple package `HISTABRUT` available from

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

can handle *any* sequence of discrete probability distributions, that the users have to program themselves. There are quite a few ones pre-programmed, (type `EzraPGF();` in the Maple package `HISTABRUT` for a list). It can also sketch the limiting distributions, using procedure `plotDist` (see the on-line help).

Another new feature is that it can handle *directly* sequences of probability distributions defined in terms of rational generating functions, $R(t, s)$, where the coefficient of s^n in the power-series expansion of $R(t, s)$ in terms of s is the probability generating function (in t) for a typical memebr

[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: Aug. 25, 2010. This article accompanies the Maple package `http://www.math.rutgers.edu/~zeilberg/tokhniot/HISTABRUT` . Exclusively published in the Personal Journal of Shalosh B. Ekhad and Doron Zeilberger (`http://www.math.rutgers.edu/~zeilberg/pj.html`). Supported in part by the NSF.

of a sequence of random variables parametrized by n. For example, for tossing a fair coin n times $R(t,s) = 1/(1-s(t+1/t)/2)$. Recall that the Goulden-Jackson[GJ] method (beautifully exposited and extended in [NZ]), and also included in `HISTABRUT`, outputs such rational functions for the random variable "number of occurrences of a prescribed (consecutive) subword". First `HISTABRUT` quickly and effortlessly computes *explicit* (symbolic) expressions for the mean and variance. This is no big deal, and even *you*, my dear human readers, can probably do it in many cases. Having done this easy task, `HISTABRUT` goes on and computes the (normalized) even and odd ($2r$-th and $(2r+1)$-th respectively), to *any* desired order, as expressions in **both** n and r. Now this is really impressive, and a triumph to experimental mathematics. It first "just" guesses such expressions, but *a posteriori*, just by (fully rigorous!) "hand-waving" justifies its guesses, by saying that checking a certain number of special cases suffices to prove the conjectured explicit formulas rigoroulsy. The justification is that at the *end of the day*, everything boils down to *polynomial identities*, and we all know that two polynomials of degree $\leq d$ are identically equal if they coincide in $d+1$ different values. In particular it, in any given case, *rigorously* reproves the well-known fact that the distribution is asymptotically normal, but *in addition* supplies much more information, by outputting higher-order asymptotics.

But the most salient new feature is the handling of sequences of *bi-variate* discrete random variables, for example the number of occurrences of two different words as (consecutive) subwords. Here it only gives polynomial expressions, in n, for the (r,s)-mixed moments, for $r,s \leq R$, and R is a numeric positive integer inputted by the user, but is unable (yet) to find general expressions in terms of r and s. Here, too, the Goulden-Jackson method, that is built-in, supplies lots of examples.

Whenever the sequence of bivariate discrete probability distributions happens to be asymptoically *independently* normal, procedure `AnalyseMoms2` can find explicit expressions, in n, of course, but also in *both* r, and s, for the asymptotic order, to any desired order, for the normalized mixed (r,s) moments. [More precisely, it finds four distinct expressions for the $(2r,2s)$, $(2r+1,2s)$, $(2r,2s+1)$, and $(2r+1,2s+1)$ mixed moments.]

Sample input and output

The "front" of the present article

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

has numerous sample input and output files. The readers are welcome to edit the input files in order to produce their own output.

Future Directions

Procedure `AnalyseMoms2` (and the verbose version `AnalyseMoms2V`) can only handle sequences of bivariate discrete distributions that are asymptotically *independently* normal. It would be nice to extend it to pairs of random variables that are non-independently asymptotically normal. This would first require finding the asymptotic correlation (already done!), and then finding expressions for the mixed moments for the limiting continuous bi-variate distributions $exp(-x^2/2 - y^2/2 + bxy)$ where $c = b/(1-b^2)$ is the limiting correlation coefficient. These are all things that I know how to teach the computer how to do, but I currently don't have time.

Another worthwhile extension is to consider *tri*-variate, *quad*-variate, and in general, *multi*-variate sequences of discrete probability distributions.

References

[GJ] Ian Goulden and David M. Jackson, *An inversion theorem for cluster decompositions of sequences with distinguished subsequences*, J. London Math. Soc.(2)**20** (1979), 567-576.

[NZ] John Noonan and Doron Zeilberger, *The Goulden-Jackson cluster method: extensions, applications, and implementations*, J. Difference Eq. Appl. **5** (1999), 355-377.
Available from `http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/gj.html` .

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

[Z2] Doron Zeilberger, *A new proof that there are* 2^n *ways to toss a coin n times*, Personal Journal of Shalosh B. Ekhad and Doron Zeilberger,
`http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/toss.html` .

[Z3] Doron Zeilberger, *Fully AUTOMATED Computerized Redux of Feller's (v.1) Ch. III (and Much More!)*, Personal Journal of Shalosh B. Ekhad and Doron Zeilberger,
`http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/feller.html` .

[Z4] Doron Zeilberger, *Symbol Crunching with the Gambler's Ruin Problem*, "Tapas in Experimental Mathematics" (Tewodros Amdeberhan and Victor Moll, eds.), Contemporary Mathematics, **457** (2008), 285-292.
Available from `http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/ruin.html` .