*The Mathematica® Journal*

# *The Combinatorics of Distributions*

**Bernd Günther**

Distributions, which are the various ways of distributing a certain number of objects of different classes among a collection of targets, have been the subject of combinatorial investigations since MacMahon's 1917 monograph. In this paper we apply them to a simulation of superimposed random coding. Furthermore, asymptotic estimates are provided using logarithmic polynomials (related to the well-known Bell polynomials) for symbolic and numeric calculation.

## Introduction

We denote by $B(m, w, k)$ the number of ways $m$ classes of $w$ indistinguishable objects can be distributed among $k$ different boxes, leaving no box empty and such that no box contains more than one object from the same class. In MacMahon's terms [1] (cf. also [2]) we speak of the distribution of objects of specification $\{w^m\}$ into boxes of specification $\{1^k\}$.

Relaxing the requirement that no box should remain empty gives $\binom{k}{w}^m$ distributions. Since each subcollection consisting of $j$ boxes has $B(m, w, j)$ permissible distributions, we obtain

$$\binom{k}{w}^m = \sum_{j=w}^{k} \binom{k}{j} B(m, w, j). \tag{1}$$

Equation (1) can be readily solved:

$$B(m, w, k) = \sum_{j=w}^{k} (-1)^{k-j} \binom{k}{j} \binom{j}{w}^m, \tag{2}$$

which holds for $m > 0$ with $B(1, w, k) = \delta_{w\,k}$ for $m = 1$.

```
numdistrib[m_Integer, w_Integer, k_Integer] :=
  Sum[(-1)^(k - j) Binomial[k, j] Binomial[j, w]^m, {j, w, k}] /; m > 0
```

It is convenient to extend the definition to the case $m = 0$ by setting $B(0, w, k) = \delta_{k0}$, where, of course, we tacitly assume $w > 0$.

```
numdistrib[0, w_Integer, k_Integer] := KroneckerDelta[k, 0]
```

For example, with $m = 2, w = 3, k = 4$, there are 12 admissible distributions.

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | {1, 2} | {1, 2} | {1} | {2} |
| 2 | {1, 2} | {1} | {1, 2} | {2} |
| 3 | {1} | {1, 2} | {1, 2} | {2} |
| 4 | {1, 2} | {1, 2} | {2} | {1} |
| 5 | {1, 2} | {1} | {2} | {1, 2} |
| 6 | {1} | {1, 2} | {2} | {1, 2} |
| 7 | {1, 2} | {2} | {1, 2} | {1} |
| 8 | {1, 2} | {2} | {1} | {1, 2} |
| 9 | {1} | {2} | {1, 2} | {1, 2} |
| 10 | {2} | {1, 2} | {1, 2} | {1} |
| 11 | {2} | {1, 2} | {1} | {1, 2} |
| 12 | {2} | {1} | {1, 2} | {1, 2} |

```
numdistrib[2, 3, 4]
```

```
12
```

Applications of distributions are abundant. Markov [3] considered a lottery performing regular drawings of $w$ out of $n$ lots, and asked for the probability $p_k$ that the number of lots that have shown up at least once after $m$ drawings equals $k$. Here the lots take over the role of the boxes, which are occupied by an object of class $j$ if it is drawn in the $j^{\text{th}}$ round. Hence

$$p_k = B(m, w, k) \binom{n}{k} \binom{n}{w}^{-m}. \tag{3}$$

*Superimposed Random Coding* [4, 5] is an application to coding theory. Consider a collection of randomly generated binary code words of length $n$ and weight $w$ (i.e., the number of 1-bits equals $w$). A set of code words $c_1, \ldots, c_m$ can be combined by superposition $c = c_1 \bigvee \cdots \bigvee c_m$, so that $c$ contains a 1-bit in all those places where at least one of the $c_i$ has one. Notice that $c$ has weight $k$ with probability $p_k$ as in (3). For "decoding," we tentatively assume that the $j^{\text{th}}$ code word is present if its 1-bits are among those of $c$. This way

we will certainly not miss any correct matches but may encounter a few false ones; the error probability must be investigated.

In practice we are testing for the presence of a set of code words $c_1', \ldots, c_s'$ simultaneously by checking if its superposition $c' = c_1' \bigvee \cdots \bigvee c_s'$ is covered by $c$. $c'$ is of weight $\ell$ with probability

$$p_\ell' = B(s, w, \ell)\binom{n}{\ell}\binom{n}{w}^{-s}. \tag{4}$$

Since there are $\binom{n}{\ell, k-\ell, n-k}$ pairs of subsets of size $\ell$ and $k$ (the first one contained in the second), the probability of a purely accidental match evaluates to

$$P_{\text{error}} = \sum_{\ell \le w} \binom{n}{\ell, k-\ell, n-k}\binom{n}{\ell}^{-1}\binom{n}{k}^{-1} p_\ell' \, p_k. \tag{5}$$

More details and a simulation are given in the supplemental notebook, which is available from www.mathematica-journal.com/data/uploads/2011/07/CombDistrib3Suppl.nb.

## Asymptotics

Even for moderate values of $m$, $w$, and $k$, the number of distributions $B(m, w, k)$ increases very quickly and soon exceeds the possibilities of ordinary 64-bit arithmetic. Fortunately, an asymptotic formula is available.

**Theorem 1**

*If $k, w \to \infty$ such that the ratio $\frac{k}{m\,w} = t$ remains constant with $0 < t < 1$ and if $m$ either remains constant or also $m \to \infty$, but $m \in O(w^\alpha)$ for suitable $\alpha < \frac{1}{5}$, then*

$$B(m,\ w,\ k) \sim \frac{m^{m\,w}\left(\left(\frac{1}{m(1-\zeta)} - \frac{1}{m}\right)\left(\frac{1}{1-\zeta^m} - 1\right)^{-\frac{1-\zeta^m}{m(1-\zeta)}}\right)^{m\,w}}{\sqrt{(2\,\pi\,w)^m\,\zeta^m\left(1 - \frac{1-\zeta^m}{m(1-\zeta)}\,\zeta^{m-1}\right)}}, \tag{6}$$

*where the new parameter $0 < \zeta < 1$ is defined by*

$$\frac{1-\zeta^m}{m(1-\zeta)} = \frac{1}{m}\sum_{j=0}^{m-1}\zeta^j = t. \tag{7}$$

**Sketch of Proof**

The number of distributions can be computed by induction on $m$ as

$$B(m+1, w, k) = \sum_{j=k-w}^{k} \binom{k}{k-j, w-k+j, k-w} B(m, w, j), \tag{8}$$

which readily leads to

$$B(m, w, k) = \sum_{w=j_1 \le j_1 \le \cdots \le j_m = k} \prod_{i=2}^{m} \binom{j_i}{j_i - j_{i-1}, w - j_i + j_{i-1}, j_i - w}. \tag{9}$$

We resolve the multinomial coefficient as a product of factorials, apply Stirling's formula, and approximate the sum by an integral over the variables $\xi_i = \frac{j_i}{w}$:

$$B(m, w, k) \sim \int \cdots \int_{\xi_2 \le \cdots \le \xi_{m-1}} \varphi_w(\xi_2, \cdots, \xi_{m-1})\, d\xi_2 \cdots d\xi_{m-1}, \tag{10}$$

$$\varphi_w(\xi_2, \cdots, \xi_{m-1}) =$$

$$(2\pi w)^{-\frac{m-1}{2}} \sqrt{\prod_{i=2}^{m} \frac{\xi_i}{(\xi_i - \xi_{i-1})(1 - \xi_i + \xi_{i-1})(\xi_i - 1)}}\ (\psi(\xi_2, \cdots, \xi_{m-1}))^w, \tag{11}$$

$$\psi(\xi_2, \cdots, \xi_{m-1}) = \prod_{i=2}^{m} \frac{\xi_i^{\xi_i}}{(\xi_i - \xi_{i-1})^{\xi_i - \xi_{i-1}} (1 - \xi_i + \xi_{i-1})^{1 - \xi_i + \xi_{i-1}} (\xi_i - 1)^{\xi_i - 1}}. \tag{12}$$

The function $\psi$ has a global maximum in the interior of the integration domain at $\xi_i^* = \frac{1-\zeta^i}{1-\zeta}$ with $\zeta$ chosen as in the theorem. Under such circumstances, $\psi^w$ behaves like a Dirac delta distribution as $w \to \infty$:

$$(\psi(\xi_2, \cdots, \xi_{m-1}))^w \sim \beta_w\, \delta(\xi_2 - \xi_2^*, \ldots, \xi_{m-1} - \xi_{m-1}^*), \tag{13}$$

$$\beta_w = \sqrt{\frac{(2\pi)^{m-2}\, \psi(\xi_2^*, \ldots, \xi_{m-1}^*)^{m+2w-2}}{w^{m-2} \left| \frac{\partial^2 \psi}{\partial \xi_i\, \partial \xi_j} \right|}}, \tag{14}$$

where $\left| \frac{\partial^2 \psi}{\partial \xi_i\, \partial \xi_j} \right|$ denotes the absolute value of the Hessian of $\psi$ at the location of the maximum. The validation of our estimates requires the interchange of limit operations at several instances, which can be justified if $m \in O(w^\alpha)$. More details of the proof cannot be given within the present scope. We recommend [6] as general reference for asymptotics.

Notice that the maximal possible value $t = 1$ corresponding to $k = m\,w$, where every object is assigned to a different box, is not covered by the theorem.

We are interested in the limit case $m \to \infty$, where the number of classes used to typify our objects is very large; imagining that the different classes are distributed one after another

($m$ can be interpreted as the time parameter of a Markov chain), this means that we want to anticipate the far future. However, the case $m \to \infty$, though included in the theorem, is not handled very conveniently, most notably because of the necessity of solving the $(m-1)$-degree polynomial equation (7); some other deficiencies are described in the supplemental notebook.

Setting $\varrho = m\,(1-\zeta)$ and $\varepsilon = \frac{1}{m}$ in equation (7) is equivalent to

$$\log[1 - t\,\varrho] = \frac{1}{\varepsilon}\,\log[1 - \varepsilon\,\varrho], \tag{15}$$

which is an analytic equation, since the singularity at $\varepsilon = 0$ is inessential. By the implicit function theorem $\varrho$ can be expressed as a power series in $\varepsilon$, and we are going to compute the coefficients. The logarithmic polynomials will be required as tools, for whose introduction we make a small digression.

## □ Logarithmic Polynomials

We recall that the *logarithmic polynomials* $L_n[x_1, \ldots x_n]$ are defined by

$$L_n[x_1, \ldots x_n] = \sum_{k=1}^{n} (-1)^{k-1}\,(k-1)!\,B_{n,k}[x_1, \ldots x_{n-k+1}], \tag{16}$$

where $B_{n,k}$ denotes the Bell polynomials familiar from Faà di Bruno's formula [7]. Observe that $L_n$ is a polynomial in $n$ indeterminates with integer coefficients. Their sequence satisfies the condition

$$\log\Big[1 + \sum_{n=1}^{\infty} \frac{x_n\,t^n}{n!}\Big] = \sum_{n=1}^{\infty} \frac{t^n\,L_n[x_1, \ldots x_n]}{n!}. \tag{17}$$

The following implementation is basically an adaption of [8], exercise 16b (we are indebted to the referee).

```
logpoly[x_List] :=
 Module[{n, listOfPartitions, monomial, homdeg, aFactor},
   n = Length[x];
   listOfPartitions = Map[Tally, IntegerPartitions[n]];
   aFactor[anEntry_] :=
    (x[[anEntry[[1]]]] / anEntry[[1]]!)^(anEntry[[2]]) /
     anEntry[[2]]!;
   monomial[aPartition_] :=
    (homdeg = Plus @@ Map[Last, aPartition];
     (-1)^(homdeg - 1) n! (homdeg - 1)!
      Times @@ Map[aFactor, aPartition]);
   Total@ Map[monomial, listOfPartitions]] /; Length[x] > 0
```

Here are the first five log polynomials.

```
TableForm[Table[logpoly[Take[{x1, x2, x3, x4, x5}, n]],
  {n, 5}], TableHeadings → Automatic]
```

| | |
|---|---|
| 1 | $x_1$ |
| 2 | $-x_1^2 + x_2$ |
| 3 | $2\,x_1^3 - 3\,x_1\,x_2 + x_3$ |
| 4 | $-6\,x_1^4 + 12\,x_1^2\,x_2 - 3\,x_2^2 - 4\,x_1\,x_3 + x_4$ |
| 5 | $24\,x_1^5 - 60\,x_1^3\,x_2 + 30\,x_1\,x_2^2 + 20\,x_1^2\,x_3 - 10\,x_2\,x_3 - 5\,x_1\,x_4 + x_5$ |

Substituting $t^k\,x_k$ for $x_k$ in (17) one easily derives

$$L_n\left[t\,x_1, t^2\,x_2, \ldots, t^n\,x_n\right] = t^n\,L_n[x_1, \ldots, x_n], \tag{18}$$

thus the logarithmic polynomials are not homogeneous but of uniform weight, where the variable $x_k$ counts with weight $k$. Furthermore, taking partial derivatives with respect to the highest argument,

$$\frac{\partial}{\partial x_n}\,L_n[x_1, \ldots x_n] \equiv 1, \tag{19}$$

which means that $L_n$ depends additively on the last variable:

$$L_n[x_1, \ldots x_n] = L_n[x_1, \ldots x_{n-1}, 0] + x_n. \tag{20}$$

## □ Asymptotics, Resumed

Thus equipped, we set out to solve equation (15) for $\varrho = \sum_{n=0}^{\infty} \frac{\varrho_n[t]\,\varepsilon^n}{n!}$. The constant coefficient $\varrho_0[t]$ can be obtained from (15) by taking limits:

$$\log(1 - t\,\varrho_0) = -\varrho_0. \tag{21}$$

It will be convenient to introduce $\eta = \exp(-\varrho_0) = 1 - t\,\varrho_0$, $0 < \eta < 1$, as a new independent parameter, so that

$$t = -\frac{1-\eta}{\log\eta}. \tag{22}$$

This puts us in a position to put (15) into a form where (17) can be applied, namely where the constant coefficients of the involved power series equal 1:

$$\log\left[1 - \sum_{n=1}^{\infty} \frac{t\,\varrho_n\,\varepsilon^n}{\eta\,n!}\right] = \frac{1}{\varepsilon}\,\log\left[1 - \sum_{n=1}^{\infty} \frac{\varrho_{n-1}\,\varepsilon^n}{(n-1)!}\right] + \varrho_0. \tag{23}$$

By comparing coefficients we obtain

$$L_n\Big[-\frac{t\,\varrho_1}{\eta},\ldots,-\frac{t\,\varrho_n}{\eta}\Big]=\frac{1}{n+1}\,L_{n+1}[-\varrho_0,-2\,\varrho_1,\ldots,-(n+1)\,\varrho_n]. \tag{24}$$

It will be seen presently that this determines $\varrho_n$ as a rational expression in $t$ and $\eta$; we observe that these two quantities, transcendentally related by (22), are algebraically independent. To factor out the denominator, we set $\varrho_n=\frac{\eta\,(1-\eta)^{n+1}}{t^{n+1}\,(t-\eta)^{2\,n-1}}\,\sigma_n$ with $\sigma_0=\frac{1}{\eta\,(t-\eta)}$. Observing (18), we obtain

$$\begin{aligned}&L_n[-(t-\eta)\,(1-\eta)\,\sigma_1,\ldots,-(t-\eta)\,(1-\eta)\,\sigma_n]\ =\\&\frac{(1-\eta)}{(n+1)\,t\,(t-\eta)^2}\,L_{n+1}\Big[-(t-\eta)^2,-2\,\eta\,(t-\eta)^3\ \sigma_1,\ldots,-(n+1)\,\eta\,(t-\eta)^3\ \sigma_n\Big].\end{aligned} \tag{25}$$

Now from (20):

$$\begin{aligned}\sigma_n=&\frac{t}{(1-\eta)\,(t-\eta)^2}\,L_n[-(t-\eta)\,(1-\eta)\,\sigma_1,\ldots,-(t-\eta)\,(1-\eta)\,\sigma_{n-1},0]\\&-\frac{1}{(n+1)\,(t-\eta)^4}\,L_{n+1}\Big[-(t-\eta)^2,-2\,\eta\,(t-\eta)^3\ \sigma_1,\ldots,-n\,\eta\,(t-\eta)^3\ \sigma_{n-1},0\Big].\end{aligned} \tag{26}$$

By uniformity of weight (18), the denominators cancel, thus revealing $\sigma_n$ for $n\geq 1$ as a polynomial.

```
sigmapoly[k_, η_, t_] :=
 sigmapoly[k, η, t] =
  Collect[
   Simplify[
    t
      logpoly[Append[Table[-(t - η) (1 - η) sigmapoly[j, η, t],
          {j, 1, k - 1}], 0]] / ((1 - η) (t - η)^2) -
     logpoly[
       Append[Table[If[j == 1, -(t - η)^2,
          -j η (t - η)^3 sigmapoly[j - 1, η, t]], {j, 1, k}],
        0]] / ((k + 1) (t - η)^4)], t]
```

Here are the first three cases of $\sigma_n$.

```
TableForm[Table[sigmapoly[k, η, t], {k, 3}],
 TableHeadings → Automatic]
```

| | |
|---|---|
| 1 | $\frac{1}{2}$ |
| 2 | $\frac{2\,t^2}{3}+\frac{1}{12}\,t\,(-3-\eta)-\frac{\eta^2}{3}$ |
| 3 | $-t^3+\frac{3\,t^4}{2}+\frac{\eta^4}{4}+\frac{1}{8}\,t^2\,\left(1-4\,\eta-15\,\eta^2\right)+\frac{1}{4}\,t\,\eta\,\left(1+3\,\eta+2\,\eta^2\right)$ |

We can now check the correctness of our coefficients by inserting them into equation (15).

```
maxDim = 5;
rhotest =
 (1 - η) / t +
  Sum[(η * (1 - η) ^ (k + 1) * sigmapoly[k, η, t] * ε^k) /
     (k ! t^(k + 1) * (t - η) ^ (2 * k - 1)) , {k, maxDim}];
Simplify[Log[1 - t rhotest + O[ε]^(maxDim + 1)] -
  Log[1 - ε rhotest + O[ε]^(maxDim + 2)] / ε]
```

$$\left(\frac{1-\eta}{\mathtt{t}} + \mathtt{Log}[\eta]\right) + \mathtt{O}[\varepsilon]^6$$

This is 0 because of (22). Let us summarize the results achieved so far.

**Lemma 1**

*The unique solution $\zeta > 0$ of the equation $\frac{1-\zeta^m}{m(1-\zeta)} = t$ is given by the power series*

$$\zeta = 1 - \frac{1-\eta}{m\,t} - \sum_{n=1}^{\infty} \frac{\eta\,(1-\eta)^{n+1}\,\sigma_n}{n!\,t^{n+1}\,(t-\eta)^{2n-1}\,m^{n+1}}, \tag{27}$$

*with $\eta$ determined from $t = -\frac{1-\eta}{\log \eta}$ and with the polynomials $\sigma_n$ determined inductively by* (26).

The rest now is fairly standard. We observe that in (6),

$$\log\left(\left(\frac{1}{m(1-\zeta)} - \frac{1}{m}\right)\left(\frac{1}{1-\zeta^m} - 1\right)^{-\frac{1-\zeta^m}{m(1-\zeta)}}\right) = \log\left(\frac{1}{\varrho} - \varepsilon\right) - t\log\left(\frac{1}{t\,\varrho} - 1\right]\right), \tag{28}$$

and we can easily compute a power series expansion of the expression $\log\left(\frac{1}{\varrho} - \varepsilon\right) - t\log\left(\frac{1}{t\,\varrho} - 1\right) = \sum_{n=0}^{\infty} \alpha_n\,\varepsilon^n$ from the one of $\varrho$.

```
maxDim = 5;
rhotest =
 (1 - η) / t +
  Sum[(η * (1 - η) ^ (k + 1) * sigmapoly[k, η, t] * ε^k) /
     (k ! t^(k + 1) * (t - η) ^ (2 * k - 1)) , {k, maxDim}];
pbase =
 Simplify[Log[1 / rhotest - ε + O[ε]^(maxDim + 1)] -
   t Log[1 / (t rhotest) - 1 + O[ε]^(maxDim + 1)]];
α[n_] := Coefficient[pbase, ε, n];
```

The first two coefficients are as follows.

```
Exp[α[0]] /. t → -(1 - η) / Log[η]
```

$$-\frac{\left(-\frac{\eta}{-1+\eta}\right)^{-\frac{-1+\eta}{\text{Log}[\eta]}}}{\text{Log}[\eta]}$$

```
Exp[α[1]] /. t → -(1 - η) / Log[η]
```

$$\sqrt{\eta}$$

This provides us with complete control over the numerator of (6). To handle the denominator we observe $\zeta^m = 1 - t\,m\,(1-\zeta)$ and therefore, from (27), $\lim_{m\to\infty} \zeta^m = \eta$. We also observe $\lim_{m\to\infty} \zeta = 1$, because $\zeta$ depends implicitly on $m$. In consequence,

$$\lim_{m\to\infty} \sqrt{\zeta^m \left(1 - \frac{1-\zeta^m}{m(1-\zeta)}\,\zeta^{m-1}\right)} = \sqrt{\eta(1-t\,\eta)}\,. \tag{29}$$

In using this limit approximation we accept a slight penalty of accuracy in favor of a simpler formula. Assembling all parts, we have proved the following theorem.

**Theorem 2**

*If $k, m, w \to \infty$ such that the ratio $\frac{k}{m\,w} = t$ remains constant with $0 < t < 1$ and $m \in O(w^\alpha)$ for suitable $\alpha < \frac{1}{5}$, then*

$$B(m,\ w,\ k) \sim \sqrt{\frac{\eta^{w-1}}{(2\,\pi\,w)^m\,(1-t\,\eta)}} \left(m\,\frac{\left(\frac{1-\eta}{\eta}\right)^{\frac{1-\eta}{-\log\eta}}}{-\log\eta}\right)^{m\,w} \exp\left(\sum_{n=2}^{\infty} \frac{\alpha_n\,w}{m^{n-1}}\right), \tag{30}$$

*with $\eta$ determined from $t = -\frac{1-\eta}{\log\eta}$ and with the power series coefficients $\alpha_n$ as above.*

Since $m$ is not supposed to grow faster than $\sqrt[5]{w}$, at least the first five terms of the power series in (30) will be significant, but higher terms may be negligible depending on the actual growth rate.

The derivation of error bounds for the asymptotic relation (30) is beyond our present scope. A numerical example showing that the approximation is rather good is contained in the supplemental notebook.

## References

[1] P. A. MacMahon, *Combinatory Analysis,* 3rd ed. reprint, Providence, RI: AMS Chelsea, 2001.

[2] J. Riordan, *Introduction to Combinatorial Analysis*, Mineola, NY: Dover Publications, 2002.

[3] A. A. Markov, *Wahrscheinlichkeitsrechnung*, Leipzig, Berlin: Teubner, 1912. catalog.hathitrust.org/Record/012107751.

[4] B. Günther, "On the Probability Distribution of Superimposed Random Codes," *IEEE Transactions on Information Theory*, **54**(7), 2008 pp. 3206–3210. doi:10.1109/TIT.2008.924658.

[5] C. S. Roberts, "Partial-Match Retrieval via the Method of Superimposed Codes," *Proceedings of the IEEE*, **67**(12), 1979 pp. 1624–1642. ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1455812.

[6] N. G. de Bruijn, *Asymptotic Methods in Analysis*, Amsterdam: North-Holland Publishing Co.; New York: Interscience Publishers, 1958.

[7] L. Comtet, *Advanced Combinatorics—The Art of Finite and Infinite Expansions*, rev. and enl. ed. (J. W. Nienhuys, trans.), Dordrecht, Boston: D. Reidel Publishing Company, 1974 p. 140.

[8] M. Trott, *The Mathematica GuideBook for Numerics*, Berlin: Springer, 2005.

B. Günther, "The Combinatorics of Distributions," *The Mathematica Journal*, 2011. dx.doi.org/doi:10.3888/tmj.13–10.

### About the Author

The author obtained his Ph.D. in mathematics at the University of Frankfurt, Germany in 1989 and has lectured and done research in Frankfurt and Seattle, Washington. Since then he has worked for Oracle, the Beilstein Institute, and currently for Deutsche Bahn AG. He has published several papers in pure and applied mathematics.

**Bernd Günther**
*DB-Systel GmbH*
*Helpertseestrasse 21*
*63165 Muehlheim*
*Germany*
*dr.bernd.guenther@t-online.de*