Binomial Tails Domination for Random Graphs via Bell Polynomials

Andrzej Korzeniowski
University of Texas at Arlington

ABSTRACT  We obtain an inequality between tails of binomial distributions by establishing a combinatorial identity for a sum involving binomial and multinomial coefficients associated with the Bell polynomials.  As an application a connection to random graphs is presented.

Keywords  Bell polynomials, Di Bruno formula, Random graphs.

1. Introduction

As remarked in [5] the task of reducing combinatorial sums to a single closed form expression is an art in itself.  Indeed, despite various algorithms and techniques developed over the years, quite often one encounters a situation in which the known methodology does not apply and requires a novel approach.  This paper is an example of such a case in which a combinatorial identity emerged as a by-product of studying polynomials associated with binomial random graphs.  Namely, let $Q(x)$ be a binomial-type polynomial defined by

$$Q(x) = \sum_{i=0}^{m} \binom{n}{i} (x^k - 1)^i, \quad 0 \leq m \leq n, \quad k \geq 1$$

and consider

$$Q^j(1) = \frac{d^j Q(x)}{dx^j} \bigg|_{x=1}, \quad 0 \leq j \leq m. \quad (2)$$

Evaluating $Q^j(1)$ when $k = 1$ is trivial however for $k \geq 2$ the matter is far less obvious.
Consequently it is of interest to find an explicit representation for $Q^{(j)}(1)$ and explore the possibility of reducing it to a simple closed form. It turns out that the first part can be achieved by applying di Bruno formula for differentiation of composite function while the second part amounts to showing that so obtained representation can in fact be simplified to a single expression referred to as the combinatorial identity for Bell polynomials.

2. The Bell Polynomials Identity

We begin by recalling some facts concerning the Bell polynomials [6]. Let $B_j(x_1, \ldots, x_j)$ be the $j$-th exponential Bell polynomial defined by the generating function

$$e^{\sum_{i=1}^{\infty} x_i t^i} = 1 + \sum_{j=1}^{\infty} B_j(x_1, \ldots, x_j) \frac{t^j}{j!}$$

where

$$B_j(x_1, \ldots, x_j) = \sum_{l_1+2l_2+\ldots+j_l=j} \left( \prod_{i=1}^{j} \frac{j!}{l_1! l_2! \cdots l_j!} x_1^{l_1} \cdots x_j^{l_j} \right) \prod_{i=1}^{j} l_i!$$

and \{l_1, l_2, \ldots, l_j\} are all nonnegative integer solutions of the Diophantine equation $l_1 + 2l_2 + \ldots + jl_j = j$. Below are the first few polynomials

$$B_1(x_1) = x_1$$
$$B_2(x_1, x_2) = x_2 + x_1^2$$
$$B_3(x_1, x_2, x_3) = x_3 + 3x_2x_1 + x_1^3$$
$$B_4(x_1, x_2, x_3, x_4) = x_4 + (4x_3x_1 + 3x_2^2) + 6x_2x_1^2 + x_1^4$$
$$B_5(x_1, x_2, x_3, x_4, x_5) = x_5 + (5x_4x_1 + 10x_3x_2) + (10x_3x_1^2 + 15x_2x_1) + 10x_2^2x_1 + x_1^5$$

where the right hand side is written as a sum $B_{j,1}(x_1, \ldots, x_j) + \ldots + B_{j,j}(x_1, \ldots, x_j)$, with parenthesis used as needed, of the associated partial Bell polynomials \{B_{j,l}(x_1, \ldots, x_j), l = 1, \ldots, j\}. Notice that for $x_1 = \ldots = x_j = x$, $B_j(x_1, \ldots, x_j)$ turns into the classical Bell polynomial and when $x \equiv 1$ then $B_{j,1}(1, \ldots, 1) = S(j, 1)$, $B_j(1, \ldots, 1) = \sum_{l=1}^{j} S(j, l) = B_j$ which are the Stirling numbers of the second kind and the Bell numbers respectively corresponding to all possible partitions of a $j$-element set into $l$ blocks of its non-empty disjoint subsets, $l = 1, \ldots, j$. Furthermore, di Bruno formula for the $j$-th derivative of the composite function $f(g(x))$ reads

$$f(g(x))^{(j)} = \sum_{l_1+2l_2+\ldots+j_l=j} \frac{j!}{(l_1)! (l_2)! \cdots (l_j)!} (g^{(1)})^{l_1} (g^{(j)})^{l_j} f^{(l_1+l_2+\ldots+l_j)}$$
= \sum_{l=1}^{j} \left( \sum_{l_1+l_2+...+l_l=j \atop l_1+2l_2+...+jl_j=j} \frac{j!}{(1)l_1!(j)l_j!} (g^{(1)})^{l_1}...g^{(j)})^{l_j} \right) \ f^{(l)}
\equiv \sum_{l=1}^{j} f^{(l)} B_{j,l}(g^{(1)},...,g^{(j)})

where the last equality comes from (3) while \( f^{(l)} = \frac{d^lf(x)}{dy^l}|_{y=g(x)} \) and \( g^{(l)} = \frac{d^lg(x)}{dx^l}, l = 1,2,...j. \)

**Lemma 1** Let \((k)_l = k(k-1)...(k-l+1)\) denote the falling factorial. Then for any positive integers \(i, j, k\) we have

\[
((x^k - 1)^j)^{(j)}(1) = \begin{cases} 
  i!B_{j,i}((k)_1,...,(k)_j), & j \geq i \\
  0, & j < i 
\end{cases}
\]  

**(Proof)** Applying (4) for \( f(y) = y^i \) and \( y = g(x) = x^k - 1 \) gives

\[
((x^k - 1)^j)^{(j)} = \sum_{l=1}^{j} f^{(l)} B_{j,l}(g^{(1)},...,g^{(j)})
\]

where \( g^{(l)} = (k)_l x^{k-l} \) for \( l \leq k \) and \( g^{(l)} = 0 = (k)_l \) for \( l > k. \) Since

\[
f^{(l)} = ((x^k - 1)^i)^{(l)} = \begin{cases} 
  (i)_l(x^k - 1)^{i-l}, & l = 1,...,j, \quad j < i \\
  (i)_l(x^k - 1)^{i-l}, & l = 1,...,i, \quad j \geq i \\
  0, & l = i + 1,...,j, \quad j \geq i 
\end{cases}
\]

therefore

\[
f^{(l)}(1) = \begin{cases} 
  0, & l = 1,...,j, \quad j < i \\
  i!, & l = i, \quad l \leq j, \quad j \geq i \\
  0, & l \neq i, \quad l \leq j, \quad j \geq i 
\end{cases}
\]

which combined with (6) yields

\[
((x^k - 1)^j)^{(j)}(1) = \begin{cases} 
  i!B_{j,i}((k)_1,...,(k)_j), & j \geq i \\
  0, & j < i 
\end{cases}
\]

and the proof is complete.

**Lemma 2** Let \( 1 \leq j \leq m \leq n, \ k \geq 1 \) and \( Q(x) \) and \( Q^{(j)}(1) \) be as defined by (1)-(2). Then

\[
Q^{(j)}(1) = (nk)_j
\]  

(7)
Proof. Since
\[ x^{nk} = (1 + x^k - 1)^n = \sum_{i=0}^{m} \binom{n}{i} (x^k - 1)^i + \sum_{i=m+1}^{n} \binom{n}{i} (x^k - 1)^i \]
it follows
\[ Q(x) = x^{nk} - \sum_{i=m+1}^{n} \binom{n}{i} (x^k - 1)^i = x^{nk} - (x - 1)^{m+1} P(x) \]
whence Leibnitz’s differentiation \((x - 1)^{m+1} P(x)^{(j)}(1) = 0\) for \(1 \leq j \leq m\) gives
\[ Q^{(j)}(1) = (nk)_j \]
as desired.

**Theorem 1** (Bell polynomials identity). For any positive integers \(i, j, k\) the following combinatorial identity holds true
\[ \sum_{i=1}^{j} \binom{n}{i} \sum_{i_1+i_2+\ldots+i_j = i} \binom{i}{i_1, i_2, \ldots, i_j} \binom{k}{i_1} \binom{k}{i_2} \ldots \binom{k}{i_j} = \binom{nk}{j} \] (8)
or equivalently upon multiplying (8) by \(j!\)
\[ \sum_{i=1}^{j} (n)_i B_{j,i}((k)_1, \ldots, (k)_j) = (nk)_j \] (9)
where \(B_{j,i}(x_1, \ldots x_j) = \sum_{i_1+i_2+\ldots+i_j = i} \frac{j!}{(i_1)! i_1! \ldots (i_j)! i_j!} x_1^{i_1} \ldots x_j^{i_j}\) are the partial Bell polynomials defined in (3).

Proof. It suffices to verify (9) but it is a straightforward consequence of Lemma 1 and Lemma 2 because for \(1 \leq j \leq m \leq n\) we have
\[ (nk)_j = Q^{(j)}(1) = \sum_{i=1}^{m} \binom{n}{i} (x^k - 1)^i (1) = \sum_{i=0}^{j} \binom{n}{i} (x^k - 1)^i (1) \]
\[ = \sum_{i=1}^{j} (n)_i B_{j,i}((k)_1, \ldots, (k)_j) \]
This concludes the proof of identities (8)-(9).

### 3. Domination of Binomial Tails

In the sequel we shall focus on comparison of random variables \(X\) and \(Y\) in the following [7] sense.
Definition  We say that \( X \) is stochastically larger than \( Y \) (\( X \) dominates \( Y \)), written \( X \gtrsim Y \), if the tail probability of \( X \) dominates the tail probability of \( Y \) according to
\[
P(X > a) \geq P(Y > a) \text{ for all } a, \text{ written } F \gtrsim G
\]
where \( F \) and \( G \) are the distributions of \( X \) and \( Y \) respectively.

Notice that for nonnegative integer-valued random variables, thanks to \( P(X \geq 0) = P(Y \geq 0) = 1 \), (10) is equivalent to \( P(X \geq k) \geq P(Y \geq k), k = 1, 2, \ldots \) and implies \( \text{EX} = \sum_{k \geq 1} P(X \geq k) \geq \sum_{k \geq 1} P(Y \geq k) = EY \) for the expected values.

In what follows \( X \sim b(m, \theta) \) stands for binomial distribution with parameter \( \theta \), i.e., \( P(X = i) = \binom{m}{i} \theta^i (1 - \theta)^{m-i}, i = 0, 1, \ldots, m \).

**Theorem 2**  Let \( X \sim b(nk, p) \) and \( Y \sim b(n, 1 - (1 - p)^k) \) then \( X \gtrsim Y \) or equivalently \( b(nk, p) > b(n, 1 - (1 - p)^k) \) (11) for any positive integers \( n, k \) and \( 0 \leq p \leq 1 \).

**Proof.**  Since \( p = 1 \) is trivial, it suffices to show that for \( 0 \leq p < 1, 0 \leq m \leq n, k \geq 1 \)
\[
P(X \geq m + 1) = \sum_{i = m+1}^{nk} \binom{nk}{i} p^i (1 - p)^{nk-i} \geq P(Y \geq m + 1)
\]
\[
= \sum_{i = m+1}^{n} \binom{n}{i} (1 - (1 - p)^k)^i (1 - p)^{n-i}
\]
or equivalently, by \( P(X \leq m) = 1 - P(X \geq m + 1) \leq 1 - P(Y \geq m + 1) = P(Y \leq m) \),
\[
\sum_{i=0}^{m} \binom{n}{i} (1 - (1 - p)^k)^i (1 - p)^{n-i} \geq \sum_{i=0}^{m} \binom{nk}{i} p^i (1 - p)^{nk-i} \tag{12}
\]
Dividing (12) by \( (1 - p)^{nk} \), setting \( x = \frac{1}{1-p} \geq 1 \), and using (1) gives
\[
Q(x) = \sum_{i=0}^{m} \binom{n}{i} (x^k - 1)^i \geq \sum_{i=0}^{m} \binom{nk}{i} (x-1)^i, \ x \geq 1 \tag{13}
\]
To verify (13), following the proof of Lemma 2, we express \( Q(x) \) in two different ways
\[
Q(x) = x^{nk} - \sum_{i = m+1}^{n} \binom{n}{i} (x^k - 1)^i = \sum_{j=0}^{m} a_j (x-1)^j + \sum_{j=m+1}^{nk} a_j (x-1)^j \tag{14}
\]
where the last equality is the representation of the \( mk \)-th degree polynomial \( Q(x) \) in the basis \( \{(x-1)^j, j = 0, 1, \ldots, mk\} \). Now by Taylor’s expansion combined with (2) and (7) of Lemma 2 we have
\[
a_j = \frac{Q^{(j)}(1)}{j!}, \ j = 0, 1, \ldots, mk \text{ and } a_j = \binom{nk}{j}, \ j = 0, 1, \ldots, m \tag{15}
\]
whence (1), (14)-(15) gives
\[
\sum_{i=0}^{m} \binom{n}{i} (x^i - 1)^j = \sum_{i=0}^{m} \binom{mk}{i} (x - 1)^i + \sum_{j=m+1}^{mk} a_j (x - 1)^j, \quad x \geq 1
\] (16)
which proves (13) because the second sum in the right hand side of (16) is nonnegative thanks to
\[
a_j = \frac{Q^{(j)}(1)}{j!} = \frac{\left(\sum_{i=0}^{m} \binom{n}{i} (x^i - 1)^j\right)(1)}{j!} \geq 0, \quad j = m + 1, \ldots, mk
\]
which follows from the fact that \((x^i - 1)^j\) is a polynomial with positive integer coefficients satisfies \(B_{j,i}(x_1, \ldots, x_j) \geq 0\) whenever \(x_1, \ldots, x_j \geq 0\).

Before we present an application of (10)-(11) domination, let us remark that it extends to the case where \(G\) is a conditional distribution of \(Y\) given a random vector, say \(Z\), since \(F \sim G\) is the distribution property.

4. Application to Random Graphs

For a reference on the subject of random graphs the reader may consult a monograph [1] while some recent findings concerning binomial random graphs and their representations can be found in [3], [4]. A binomial random graph \(G[V, p]\) on the set of vertices \(V = \{v_1, \ldots, v_N\}\) is built by placing a connecting edge between any pair of \(\binom{N}{2}\) vertices independently at random with probability \(p\) and placing no edge with probability \(1 - p\). Let \(Z = \#\) edges in \(G[V, p]\), then \(Z\) has the binomial distribution \(b\left(\binom{N}{2}, p\right)\) according to \(P(Z = i) = \binom{\binom{N}{2}}{i} p^i (1 - p)^{\binom{N}{2} - i}, i = 0, 1, \ldots, \binom{N}{2}\).

Let \(V_1\) and \(V_2\) be two disjoint non-empty subsets of \(V\) such that \(|V_1| = n, |V_2| = k, n + k \leq N\). A bipartite random graph \(G[V, V_1, V_2, p]\) is built by placing connecting edges between \(nk\) pairs of vertices from \(V_1\) and \(V_2\) with probability \(p\). We define two random variables associated with \(G[V, V_1, V_2, p]\) First of which.

\[
X = \# \text{ of edges between } V_1 \text{ and } V_2
\] (17)
has the binomial distribution \(b(nk, p)\) whereas the second is defined as follows. For any vertex \(v \in V_1\) \(P\)(there is a connecting edge from \(v\) to at least one vertex from \(V_2\)) = \(1 - (1 - p)^k\). Then

\[
Y = \# \text{ of vertices in } V_1 \text{ that connect to at least one vertex from } V_2
\] (18)
has the binomial distribution \(b\left(n, 1 - (1 - p)^k\right)\).

**Theorem 3** Consider \(G[V, p]\) on the set of vertices \(V = \{v_1, \ldots, v_n\}\). Given \(x \in V\), let \(\Gamma_k(x) = \{y \in V \mid d(x, y) = k\}\), where \(d(x, y)\) is the distance between the vertices \(\{x, y\}\) (\(=\) length of the shortest chain of edges connecting \(x\) and \(y\)) and let \(N_k(x) = \bigcup_{i=0}^{k} \Gamma_k(x) \subset V\). Let
$Z = (|\Gamma_k(x)|, |N_k(x)|)$, $Y = |\Gamma_{k+1}(x)|$ and $G(a) = G_{Y|Z}(a|Z) = P(Y \leq a|Z)$ be the conditional distribution. Then we have the domination

$$b\left(|\Gamma_k(x)|(n - |N_k(x)|)), p\right) > G_{|\Gamma_{k+1}(x)|, |(\Gamma_k(x), N_k(x))|}
= b\left(n - |N_k(x)|, 1 - (1 - p)^{|\Gamma_k(x)|}\right)$$

for $0 \leq k < n$, $0 \leq p \leq 1$.

**Proof.** Apply Theorem 2 with $|V_1| = n - |N_k(x)|$, $|V_2| = |\Gamma_k(x)|$ and observe that the distribution of $|\Gamma_{k+1}(x)|$, given $(|\Gamma_k(x)|, |N_k(x)|)$, corresponds to the bipartite random graph $G[V, V \setminus N_k(x), \Gamma_k(x), p]$. □

A special case of domination (19) under specific restrictions on the range of $p = p(n)$ was a basis for several results in [2]. Our theorem shows that (19) holds true universally without any restrictions on $0 \leq p \leq 1$ whatsoever and may pave the way toward new probability estimates for the upper bound of $\Gamma_k(x)$ for the associated random graphs.

**References**