Math 162: Differentiating Identities

Steven J. Miller
Brown University

April 20, 2008

Abstract

We give some examples of differentiating identities to prove formulas in probability theory and combinatorics. The main result we prove concerns the number of alternating strings of heads and tails in tossing a coin. Specifically, if we toss a coin $n_1 + n_2$ times and see n_1 heads and n_2 tails, the mean of the number of runs is $\frac{2n_1n_2}{n_1+n_2} + 1$ and the variance is $\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$. For example, if we observed $HHHTHHTTTTHTT$ then $n_1 = 6$, $n_2 = 7$ and there would be 6 alternating strings or 6 runs.

More generally, assume we toss a coin with probability p of heads a total of N times. The expected number of runs is $2p(1-p)(N-1)+1$. In particular, if the coin is fair (so $p = \frac{1}{2}$) then the expected number of runs is $\frac{N+1}{2}$.

Contents

1 Simple Examples

We give a standard example illustrating the key idea of differentiating identities. Assume, for some reason (perhaps because of the tantalizing simplicity of the expression), that we want to evaluate

$$\frac{1}{1}+\frac{2}{2}+\frac{3}{4}+\frac{4}{8}+\frac{5}{16}+\frac{6}{32}+\frac{7}{64}+\cdots. \tag{1.1}$$

After some thought we might realize that this is the same as

$$\sum_{n=0}^{\infty}\frac{n}{2^{n-1}}. \tag{1.2}$$

The series does converge by the comparison test (for n large, compare $\frac{n}{2^n}$ to $\frac{1}{(3/2)^n}$).

Abstraction actually helps us. It is easier to study

$$\sum_{n=0}^{\infty} n\cdot x^{n-1}. \tag{1.3}$$

Using the comparison test, one can show this series converges for $|x|<1$. If we didn't have the n's above, the series would be easily summable: the geometric series formula gives

$$\sum_{n=0}^{\infty} x^n \;=\; \frac{1}{1-x}. \tag{1.4}$$

If we could differentiate both sides of the above equation *and* interchange the order of summation and differentiation we would have

$$\begin{aligned}\frac{d}{dx}\sum_{n=0}^{\infty}x^n &= \frac{d}{dx}\frac{1}{1-x}\\ \sum_{n=0}^{\infty}\frac{d}{dx}x^n &= \frac{1}{(1-x)^2}\\ \sum_{n=0}^{\infty}nx^{n-1} &= \frac{1}{(1-x)^2}.\end{aligned} \tag{1.5}$$

Now all we have to do is take $x=\frac{1}{2}$ above to solve the original problem. For this problem, as long as $|x|<1$ we can justify interchanging the order of summation and differentiation. See Appendix E for some results about interchanging orders of differentiation and summation.

The above is a standard example of **Differentiating Identities**. We give an interesting application of a related problem in Appendix C; namely, by considering a finite geometric sum and differentiating the resulting identity we obtain formulas for the sums of powers of integers. Typically such formulas are proved by induction; this presents an alternative approach. As another example, in Appendix D we use this method to show that the harmonic series (the sum of the reciprocals of the integers) diverges.

We give another common example, this time from basic probability. Consider a binomial distribution with n trials, where each trial has probability p of being a success (coded as 1) and probability $1-p$ of being a failure (coded as 0). For example, consider n tosses of a coin with probability p of heads and $1-p$ of tails. Thus

$$\text{Prob}(k) \;=\; \begin{cases} \binom{n}{k}p^k(1-p)^{n-k} & \text{if } k \in \{0, 1, \dots, n\} \\ 0 & \text{otherwise.} \end{cases} \tag{1.6}$$

What is the expected number of successes (or heads)? What is the variance? One simple way to solve this is by linearity of expectation. Namely, consider n independent trials, where X_i is a random variable denoting the outcome of the i^{th} trial. Specifically X_i is 1 for a success (which occurs with probability p) and 0 for a failure (which occurs with probability $1-p$). If $X = X_1 + \cdots + X_n$, then X has the binomial distribution with parameters n and p and we have

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X_1 + \cdots + X_n] \\ &= \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]. \end{aligned} \tag{1.7}$$

As

$$\mathbb{E}[X_i] \;=\; 1 \cdot p + 0 \cdot (1-p), \tag{1.8}$$

we find that

$$\mathbb{E}[X] \;=\; np. \tag{1.9}$$

Similarly, using

$$\text{Var}(X) \;=\; \text{Var}(X_1) + \cdots + \text{Var}(X_n) \tag{1.10}$$

and

$$\text{Var}(X_i) \;=\; \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 \;=\; \left(1^2 \cdot p + 0^2 \cdot (1-p)\right) - (p)^2 \;=\; p(1-p), \tag{1.11}$$

we see that

$$\text{Var}(X) \;=\; np(1-p). \tag{1.12}$$

We now show how these formulas can be derived by differentiating identities. Similar to the geometric series formulas above, it is much easier to work with a free parameter (such as p) and then set it equal to a desired probability at the end; if we didn't have a free parameter, we couldn't differentiate! Thus, even if a problem gives a particular value for p, it is easier to derive formulas for arbitrary p and then set p equal to the given value at the end. This allows us to use the tools of calculus.

Thus to study the binomial function we should consider

$$(p+q)^n \;=\; \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}. \tag{1.13}$$

(*Aside:* for those knowing moment generating functions, think about connections between moment generating functions and differentiation.) If we take $p \in [0,1]$ and $q = 1-p$, then we have a binomial distribution and $(p+q)^n = 1$. We now differentiate the above with respect to p. While we will eventually set $q = 1-p$, for now we consider p and q independent variables.

In fact, instead of $\frac{\partial}{\partial p}$ we apply $p\frac{\partial}{\partial p}$. The advantage of this is that we do not change the powers of p and q in our expressions, and we find

$$\begin{aligned} p\frac{\partial}{\partial p}\left(\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\right) &= p\frac{\partial}{\partial p}(p+q)^n \\ p\sum_{k=0}^{n}\binom{n}{k}kp^{k-1}q^{n-k} &= p\cdot n(p+q)^{n-1} \\ \sum_{k=0}^{n}k\binom{n}{k}p^k q^{n-k} &= np(p+q)^{n-1}; \end{aligned} \tag{1.14}$$

interchanging the differentiation and summation is trivial to justify because we have a finite sum. The expected number of successes (when each trial has probability p of success) is obtained by now setting $q = 1-p$, which yields

$$\sum_{k=0}^{n}k\binom{n}{k}p^k(1-p)^n \;=\; np. \tag{1.15}$$

To determine the various, we differentiate again. Hence applying the operator $p^2\frac{\partial^2}{\partial p^2}$ to (1.13) gives

$$p^2\frac{\partial^2}{\partial p^2}\left(\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\right) \;=\; p^2\frac{\partial^2}{\partial p^2}(p+q)^n; \tag{1.16}$$

again, we apply $p^2\frac{\partial^2}{\partial p^2}$ as this keeps the powers of p and q the same before and after the differentiation. After some simple algebra we find

$$\sum_{k=0}^{n}k(k-1)\binom{n}{k}p^k q^{n-k} \;=\; p^2\cdot n(n-1)(p+q)^{n-2}. \tag{1.17}$$

Unfortunately, to find the variance we need to study

$$\sum_{k=0}^{n}(k-\mu)^2\binom{n}{k}p^k q^{n-k}, \tag{1.18}$$

where $\mu = np$ is the mean of the binomial random variable X. This is not a serious problem, as we can determine the variance from $\mathbb{E}[X^2]-\mathbb{E}[X]^2$ and write $k(k-1)$ as k^2-k; note the sum of k^2 will be $\mathbb{E}[X^2]$. Thus

$$n(n-1)p^2(p+q)^{n-2} \;=\; \sum_{k=0}^{n}k^2\binom{n}{k}p^k q^{n-k} - \sum_{k=0}^{n}k\binom{n}{k}p^k q^{n-k}. \tag{1.19}$$

But we have already determined the second sum – it is just np when $q = 1-p$. Setting $q = 1-p$ we thus find

$$\sum_{k=0}^{n}k^2\binom{n}{k}p^k(1-p)^{n-k} \;=\; n(n-1)p^2 + np \;=\; n^2p^2 + np(1-p). \tag{1.20}$$

Therefore the variance is just

$$\begin{aligned}\text{Var}(X) &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} - \left(\sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \right)^2 \\ &= n^2p^2 + np(1-p) - (np)^2 \\ &= np(1-p). \end{aligned} \tag{1.21}$$

As a final remark, consider again (1.15) and (1.20). If we set $p = q = \frac{1}{2}$ and then move those factors to the right hand side, we obtain

$$\sum_{k=0}^{n} k \binom{n}{k} \ = \ 2^n, \quad \sum_{k=0}^{n} k^2 \binom{n}{k} \ = \ n(n+1)2^{n-2}. \tag{1.22}$$

Thus we can find nice expressions for sums of products of binomial coefficients and their indices.

Remark 1.1. *It is interesting to note that even if we only want to evaluate sums of integers or rationals, we need to have* continuous *variables so that we can use the tools of calculus.*

Remark 1.2. *Instead of applying $p^2 \frac{\partial}{\partial p}$, it is easier to apply $p \frac{\partial}{\partial p}$ twice. The advantage of this is that we have k^2 coming down and not $k(k-1)$. Specifically, we start with*

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \ = \ (p+q)^n. \tag{1.23}$$

Applying $p \frac{\partial}{\partial p}$ once yields

$$\sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} \ = \ p \cdot n(p+q)^{n-1}. \tag{1.24}$$

Applying $p \frac{\partial}{\partial p}$ again gives

$$\sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} \ = \ p \left[1 \cdot n(p+q)^{n-1} + p \cdot n(n-1)(p+q)^{n-2} \right]. \tag{1.25}$$

By letting $q = 1 - p$ and subtracting the square of the mean, we regain the variance in (1.21).

2 Matching Coefficients

Sometimes we can derive identities of binomial coefficients without differentiating – one common technique is matching coefficients. For example, consider

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \sum_{k=0}^{n} \binom{n}{k} \binom{n}{n-k}, \tag{2.1}$$

because $\binom{n}{k} = \binom{n}{n-k}$. Consider now the following sum

$$\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \cdot \binom{n}{n-k} x^{n-k} y^k, \tag{2.2}$$

as well as

$$(x+y)^n (x+y)^n. \tag{2.3}$$

Expanding the product gives

$$(x+y)^n (x+y)^n \;=\; (x+y)^{2n} \;=\; \sum_{j=0}^{2n} \binom{2n}{j} x^j y^{2n-j}; \tag{2.4}$$

note the coefficient of $x^n y^n$ in this product is $\binom{2n}{n}$. The key observation is that (2.2) is just the $x^n y^n$ term of $(x+y)^{2n}$. This is because it can be interpreted as taking the $x^n y^n$ term of $(x+y)^n (x+y)^n$. How do we get an $x^n y^n$ term from multiplying $(x+y)^n$ with $(x+y)^n$? Well, the two factors $(x+y)^n$ give terms like $\binom{n}{i} x^i y^{n-i}$ and $\binom{n}{j} x^j y^{n-j}$, which are then multiplied together. The only way we get an $x^n y^n$ is when $j = n - i$, and we can do this for any $j \in \{0, 1, \dots, n\}$. Thus the $x^n y^n$ term in $(x+y)^{2n}$ is

$$\binom{2n}{n} x^n y^n \;=\; \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \cdot \binom{n}{n-k} x^{n-k} y^k. \tag{2.5}$$

The proof is completed by taking $x = y = 1$.

The reason arguments like this work is because if we have two polynomials of finite degree in finitely many variables, then if they take on identical values for all values of the parameters then all the coefficients of the two polynomials are equal. This allowed us to take two expressions and equate the coefficients of terms. Without this observation, the equality of two polynomials (at all values of the parameters) would not imply the equality of the coefficients. For example, assume $x^2 + 2xy - 7y = x^2 + 3xy - 5y^2 + y$ for all $x, y \in \mathbb{C}$ (of course these two polynomials are not always equal); however, if this *were* to happen, we would be in trouble as in the first we have $2xy$ and the second we have $3xy$. Thus while some terms (such as x^2) have the same coefficient, others do not.

Specifically, say $F(x, y)$ and $G(x, y)$ are two polynomials of finite degree with complex coefficients. Then if they are equal for all choices of $x, y \in \mathbb{C}$ we have $F(x, y) - G(x, y)$ is a polynomial of finite degree and it is zero for all $x, y \in \mathbb{C}$. It is an easy exercise to show this implies all the coefficients of $F(x, y) - G(x, y)$ are zero (i.e., all the coefficients of $F(x, y)$ equal those of $G(x, y)$). One way to see this is to choose fixed values of x. Say $x = a$. Except for finitely many choices of a, we would get $F(a, y) - G(a, y)$ is a finite degree polynomial and it has some non-zero coefficient but it vanishes for all $y \in \mathbb{C}$. This is absurd as a polynomial of degree d has at most d complex roots. We do not need to have x and y range over all of $\mathbb{C}$; it suffices to have them range over a large enough set, for example $|x|, |y| \le R$ for some $R > 0$.

The biggest difficulty in successfully applying arguments of this nature is figuring out what to compare the observed sum to. Here we needed to see that we should compare $\sum_{k=0}^{n} \binom{n}{k}^2$ to the coefficient of $x^n y^n$ in $(x+y)^{2n}$. Writing $\binom{n}{k}$ as $\binom{n}{k} \cdot \binom{n}{n-k}$ suggests that we should compare it to a coefficient of $(x+y)^n (x+y)^n$.

3 Combinatorics and Partitions

We review some needed results on combinatorics and partitions before tackling the number of alternating strings of coin tosses.

3.1 The Cookie Problem

We describe a combinatorial problem which contains many common features of the subject. Assume we have 10 identical cookies and 5 distinct people. How many different ways can we divide the cookies among the people, such that all 10 cookies are distributed? Since the cookies are identical, we cannot tell which cookies a person receives; we can only tell how many. We could enumerate all possibilities (there are 5 ways to have one person receive 10 cookies, 20 ways to have one person receive 9 and another receive 1, and so on). While in principle we can solve the problem, in practice this computation becomes intractable, especially as the number of cookies and people increase.

We introduce common combinatorial functions. The first is the **factorial function**: for a positive integer n, set $n! = n \cdot (n-1) \cdots 2 \cdot 1$. The number of ways to choose r objects from n when order matters is $n \cdot (n-1) \cdots (n-(r-1)) = \frac{n!}{(n-r)!}$ (there are n ways to choose the first element, then $n-1$ ways to choose the second element, and so on). The **binomial coefficients** $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ is the number of ways to choose r objects from n objects when order does not matter. The reason is once we've chosen r objects, there are $r!$ ways to order them. For convenience, we define $0! = 1$ (thus $\binom{n}{0} = 1$, which may be interpreted as saying there is one way to choose zero elements from a set of n objects). For more on binomial coefficients, see §B.

We show the number of ways to divide 10 cookies among 5 people is $\binom{10+5-1}{5-1}$. In general, if there are C cookies and P people,

Lemma 3.1. *The number of distinct ways to divide C identical cookies among P different people is $\binom{C+P-1}{P-1}$.*

Proof. Consider $C+P-1$ cookies in a line, and number them 1 to $C+P-1$. Choose $P-1$ cookies. There are $\binom{C+P-1}{P-1}$ ways to do this. This divides the cookies into P sets: all the cookies up to the first chosen (which gives the number of cookies the first person receives), all the cookies between the first chosen and the second chosen (which gives the number of cookies the second person receives), and so on. This divides C cookies among P people. Note different sets of $P-1$ cookies correspond to different partitions of C cookies among P people, and every such partition can be associated to choosing $P-1$ cookies as above. $\square$

Remark 3.2. *In the above proof, we do not care* which *cookies a person receives. We introduced the numbers for convenience: now cookies 1 through i_1 (say) are given to person 1, cookies i_1+1 through i_2 (say) are given to person 2, and so on.*

For example, if we have 10 cookies and 5 people, say we choose cookies 3,4,7, and 13 of the 10+5-1 cookies:

$$\odot \ \odot \ \otimes \ \otimes \ \odot \ \odot \ \otimes \ \odot \ \odot \ \odot \ \odot \ \odot \ \otimes \ \odot$$

This corresponds to person 1 receiving 2 cookies, person 2 receiving 0, person 3 receiving 2, person 4 receiving 5, and person 5 receiving 1.

The above is an example of a partition problem: we are solving $x_1 + x_2 + x_3 + x_4 + x_5 = 10$, where x_i is the number of cookies person i receives. We may interpret Lemma 3.1 as the number of ways to divide an integer N into k non-negative integers is $\binom{N+k-1}{k-1}$.

Exercise 3.3. *Show*

$$\sum_{n=0}^{N} \binom{n+k-1}{k-1} = \binom{N+k}{k}. \tag{3.1}$$

One can interpret the above as dividing N cookies among k people, where we do not assume all cookies are distributed. Note here we have a sum of binomial coefficients where both the top and *the bottom index are varying. In general such sums are difficult* unless *you can find a nice way to interpret such a sum.*

Exercise 3.4. *In partition problems, often there are requirements such as everyone receives at least one cookie. How many ways are there to write N as a sum of k non-negative integers? How many solutions of $x_1+x_2+x_3 = 2005$ are there if each x_i is an integer and $x_1 \geq 5$, $x_2 \geq 7$, and $x_3 \geq 1000$?*

3.2 The Alternating Strings Problem

Consider a string of $n_1 + n_2$ coin tosses with n_1 heads and n_2 tails. There are $\binom{n_1+n_2}{n_2}$ ways to order the n_1 heads and n_2 tails. Assume all orderings are equally likely. Our goal is to eventually study the number of alternating strings of heads and tails. We start with a simpler problem, namely trying to figure out how many ways there are to arrange n_1 heads and n_2 tails and observe u runs (again, $HHTTHTTTH$ would have 5 runs and 4 alterations).

For example, let us say $n_1 = n_2 = 3$ and we want to have 3 runs. If we assume we start with a head we could have $HTTTHH$ or $HHTTTH$, and by symmetry if we start with a tail we could have $THHHTT$ or $TTHHHT$.

In general, we have

Theorem 3.5. *Let there be n_1 heads and n_2 tails, and assume each of the $\binom{n_1+n_2}{n_1}$ arrangements are equally likely. Let there be u runs of heads and tails. Then*

$$u = \begin{cases} 2\binom{n_1-1}{k-1}\binom{n_2-1}{k-1} & \text{if } u = 2k \text{ for a positive integer } k \\ \binom{n_1-1}{k}\binom{n_2-1}{k-1} + \binom{n_1-1}{k-1}\binom{n_2-1}{k} & \text{if } u = 2k+1 \text{ for a positive integer } k. \end{cases} \tag{3.2}$$

Proof. We consider $u = 2k$ and leave the other case as an exercise. As there are an even number of runs, we must either begin with a head and end with a tail, or we must begin with a tail and end with a head. By symmetry, it is enough to consider just the case when we start with a head and then multiply by 2. The reason is if we have a sequence like $HHHTTHTTTHTHT$ we can reverse it and obtain a sequence that starts with a tail and ends with a head.

Let us assume we will start with a head and end with a tail. Consider a string of n_1 heads. If we partition it into k strings of heads, we can then put tails in after the partitions, and we will have $2k$ runs; however, we *must* put a partition after the final head, as we must end with a tail. Further, we cannot put a partition before the first head as we *must* start with a head. For example, if we partition $HHHHH$ by adding partitions $|$ to get $H|HHH|H|$, then we can add strings of tails after the partitions to get $HT \cdots THHHT \cdots THT \cdots T$ for a total of 6 runs. How many ways are there to partition n_1 heads into k groups *with a partition occurring after the final head and no partition allowed before the first head*? Note there are $n_1 + 1$ positions where we can put a partition (before the first head, after the first head, after the second head, ..., after the last head); however, we shall see that two of these positions have their values forced.

We must choose the last place for one partition, we cannot choose the place before the first head, and then we must choose $k - 1$ of the remaining $n_1 - 1$ positions for the other partitions. Thus the number of ways to add k partitions when we must add a partition after the final head and we cannot add one before the first head is just $\binom{1}{1}\binom{1}{0}\binom{n_1-1}{k-1} = \binom{n_1-1}{k-1}$. A similar argument shows there are $\binom{n_2-1}{k-1}$ ways to partition n_2 tails into k groups, assuming we must have a partition before the first tail and we are not allowed to have a partition after the final tail.

We now intersperse the partitioned heads and tails. Consider any of the $\binom{n_1-1}{k-1}$ partitions of the n_1 heads and any of the $\binom{n_2-1}{k-1}$ partitions of the n_2 tails. Each such pair gives rise to a sequence of n_1 heads and n_2 tails with exactly $2k$ runs, and any such sequence corresponds to a unique pair. For example, say we have $H|HH|HHH|$ and $|TTTT|T|TT$; these unite to become $HTTTTHHTHH$.

Thus the number of partitions leading to $2k$ runs where the first coin is a head and the last is a tail is just $\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}$. By symmetry this is the same as the number of partitions where the first coin is a tail and the last is a head, which completes the proof of the theorem in the case of an even number of runs. □

Of course, in the arguments above $1 \le k \le \min(n_1, n_2)$; for other k the number of strings with $2k$ runs is zero.

4 Determining How Often There are an Even Number of Runs

By differentiating identities we determine how often there are an *even* number of runs when there are n_1 heads and n_2 tails and each of the $\binom{n_1+n_2}{n_1}$ strings are equally likely. A similar argument is applicable for the case when there are an odd number of runs; we concentrate here on the case of an even number to highlight the methods.

If $u = 2k$ is the number of runs, then we know the number of ways to have $2k$ runs is just

$$2\binom{n_1 - 1}{k - 1}\binom{n_2 - 1}{k - 1}. \tag{4.1}$$

Without loss of generality, for notational convenience let us assume $n_1 \ge n_2$, so k runs from 1 to n_2. Thus the number of strings with an even number of runs is just

$$\sum_{k=1}^{n_2-1} 2\binom{n_1 - 1}{k - 1}\binom{n_2 - 1}{k - 1}, \tag{4.2}$$

as there must be at least two runs (there is no way to have zero runs unless $n_1 = n_2 = 0$, which we shall assume we do not have). We first need to determine what this sum is, and then to determine the expected number of u (when $u = 2k$ is even) we will need to sum

$$\sum_{k=1}^{n_2-1} (2k) \cdot 2\binom{n_1 - 1}{k - 1}\binom{n_2 - 1}{k - 1}. \tag{4.3}$$

4.1 Determining the number of strings with $u = 2k$ runs

Consider the polynomial

$$(x_1 + y_1)^{n_1-1}(x_2 + y_2)^{n_2-1}; \tag{4.4}$$

we shall see very shortly why this is a "natural" polynomial to examine. Using the Binomial Theorem (Theorem B.4) we have

$$\begin{aligned} (x_1 + y_1)^{n_1-1} &= \sum_{k_1=0}^{n_1-1} \binom{n_1 - 1}{k_1} x_1^{n_1-1-k_1} y_1^{k_1} = \sum_{k_1=1}^{n_1} \binom{n_1 - 1}{k_1 - 1} x_1^{n_1-k_1} y_1^{k_1-1} \\ (x_2 + y_2)^{n_2-1} &= \sum_{k_2=0}^{n_2-1} \binom{n_2 - 1}{k_2} x_2^{k_2} y_2^{n_2-1-k_2} = \sum_{k_2=1}^{n_2} \binom{n_2 - 1}{k_2 - 1} x_2^{k_2-1} y_2^{n_2-k_2}; \end{aligned} \tag{4.5}$$

we will see later why it is convenient to have $x_1^{n_1-k_1}$ but $x_2^{k_2-1}$; we can write the binomial theorem this way as $\binom{m}{r} = \binom{m}{m-r}$. Therefore

$$(x_1 + y_1)^{n_1-1}(x_2 + y_2)^{n_2-1} = \left[\sum_{k_1=1}^{n_1} \binom{n_1 - 1}{k_1 - 1} x_1^{n_1-k_1} y_1^{k_1-1}\right] \cdot \left[\sum_{k_2=1}^{n_2} \binom{n_2 - 1}{k_2 - 1} x_2^{k_2-1} y_2^{n_2-k_2}\right]. \tag{4.6}$$

Consider what happens if we set $x_1 = x_2 = x$ and $y_1 = y_2 = y$. Then the above becomes

$$\begin{aligned} (x + y)^{n_1+n_2-2} &= \left[\sum_{k_1=1}^{n_1} \binom{n_1 - 1}{k_1 - 1} x^{n_1-k_1} y^{k_1-1}\right] \cdot \left[\sum_{k_2=1}^{n_2} \binom{n_2 - 1}{k_2 - 1} x^{k_2-1} y^{n_2-k_2}\right] \\ &= \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \binom{n_1 - 1}{k_1 - 1} \binom{n_2 - 1}{k_2 - 1} x^{n_1-1-k_1+k_2} y^{n_2-1+k_1-k_2}. \end{aligned} \tag{4.7}$$

Now we use the uniqueness of polynomial expansions and equate coefficients. Consider the $x^{n_1-1}y^{n_2-1}$ term in (4.7). There are two ways we can calculate it. Looking at the left hand side, we have $(x + y)^{n_1+n_2-2}$, and thus the term is just $\binom{n_1+n_2-2}{n_1-1} x^{n_1-1}y^{n_2-1}$. Looking at the right hand side we see the term we desire occurs when $k_1 = k_2$. We see now why we wrote $x_1^{n_1-k_1}$ and $x_2^{k_2-1}$; this made it easy to combine the terms. Denoting the common value of k_1 and k_2 by k we obtain

$$\binom{n_1 + n_2 - 2}{n_1 - 1} x^{n_1-1} y^{n_2-1} = \sum_{k=1}^{n_2} \binom{n_1 - 1}{k - 1} \binom{n_2 - 1}{k - 1} x^{n_1-1} y^{n_2-1}, \tag{4.8}$$

or cancelling the x's and the y's

$$\binom{n_1 + n_2 - 2}{n_1 - 1} = \sum_{k=1}^{n_2} \binom{n_1 - 1}{k - 1} \binom{n_2 - 1}{k - 1}. \tag{4.9}$$

We have determined the sum in (4.2), the sum we needed to figure out how many different strings there are with n_1 heads, n_2 tails and $u = 2k$ runs! Namely, we have shown

Lemma 4.1. *The number of strings with n_1 heads, n_2 tails and $u = 2k$ runs is*

$$\sum_{k=1}^{n_2-1} 2\binom{n_1-1}{k-1}\binom{n_2-1}{k-1} \;=\; 2\binom{n_1+n_2-2}{n_1-1}. \tag{4.10}$$

Some discussion is clearly in order as to how we knew we should consider $(x_1+y_1)^{n_1-1}(x_2+y_2)^{n_2-1}$. This is the hardest step in all such proofs by matching or proofs by differentiating identities, namely figuring out *where* to start. The answer is usually suggested by trying to analyze the quantity being studied, looking for clues as to what series or products we should consider.

In this case, we knew that we had to eventually have products like $\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}$. How can we get such terms? Well, the $\binom{n_1-1}{k-1}$ are the coefficients when we expand $(A+B)^{n_1-1}$; we chose $A = x_1$ and $B = y_1$ to have some flexibility, and to distinguish these terms from the other factors. For simply counting the number of strings with $u = 2k$ runs this extra degree of freedom or flexibility was not needed; however, it will be crucial in trying to find the mean of u when u is even. Similarly the $\binom{n_2-1}{k-1}$ are the coefficients from expanding $(A+B)^{n_2-1}$, and we choose $A = x_2$ and $B = y_2$ for the same reasons as before. By setting $x_1 = x_2 = x$ and $y_1 = y_2 = y$ in the end we are arguing in a similar manner as in §2. This is a common and powerful technique, namely writing $(x+y)^{n+m}$ and $(x+y)^n(x+y)^m$ and then deducing identities for sums involving terms like $\binom{n}{r}\binom{m}{a+r}$ for a fixed a.

4.2 Determining the expected value of u for strings with $u = 2k$ runs

We now turn to the sum in (4.2), which gives the expected value of $u = 2k$; again, remember that we are only considering strings with n_1 heads, n_2 tails and an even number $u = 2k$ of runs. As by Lemma 4.1 there are $2\binom{n_1+n_2-2}{n_1-1}$ such strings and the number of strings with $2k$ runs is $2\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}$, we need to determine

$$\frac{\sum_{k=1}^{n_2-1}(2k)\cdot 2\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}}{2\binom{n_1+n_2-2}{n_1-1}} \;=\; 2\frac{\sum_{k=1}^{n_2-1} k\cdot\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}}{\binom{n_1+n_2-2}{n_1-1}}. \tag{4.11}$$

We shall ignore the factor of $2\binom{n_1+n_2-2}{n_1-1}^{-1}$ for now and concentrate on evaluating

$$\sum_{k=1}^{n_2-1} k\cdot\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}. \tag{4.12}$$

Actually, it will be significantly easier to find, not the sum with k but the sum with $k-1$:

$$\sum_{k=1}^{n_2-1}(k-1)\cdot\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}; \tag{4.13}$$

clearly if we can evaluate this sum for $k-1$ then by adding 1 we can find the sum with k.

We have seen in §4.1 that the sum over k of $\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}$ can be obtained by looking at the $x^{n_1-1}y^{n_2-1}$ coefficient of $(x_1+y_1)^{n_1-1}(x_2+y_2)^{n_2-1}$ under $x_1 = x_2 = x$ and $y_1 = y_2 = y$. So, let us study again (4.6):

$$(x_1+y_1)^{n_1-1}(x_2+y_2)^{n_2-1} \quad = \quad \left[\sum_{k_1=1}^{n_1}\binom{n_1-1}{k_1-1}x_1^{n_1-k_1}y_1^{k_1-1}\right]\cdot\left[\sum_{k_2=1}^{n_2}\binom{n_2-1}{k_2-1}x_2^{k_2-1}y_2^{n_2-k_2}\right]. \tag{4.14}$$

We will now see the advantage of having two different x's and two different y's. Let us take the derivative with respect to y_1 and then multiply by y_1. Thus we are applying the operator $y_1\frac{\partial}{\partial y_1}$; the advantage of multiplying by y_1 after differentiating by y_1 is that we do not change the degree of any of the terms. Applying $y_1\frac{\partial}{\partial y_1}$ to the left hand side of (4.14) gives

$$(n_1-1)y_1(x_1+y_1)^{n_1-2}(x_2+y_2)^{n_2-1}, \tag{4.15}$$

because x_1, y_1, x_2 and y_2 are independent variables. When we apply $y_1\frac{\partial}{\partial y_1}$ to the right hand side of (4.14) we get

$$\left[\sum_{k_1=1}^{n_1}(k_1-1)\binom{n_1-1}{k_1-1}x_1^{n_1-k_1}y_1^{k_1-1}\right]\cdot\left[\sum_{k_2=1}^{n_2}\binom{n_2-1}{k_2-1}x_2^{k_2-1}y_2^{n_2-k_2}\right]. \tag{4.16}$$

The above shows why it is easier to study $k-1$ rather than k: when we differentiate a factor of $k-1$ comes down, not k. We have thus shown

$$\begin{aligned}&(n_1-1)y_1(x_1+y_1)^{n_1-2}(x_2+y_2)^{n_2-1}\\ &\quad=\left[\sum_{k_1=1}^{n_1}(k_1-1)\binom{n_1-1}{k_1-1}x_1^{n_1-k_1}y_1^{k_1-1}\right]\cdot\left[\sum_{k_2=1}^{n_2}\binom{n_2-1}{k_2-1}x_2^{k_2-1}y_2^{n_2-k_2}\right].\end{aligned} \tag{4.17}$$

NOW we take $x_1 = x_2 = x$ and $y_1 = y_2 = y$ and obtain

Lemma 4.2.

$$\begin{aligned}&(n_1-1)y(x+y)^{n_1+n_2-3}\\ &\quad=\left[\sum_{k_1=1}^{n_1}(k_1-1)\binom{n_1-1}{k_1-1}x^{n_1-k_1}y^{k_1-1}\right]\cdot\left[\sum_{k_2=1}^{n_2}\binom{n_2-1}{k_2-1}x^{k_2-1}y^{n_2-k_2}\right].\end{aligned} \tag{4.18}$$

It is extremely important that we waited to set x_1 equal to x_2 and y_1 equal to y_2; if we had set them equal first and then differentiated, we would have two pieces (from when the operator hit the first sum and when it hit the second). The difficulty would be the first sum would bring down a factor of k_1-1 and the second a factor of n_2-k_2. With some book-keeping this could probably be made to work, but this is easier.

We now look at the $x^{n_1-1}y^{n_2-1}$ term of both sides of Lemma 4.2. First consider the left hand side. We have one factor of y automatically because of the y outside. There are $\binom{n_1+n_2-3}{n_1-1}$ ways to choose $n_1 - 1$ factors of $(x+y)^{n_1+n_2-3}$ to give x and $n_2 - 2$ factors to give y. Thus the coefficient of $x^{n_1-1}y^{n_2-1}$ on the left hand side is

$$(n_1 - 1)\binom{n_1 + n_2 - 3}{n_1 - 1}. \tag{4.19}$$

We now determine the $x^{n_1-1}y^{n_2-1}$ term from the right hand side of Lemma 4.2. As before, this term arises from $k_1 = k_2$. Denoting this common value by k we find the coefficient of the $x^{n_1-1}y^{n_2-1}$ term from the right hand side is

$$\sum_{k=1}^{n_2}(k-1)\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}. \tag{4.20}$$

As always, the proof is concluded by the uniqueness of the coefficients. By matching we obtain

Lemma 4.3.

$$\sum_{k=1}^{n_2}(k-1)\binom{n_1-1}{k-1}\binom{n_2-1}{k-1} \;=\; (n_1-1)\binom{n_1+n_2-3}{n_1-1}. \tag{4.21}$$

We can now determine the mean of $k-1$, or better yet $2(k-1)$. From this it is trivial to determine the mean of $2k$. Specifically

Lemma 4.4.

$$\frac{\sum_{k=1}^{n_2-1} 2(k-1)\cdot 2\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}}{2\binom{n_1+n_2-2}{n_1-1}} \;=\; 2\frac{n_1n_2 - n_1 - n_2 + 1}{n_1 + n_2 - 3}. \tag{4.22}$$

Proof. The denominator comes from Lemma 4.1, where we showed this is the number of strings with n_1 heads, n_2 tails and an even number of runs. We cancel two of the factors of 2 and are left with one factor of 2 in the numerator, and then use Lemma 4.3 to evaluate the numerator. The proof is completed by expanding out the binomial coefficients. Let $\mu_{u-2,\text{even}}$ denote the mean of two less than even u (in other words, the expected value of $2(k-1)$ when $u = 2k$). Then

$$\begin{aligned}
\mu_{u-2,\text{even}} &= \frac{2(n_1-1)\binom{n_1+n_2-3}{n_1-1}}{\binom{n_1+n_2-2}{n_1-1}} \\
&= 2(n_1-1)\binom{n_1+n_2-3}{n_1-1}\cdot\binom{n_1+n_2-2}{n_1-1}^{-1} \\
&= \frac{2(n_1-1)(n_1+n_2-3)!}{(n_1-1)!(n_2-2)!}\cdot\frac{(n_1-1)!(n_2-1)!}{(n_1+n_2-2)!} \\
&= \frac{2(n_1+n_2-3)!}{(n_1-2)!(n_2-2)!}\cdot\frac{(n_1-1)!(n_2-1)!}{(n_1+n_2-2)(n_1+n_2-3)!} \\
&= \frac{2(n_1-1)(n_2-1)}{n_1+n_2-2} \\
&= 2\frac{n_1n_2 - n_1 - n_2 + 1}{n_1+n_2-2}.
\end{aligned} \tag{4.23}$$

Note that as we write $(n_1 + n_2 - 2)!$ as $(n_1 + n_2 - 2) \cdot (n_1 + n_2 - 3)!$, we are implicitly assuming that $n_1 + n_2 - 2 \geq 1$. If this fails, i.e. if $n_1 + n_2 \leq 2$, then the above algebra could be wrong and those cases should be investigated separately (though if interpreted properly, our formulas will still be correct in these cases). □

By adding 2 we get the mean of $u = 2k$ for even u.

Theorem 4.5. *Assume we have n_1 heads, n_2 tails, $u = 2k$ runs and all strings are equally likely. Then the expected number of runs is*

$$\mu_{u,\text{even}} \quad = \quad 2\left[\frac{n_1 n_2 - n_1 - n_2 + 1}{n_1 + n_2 - 2} + 1\right] \;=\; 2\frac{n_1 n_2 - 1}{n_1 + n_2 - 2}. \tag{4.24}$$

Whenever one derives a complicated formula, it is a good idea to test it in extreme cases and see if it is reasonable. For example, the formula does not make sense if $n_1 + n_2 - 2 = 0$. However, the only way that could happen, since n_1 and n_2 are non-negative integers, is if either both equal 1 or one is 0 and the other 2. If one is 0 and the other is 2 then we have an *odd* number of runs, and this formula is only for the case of an even number of runs. We are left with the case when $n_1 = n_2 = 1$. We have two runs, either HT or TH. In this case we have $2\frac{n_1 n_2 - 1}{n_1 + n_2 - 2} = 2\frac{0}{0}$; it is not unreasonable to think $\frac{0}{0}$ should be interpreted as 1 in this instance, and we would then get 2 (the correct answer). However, some care is needed in using this formula when $n_1 + n_2 = 2$, but this case can be handled directly.

Another good extreme to consider is when n_1 is much larger than n_2 (or vice-versa, but we have assumed without loss of generality earlier that $n_1 \geq n_2$). In this case, the mean for sequences with an even number of runs is approximately $2\frac{n_1 n_2}{n_1}$ or about $2n_2$. This is the correct behavior for such n_1 and n_2. Why? Imagine we have millions of time more heads (n_1) than tails (n_2). In that case it is extremely unlikely that any two tails will be adjacent. Thus there will be strings of varying lengths between the tails. As there are n_2 tails, this gives us $2n_2$ runs (the heads before a tail, a tail, another string of heads, a tail, another string of heads, a tail, and so on).

While such sanity checks are not proofs, they help us see if our formulas are reasonable, as well as possibly catching missing factors. For example, if we had dropped a factor of 2 earlier we would have found the mean was $\frac{n_1 n_2 - 1}{n_1 + n_2 - 2}$, and this would not have the right behavior for n_1 significantly larger than n_2. We also saw that the -2 in the denominator is reasonable.

We can also try a special case, for example $n_1 = 2$, $n_2 = 1$. In this case if we want an even number of runs we must have HHT or THH. Thus all strings with an even number of runs have 2 runs, and our formula does give 2 when $n_1 = 2$ and $n_2 = 1$. This helps check the -1 factor.

Thus, while it is still possible that we have made an algebra error somewhere, we should have a high degree of confidence in the result.

4.3 Determining the variance of u for strings with $u = 2k$ runs

Theorem 4.6. *Assume we have n_1 heads, n_2 tails, $u = 2k$ runs and all strings are equally likely. Then the variance in the number of runs is*

$$\sigma^2_{u,\text{even}} \quad = \quad 4\frac{(n_1 - 1)^2(n_2 - 1)^2}{(n_1 + n_2 - 2)^2(n_1 + n_2 - 3)}. \tag{4.25}$$

Proof. As $u = 2k$ is even, we need to find $\text{Var}(2k) = \mathbb{E}[(2k)^2] - \mathbb{E}[2k]^2$. We can simplify the calculations by noting that the variance of $u = 2k$ is the same as the variance of $u - 2 = 2(k-1)$. While we know the mean of both $u = 2k$ and $u - 2 = 2(k-1)$, it will turn out to be easier to calculate $\mathbb{E}[(2k-2)^2]$ than $\mathbb{E}[(2k)^2]$.

Thus we must evaluate

$$\frac{\sum_{k=1}^{n_2-1}[2(k-1)]^2 \cdot 2\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}}{2\binom{n_1+n_2-2}{n_1-1}} = 4\frac{\sum_{k=1}^{n_2-1}(k-1)^2 \cdot \binom{n_1-1}{k-1}\binom{n_2-1}{k-1}}{\binom{n_1+n_2-2}{n_1-1}}. \tag{4.26}$$

As before, the starting point is (4.6):

$$(x_1+y_1)^{n_1-1}(x_2+y_2)^{n_2-1} = \left[\sum_{k_1=1}^{n_1}\binom{n_1-1}{k_1-1}x_1^{n_1-k_1}y_1^{k_1-1}\right]\cdot\left[\sum_{k_2=1}^{n_2}\binom{n_2-1}{k_2-1}x_2^{k_2-1}y_2^{n_2-k_2}\right]. \tag{4.27}$$

We apply the operator $x_2 y_1 \frac{\partial^2}{\partial x_2 \partial y_1}$. The reason for this choice is that the two derivatives bring down a factor of $(k_1-1)(k_2-1)$; the presence of x_2y_1 means the degree of each term is unchanged (in all four variables x_1, x_2, y_1, y_2). Setting $x_1 = x_2 = x$ and $y_1 = y_2 = y$ and matching coefficients will complete the proof, as looking at the coefficient of $x^{n_1-1}y^{n_1-1}$ will cause $k_1 = k_2$, and this will give us the sum we desire.

Specifically, after applying $x_2 y_1 \frac{\partial^2}{\partial x_2 \partial y_1}$ the left hand side of (4.27) is

$$(n_1-1)(n_2-1)x_2y_1(x_1+y_1)^{n_1-2}(x_2+y_2)^{n_2-2}, \tag{4.28}$$

while the right hand side of (4.27) is

$$\left[\sum_{k_1=1}^{n_1}(k_1-1)\binom{n_1-1}{k_1-1}x_1^{n_1-k_1}y_1^{k_1-1}\right]\cdot\left[\sum_{k_2=1}^{n_2}(k_2-1)\binom{n_2-1}{k_2-1}x_2^{k_2-1}y_2^{n_2-k_2}\right]. \tag{4.29}$$

Setting $x_1 = x_2 = x$ and $y_1 = y_2 = y$, (4.28) and (4.29) give

$$(n_1-1)(n_2-1)xy(x+y)^{n_2+n_2-4} = \left[\sum_{k_1=1}^{n_1}(k_1-1)\binom{n_1-1}{k_1-1}x^{n_1-k_1}y^{k_1-1}\right] \cdot \left[\sum_{k_2=1}^{n_2}(k_2-1)\binom{n_2-1}{k_2-1}x^{k_2-1}y^{n_2-k_2}\right]. \tag{4.30}$$

We match the $x^{n_1-1}y^{n_1-1}$ term on both sides. The left hand side is easy. As we have an xy outside, we see we need to choose $n_1 - 2$ more x's and $n_2 - 2$ more y's. The right hand side is just the sum over $k_1 = k_2$. Denoting this common value by k we find

$$(n_1-1)(n_2-1)\binom{n_1+n_2-4}{n_1-2}x^{n_1-1}y^{n_2-1} = \sum_{k=1}^{n_2}(k-1)^2\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}x^{n_1-1}y^{n_2-1}, \tag{4.31}$$

or equivalently

$$\sum_{k=1}^{n_2}(k-1)^2\binom{n_1-1}{k-1}\binom{n_2-1}{k-1} \;=\; (n_1-1)(n_2-1)\binom{n_1+n_2-4}{n_1-2}. \tag{4.32}$$

Therefore we have

$$\mathbb{E}[(2k-2)^2] \;=\; \frac{4(n_1-1)(n_2-1)\binom{n_1+n_2-4}{n_1-2}}{\binom{n_1+n_2-2}{n_1-1}}. \tag{4.33}$$

We can simplify the above expression to make it easier to subtract $\mathbb{E}[(2k-2)]^2$:

$$\begin{aligned}
\mathbb{E}[(2k-2)^2] &= 4(n_1-1)(n_2-1)\frac{(n_1+n_2-4)!}{(n_1-2)!(n_2-2)!}\cdot\frac{(n_1-1)!(n_2-1)!}{(n_1+n_2-2)!} \\
&= 4(n_1-1)(n_2-1)\frac{(n_1+n_2-4)!}{(n_1-2)!(n_2-2)!}\cdot\frac{(n_1-1)(n_1-2)!(n_2-1)(n_2-2)!}{(n_1+n_2-2)(n_1+n_2-3)(n_1+n_2-4)!} \\
&= 4\frac{(n_1-1)^2(n_2-1)^2}{(n_1+n_2-2)(n_1+n_2-3)} \\
&= 4\frac{(n_1-1)^2(n_2-1)^2}{(n_1+n_2-2)^2}\cdot\frac{n_1+n_2-2}{n_1+n_2-3}.
\end{aligned} \tag{4.34}$$

We must now subtract $\mathbb{E}[(2k-2)]^2$. It is easiest algebraically to use the expression for $\mathbb{E}[(2k-2)]^2$ from the second to last line of (4.23). This yields

$$\begin{aligned}
\text{Var}(2k-2) &= 4\frac{(n_1-1)^2(n_2-1)^2}{(n_1+n_2-2)^2}\cdot\frac{n_1+n_2-2}{n_1+n_2-3}-\left[\frac{2(n_1-1)(n_2-1)}{n_1+n_2-2}\right]^2 \\
&= 4\frac{(n_1-1)^2(n_2-1)^2}{(n_1+n_2-2)^2}\left[\frac{n_1+n_2-2}{n_1+n_2-3}-1\right] \\
&= 4\frac{(n_1-1)^2(n_2-1)^2}{(n_1+n_2-2)^2}\cdot\frac{1}{n_1+n_2-3} \\
&= 4\frac{(n_1-1)^2(n_2-1)^2}{(n_1+n_2-2)^2(n_1+n_2-3)},
\end{aligned} \tag{4.35}$$

and $\text{Var}(2k-2)=\text{Var}(2k)$. □

For large n_1 and n_2,

$$\text{Var}(2k) \;\sim\; 4\frac{n_1^2n_2^2}{(n_1+n_2)^3}. \tag{4.36}$$

If n_1 is much larger than n_2, the mean is approximately $2n_2$ and the variance is approximately $4\frac{n_2^2}{n_1}$.

4.4 Behavior for all u

We briefly describe what happens if we don't restrict to the case when u, the number of runs, is even. The main result is that

Theorem 4.7. *Assume we have n_1 heads, n_2 tails, u runs and all strings are equally likely; u may be either even or odd and we assume $n_1, n_2 \geq 1$. Then the expected number of runs u is $\frac{2n_1n_2}{n_1+n_2}+1$ and the variance is $\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$. For n_1 and n_2 large, the expected number of runs is approximately $2\frac{n_1n_2}{n_1+n_2}$ and the variance is approximately $4\frac{n_1^2n_2^2}{(n_1+n_2)^3}$.*

Note our results on the expected number and variance of u (when u is forced to be even) are consistent with the above, at least when n_1 and n_2 are large. This isn't surprising, as when n_1 and n_2 are large it is reasonable to think that there are about as many strings with an odd number of runs as an even number of runs.

Sketch of the proof. To prove Theorem 4.7 we would need to investigate the case when $u = 2k+1$. The starting point is the second part of (3.2), which tells us how many ways there are to have $u = 2k+1$ runs. We need to know how many strings there are with n_1 heads and n_2 tails so that we can find the probabilities of having $u = 2k$ or $u = 2k+1$ runs. This is just $\binom{n_1+n_2}{n_1}$ as we choose n_1 of the n_1+n_2 positions to be heads.

In determining the mean and variance when $u = 2k-2$ we divided the number of strings with $2k$ runs by $2\binom{n_1+n_2-2}{n_1-1}$, which is the number of strings with n_1 heads, n_2 tails and an even number of runs. What we can do is multiply our results on the mean and variance in this case by

$$\frac{2\binom{n_1+n_2-2}{n_1-1}}{\binom{n_1+n_2}{n_1}}, \tag{4.37}$$

which now divides the contribution by the total number of strings and not just the total number of strings with an even number of runs.

The proof is completed by determining the contributions to the mean and the variance from the $u = 2k+1$ terms. These contributions are found in a similar manner (i.e. by differentiating identities) as the $u = 2k$ terms. We leave the details to the reader. □

For completeness, we sketch the key steps in the algebra to finish the proof. We need to find the mean. For the terms with an even number of runs we need to average $2k$ and for the terms with an odd number of runs we average $2k+1$.

For the even terms, we showed that there are $2\binom{n_1+n_2-2}{n-1}$ strings, and there are $\binom{n_1+n_2}{n_1}$ total strings. We multiply the mean in Theorem 4.5 by $\frac{2\binom{n_1+n_2-2}{n-1}}{\binom{n_1+n_2}{n_2}}$.

For the odd terms, from (3.2) we have two sums to study. To analyze the contribution from

$$\sum_k \binom{n_1-1}{k}\binom{n_2-1}{k-1} \tag{4.38}$$

we see this can be interpreted by looking at the $x^{n_2-2}y^{n_2}$ term of

$$\sum_{k_1} \binom{n_1-1}{k_1} x_1^{n_1-1-k_1} y_1^{k_1} \sum_{k_2} \binom{n_2-1}{k-1} x_2^{k_2-1} y_2^{n_2-k_2} \tag{4.39}$$

when we set $x_1 = x_2 = x$ and $y_1 = y_2 = y$. We see this term is the $x^{n_2-2}y^{n_2}$ term of

$$(x_1+y_1)^{n_1-1}(x_2+y_2)^{n_2-1} \tag{4.40}$$

when we set $x_1 = x_2 = x$ and $y_1 = y_2 = y$, and that term is just $\binom{n_1+n_2-2}{n_1-2} x^{n_1-2} y^{n_2}$. Note this allows us to determine the sum of these binomial coefficients. We need to evaluate the sum with a factor of $2k+1$. To evaluate the sum with a factor of k we apply the operator $y_1 \frac{\partial}{\partial y_1}$; to handle the $+1$ in $2k+1$ we just need to count the number of terms, which from above is $\binom{n_1+n_2-2}{n_1-2}$. Therefore, the contribution from these terms with odd u from (3.2) to the mean is just

$$2(n_1-1)\binom{n_1+n_2-3}{n_1-2} + \binom{n_1+n_2-2}{n_1-2} \tag{4.41}$$

while the other terms with odd u in (3.2) give (by a similar argument or by symmetry) a contribution of

$$\binom{n_1+n_2-3}{n_2-2} + \binom{n_1+n_2-2}{n_2-2}. \tag{4.42}$$

We then must go through a lot of algebra - after adding all of these contributions we divide by the number of strings, $\binom{n_1+n_2}{n_1}$. In adding the various terms it is often convenient to pull out factors of $\frac{(n_1+n_2-3)!}{(n_1-2)!(n_2-2)!}$. In the end we show the mean is $\frac{2n_1n_2}{n_1+n_2} + 1$. It is convenient to notice that

$$(n_1+n_2)(n_1+n_2-1)(n_1+n_2-2) \ = \ n_1^3+n_2^3+3n_1^2n_2+3n_1n_2^2-3n_1^2-3n_2^2-6n_1n_2+2n_1+2n_2. \tag{4.43}$$

Exercise 4.8. *Calculate the contributions from the $u = 2k+1$ terms and rescale the contributions from the $u = 2k$ terms to complete the proof of Theorem 4.7.*

4.5 Expected Number of Runs with Arbitrary Numbers of Heads and Tails

So far we assumed that there were n_1 heads, n_2 tails and all strings were equally likely. Let us assume now that we have N coin tosses where each toss has probability p of being a head and $q = 1-p$ of being a tail. Thus, $n_2 = N - n_1$. For each n_1 there are $\binom{N}{n_1}$ strings; all of these strings are equally likely, each occurring with probability $p^{n_1} q^{N-n_1}$. Our main result is

Theorem 4.9. *Assume we toss a coin with probability p of heads a total of N times. The expected number of runs $\mu_u(p)$ is $2p(1-p)(N-1)+1$. In particular, if the coin is fair (so $p = q = \frac{1}{2}$) then the expected number of runs is $\frac{N+1}{2}$.*

Proof. If there are n_1 heads then the expected number of runs is $\frac{2n_1(N-n_1)}{N} + 1$, and there are $\binom{N}{n_1}$

such strings, each occurring with probability $p^{n_1}q^{N-n_1}$. Thus the expected number of runs $\mu_u(p)$ is

$$\begin{aligned}
\mu_u &= \sum_{n_1=0}^{N}\left[\frac{2n_1(N-n_1)}{N}+1\right]\cdot\binom{N}{n_1}p^{n_1}q^{N-n_1} \\
&= 2\sum_{n_1=0}^{N}\frac{n_1(N-n_1)}{N}\frac{N!}{n_1!(N-n_1)!}p^{n_1}q^{N-n_1}+\sum_{n_1=0}^{N}\binom{N}{n_1}p^{n_1}q^{N-n_1} \\
&= 2pq\sum_{n_1=1}^{N-1}\frac{(N-1)!}{(n_! - 1)!(N-n_1-1)!}p^{n_1-1}q^{N-n_1-1}+(p+q)^N \\
&= 2pq(N-1)\sum_{n_1=1}^{N-1}\frac{(N-2)!}{(n_1-1)!(N-n_1-1)!}p^{n_1-1}q^{N-n_1-1}+(p+q)^N \\
&= 2p(1-p)(N-1)(p+q)^{N-2}+(p+q)^N. \qquad (4.44)
\end{aligned}$$

As $q = 1 - p$ the above becomes

$$\mu_u(p) = 2pq(N-1)+1. \qquad (4.45)$$

In the special case that $p = q = \frac{1}{2}$ we have

$$\mu_u\left(\frac{1}{2}\right) = \frac{N+1}{2}. \qquad (4.46)$$

□

Exercise 4.10. *Calculate the variance of* $\mu_u(p)$.

A Proofs by Induction

Assume for each positive integer n we have a statement $P(n)$ which we desire to show is true. $P(n)$ is true for all positive integers n if the following two statements hold:

- **Basis Step:** $P(1)$ is true;
- **Inductive Step**: whenever $P(n)$ is true, $P(n+1)$ is true.

This technique is called **Proof by Induction**, and is a very useful method for proving results. The reason the method works follows from basic logic. We assume the following two sentences are true:

$$\begin{aligned}
&P(1) \text{ is true} \\
\forall n \geq 1, &P(n) \text{ is true implies } P(n+1) \text{ is true.} \qquad (A.1)
\end{aligned}$$

Set $n = 1$ in the second statement. As $P(1)$ is true, and $P(1)$ implies $P(2)$, $P(2)$ must be true. Now set $n = 2$ in the second statement. As $P(2)$ is true, and $P(2)$ implies $P(3)$, $P(3)$ must be true. And so on, completing the proof. Verifying the first statement the **basis step** and the second the **inductive step**. In verifying the inductive step, note we assume $P(n)$ is true; this is called the **inductive assumption**. Sometimes instead of starting at $n = 1$ we start at $n = 0$, although in general we could start at any n_0 and then prove for all $n \geq n_0$, $P(n)$ is true.

Theorem A.1. *For n a non-negative integer,*

$$\sum_{k=1}^{n} k \;=\; \frac{n(n+1)}{2}. \tag{A.2}$$

Proof. Let $P(n)$ be the statement

$$\sum_{k=1}^{n} k \;=\; \frac{n(n+1)}{2}. \tag{A.3}$$

Basis Step: $P(1)$ is true, as both sides equal 1.

Inductive Step: Assuming $P(n)$ is true, we must show $P(n+1)$ is true. By the inductive assumption, $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$. Thus

$$\begin{aligned} \sum_{k=1}^{n+1} k &= (n+1) + \sum_{k=1}^{n} k \\ &= (n+1) + \frac{n(n+1)}{2} \\ &= \frac{(n+1)(n+1+1)}{2}. \end{aligned} \tag{A.4}$$

Thus, given $P(n)$ is true, then $P(n+1)$ is true. □

Exercise A.2. *Prove*

$$\sum_{k=1}^{n} k^2 \;=\; \frac{n(n+1)(2n+1)}{6}. \tag{A.5}$$

Find a similar formula for the sum of k^3. See also Remark A.3.

Remark A.3. *In general, $\sum_{k=0}^{n} k^p = f_p(n)$, where $f_p(x)$ is a polynomial of degree $p+1$ with leading term $\frac{x^{p+1}}{p+1}$; one can find the coefficients by evaluating the sums for $n = 0, 1, \ldots, p$ because specifying the values of a polynomial of degree p at $p+1$ points uniquely determines the polynomial.*

Exercise A.4. *Notation as in Remark A.3, assuming $f_p(n)$ is a polynomial in n, use the integral test from calculus to show the leading term is $\frac{n^{p+1}}{p+1}$.*

Exercise A.5. *Show the sum of the first n odd numbers is n^2, i.e.,*

$$\sum_{k=1}^{n} (2k-1) \;=\; n^2. \tag{A.6}$$

B The Binomial Theorem

We prove the Binomial Theorem. First, recall that

Definition B.1 (Binomial Coefficients)**.** *Let n and k be integers with $0 \le k \le n$. We set*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{B.1}$$

Note that $0! = 1$ and $\binom{n}{k}$ is the number of ways to choose k objects from n (with order not counting).

Lemma B.2. *We have*

$$\binom{n}{k} = \binom{n}{n-k}, \qquad \binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}. \tag{B.2}$$

Exercise B.3. *Prove Lemma B.2.*

Theorem B.4 (The Binomial Theorem)**.** *For all positive integers n we have*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k. \tag{B.3}$$

Proof. We proceed by induction.

Basis Step: For $n = 1$ we have

$$\sum_{k=0}^{1} \binom{1}{k} x^{1-k} y^k = \binom{1}{0} x + \binom{1}{1} y = (x+y)^1. \tag{B.4}$$

Inductive Step: Suppose

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k. \tag{B.5}$$

Then using Lemma B.2 we find that

$$\begin{aligned}
(x+y)^{n+1} &= (x+y)(x+y)^n \\
&= (x+y) \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k \\
&= \sum_{k=0}^{n} \binom{n}{k} x^{n+1-k} y^k + \binom{n}{k} x^{n-k} y^{k+1} \\
&= x^{n+1} + \sum_{k=1}^{n} \left\{ \binom{n}{k} + \binom{n}{k-1} \right\} x^{n+1-k} y^k + y^{n+1} \\
&= \sum_{k=0}^{n+1} \binom{n+1}{k} x^{n+1-k} y^k.
\end{aligned} \tag{B.6}$$

This establishes the induction step, and hence the theorem. □

C Summing p^{th} powers of integers

Using Induction (Appendix A), it is possible to prove results such as

Theorem C.1. *For p a positive integer*

$$\sum_{k=1}^{n} k^p \;=\; f_p(n), \tag{C.1}$$

where $f_p(x)$ is a polynomial of degree $p+1$ in x with rational coefficients, and the leading term is $\frac{x^{p+1}}{p+1}$.

See Remark A.3 for more details. It is also possible to prove these results *without* resorting to induction! Namely, we can prove these results by differentiating identities. We need the following result about finite geometric series:

Lemma C.2. *For any $x \in \mathbb{R}$,*

$$1 + x + x^2 + \cdots + x^n \;=\; \frac{x^{n+1}-1}{x-1}. \tag{C.2}$$

Proof. If $x = 1$ we evaluate the right hand side by L'Hospital's Rule, which gives $\frac{n+1}{1} = n+1$. For other x, let $S = 1 + x + \cdots + x^n$. Then

$$\begin{aligned} S &= 1 + x + x^2 + \cdots + x^n \\ xS &= \phantom{1 + {}} x + x^2 + \cdots + x^n + x^{n+1}. \end{aligned} \tag{C.3}$$

Therefore

$$xS - S \;=\; x^{n+1} - 1 \tag{C.4}$$

or

$$S \;=\; \frac{x^{n+1}-1}{x-1}. \tag{C.5}$$

□

We now show how to sum the p^{th} powers of the first n integers. We first investigate the case when $p = 1$ and provide an alternate proof of Theorem A.1. Consider the identity

$$\sum_{k=0}^{n} x^k \;=\; \frac{x^{n+1}-1}{x-1}. \tag{C.6}$$

We apply the operator $x\frac{d}{dx}$ to each side and obtain

$$\begin{aligned} x\frac{d}{dx}\sum_{k=0}^{n} x^k &= x\frac{d}{dx}\frac{x^{n+1}-1}{x-1} \\ \sum_{k=0}^{n} kx^k &= x\frac{(n+1)x^n\cdot(x-1) - 1\cdot(x^{n+1}-1)}{(x-1)^2} \\ \sum_{k=0}^{n} kx^k &= x\frac{nx^{n+1}-(n+1)x^n+1}{(x-1)^2}. \end{aligned} \tag{C.7}$$

If we set $x = 1$, the left hand side becomes the sum of the first n integers. To evaluate the right hand side we use L'Hospital's rule, as when $x = 1$ we get $1 \cdot \frac{0}{0}$. As long as one of the factors has a limit, the limit of a product is the product of the limits. As $x \to 1$, the factor of x becomes just 1 and we must study $\lim_{x\to 1} \frac{nx^{n+1}-(n+1)x^n+1}{(x-1)^2}$. We find

$$\lim_{x\to 1} \frac{nx^{n+1} - (n+1)x^n + 1}{(x-1)^2} = \lim_{x\to 1} \frac{n(n+1)x^n - n(n+1)x^{n-1}}{2(x-1)}. \tag{C.8}$$

As the right hand side is $\frac{0}{0}$ when $x = 1$ we apply L'Hospital again and find

$$\begin{aligned} \lim_{x\to 1} \frac{nx^{n+1} - (n+1)x^n + 1}{(x-1)^2} &= \lim_{x\to 1} \frac{n^2(n+1)x^{n-1} - n(n+1)(n-1)x^{n-1}}{2} \\ &= \frac{n(n+1)}{2}. \end{aligned} \tag{C.9}$$

Therefore, by differentiating the finite geometric series and using L'Hospital's rule we were able to prove the formula for the sum of integers *without* resorting to induction. The reason we used the operator $x\frac{d}{dx}$ and not $\frac{d}{dx}$ is this leaves the power of x unchanged. While this flexibility is not needed to compute sums of first powers of integers, if we want to calculate sums of k^p for $p > 1$, this will simplify the formulas.

Theorem C.3. *For n a positive integer,*

$$\sum_{k=0}^{n} k^2 x^k = \frac{n(n+1)(2n+1)}{6}. \tag{C.10}$$

Proof. To find the sum of k^2 we apply $x\frac{d}{dx}$ twice to (C.6) and get

$$\begin{aligned} x\frac{d}{dx}\left[x\frac{d}{dx}\sum_{k=0}^{n} x^k\right] &= x\frac{d}{dx}\left[x\frac{d}{dx}\frac{x^{n+1}-1}{x-1}\right] \\ x\frac{d}{dx}\sum_{k=0}^{n} kx^k &= x\frac{d}{dx}\left[x\frac{nx^{n+1}-(n+1)x^n+1}{(x-1)^2}\right] \\ \sum_{k=0}^{n} k^2x^k &= x\frac{d}{dx}\left[\frac{nx^{n+2}-(n+1)x^{n+1}+x}{(x-1)^2}\right] \\ \sum_{k=0}^{n} k^2x^k &= x\frac{\left[n(n+2)x^{n+1}-(n+1)^2x^n+1\right]\cdot(x-1)^2}{(x-1)^4} \\ &\quad -x\frac{\left[nx^{n+2}-(n+1)x^{n+1}+x\right]\cdot 2(x-1)}{(x-1)^4}. \end{aligned} \tag{C.11}$$

Simple algebra (multiply everything out on the right hand side and collect terms) yields

$$\sum_{k=0}^{n} k^2x^k = x\frac{n^2x^{n+2}-(2n^2+2n-1)x^{n+1}+(n^2+2n+1)x^n-x-1}{(x-1)^3}. \tag{C.12}$$

The left hand side is the sum we want to evaluate; however, the right hand side is $\frac{0}{0}$ for $x = 1$. As the denominator is $(x-1)^3$ it is reasonable to expect that we will need to apply L'Hospital's rule three times; we provide a proof of this in Remark C.4.

Applying L'Hospital's rule three times to the right hand side we find the right hand side is

$$\frac{n^2(n+2)(n+1)nx^{n-1} - (2n^2+2n-1)(n+1)n(n-1)x^{n-2} + (n^2+2n+1)n(n-1)(n-2)x^{n-3}}{3 \cdot 2 \cdot 1}. \tag{C.13}$$

Taking the limit as $x \to 1$ we obtain

$$\begin{aligned} \sum_{k=0}^{n} k^2 x^k &= \frac{n^2(n+2)(n+1)n - (2n^2+2n-1)(n+1)n(n-1) + (n^2+2n+1)n(n-1)(n-2)}{6} \\ &= \frac{n(n+1)(2n+1)}{6}, \end{aligned} \tag{C.14}$$

where the last line follows from simple algebra. □

Remark C.4. *While we are able to obtain the correct formula for the sum of squares without resorting to induction, the algebra is starting to become tedious, and will get more so for sums of higher powers. After applying $x\frac{d}{dx}$ twice we had $\frac{g(x)}{(x-1)^3}$, where $g(x)$ is a polynomial of degree $n+2$ and $g(1) = 0$. It is natural to suppose that we need to apply L'Hospital's rule three times as we have a factor of $(x-1)^3$ in the denominator. However, if $g'(1)$ or $g''(1)$ is not zero, then we do not apply L'Hospital's rule three times but rather only once or twice. Thus we really need to check and make sure that $g'(1) = g''(1) = 0$. While a straightforward calculation will show this, a moment's reflection shows us that both of these derivatives must vanish. If one of them was non-zero, say equal to a, then we would have $\frac{a}{0}$ which is undefined; however, clearly the sum of the first n squares is finite. Therefore these derivatives will be zero and we do have to apply L'Hospital's rule three times.*

Remark C.5. *For those concerned about the legitimacy of applying L'Hospital's rule and these formulas when $x = 1$, we can consider a sequence of x's, say $x_N = 1 - \frac{1}{N}$ with $N \to \infty$. Everything is then well-defined, and it is of course natural to use L'Hospital's rule to evaluate $\lim_{N\to\infty} \frac{g(x_N)}{(x_N-1)^3}$.*

D Divergence of the Harmonic Series

Instead of considering (1.3):

$$\sum_{n=1}^{\infty} n \cdot x^{n-1} \;=\; 1 + 2x + 3x^2 + 4x^3 + \cdots, \tag{D.1}$$

let us consider

$$\sum_{n=1}^{\infty} \frac{x^n}{n} \;=\; \frac{x}{1} + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \cdots. \tag{D.2}$$

Taking $x = 1$ yields the famous harmonic series

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots. \tag{D.3}$$

There are many proofs of the divergence of the harmonic series (one simple one is to group terms 2^k through $2^{k+1}-1$ together, note each term is at least $\frac{1}{2^k}$ so their sum is at least 1; another is to multiply by $\frac{1}{2}$ and subtract and note that the sum of the reciprocals of the odd numbers is greater than the sum of the reciprocals of the even numbers); we present a proof based on differentiating (or actually integrating) identities. Let

$$F(x) \ = \ \sum_{n=1}^{\infty} \frac{x^n}{n} \ = \ \frac{x}{1} + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \cdots . \tag{D.4}$$

If $|x| < 1$ then the above series converges. We now differentiate and note that if $|x| < 1$ then we can interchange differentiation and summation, and we find

$$f(x) \ = \ F'(x) \ = \ \sum_{n=1}^{\infty} x^{n-1} \ = \ 1 + x + x^2 + x^3 + \cdots \ = \ \frac{1}{1-x}, \tag{D.5}$$

where the last equality follows by the geometric series formula. We now integrate the above from 0 to u for $u \in (0,1)$ (so we may interchange integration and summation). We find

$$F(u) \ = \ \sum_{n=1}^{\infty} \frac{u^n}{n} \ = \ \int_0^u \frac{dx}{1-x} \ = \ -\log(1-u). \tag{D.6}$$

We have therefore shown that for $u \in (0,1)$ that

$$\frac{u}{1} + \frac{u^2}{2} + \frac{u^3}{3} + \frac{u^4}{4} + \cdots \ = \ -\log(1-u). \tag{D.7}$$

We now take the limit as $u \to 1$ from below; let us write u as $u = 1 - e^{-t}$ with t positive and $t \to \infty$. Then the left hand side of (D.7) as $t \to \infty$ is just the harmonic sum,

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots . \tag{D.8}$$

For a fixed t, the right hand side is $-\log(e^{-t}) = t$. Thus as $u \to 1$ from below $t \to \infty$ and $-\log(1-u) \to \infty$. Thus the harmonic series diverges.

Exercise D.1. *Use the above arguments to show that the sum of the first N terms of the harmonic series is of size* $\log N$.

E Interchanging Differentiation and Summation

We first recall Fubini's Theorem, which states when we may interchange orders of integration.

Theorem E.1 (Fubini)**.** *Assume f is continuous and*

$$\int_a^b \int_c^d |f(x,y)| dx dy \ < \ \infty. \tag{E.1}$$

Then

$$\int_a^b \left[\int_c^d f(x,y)dy\right] dx \ = \ \int_c^d \left[\int_a^b f(x,y)dx\right] dy. \tag{E.2}$$

Similar statements hold if we instead have

$$\sum_{n=N_0}^{N_1} \int_c^d f(x_n,y)dy, \qquad \sum_{n=N_0}^{N_1} \sum_{m=M_0}^{M_1} f(x_n,y_m). \tag{E.3}$$

For a proof in special cases, see [BL, VG]; an advanced, complete proof is given in [Fol]. One cannot always interchange orders of integration. For simplicity, we give a sequence a_{mn} such that $\sum_m(\sum_n a_{m,n}) \neq \sum_n(\sum_m a_{m,n})$ (although it is trivial to modify this to an example involving integrals).

Exercise E.2. *For $m,n \geq 0$ let*

$$a_{m,n} \ = \ \begin{cases} \ \ 1 & \textit{if } n = m \\ -1 & \textit{if } n = m+1 \\ \ \ 0 & \text{otherwise.} \end{cases} \tag{E.4}$$

Show that the two different orders of summation yield different answers.

We now study when we can justify interchanging orders of differentiation and summation. We know the geometric series formula gives

$$\sum_{n=0}^{\infty} x^n \ = \ \frac{1}{1-x}, \quad |x| < 1. \tag{E.5}$$

We show that we may interchange differentiation and summation above. The derivative of the right hand side (with respect to x) is just $(1-x)^{-2}$. We want to say the derivative of the left hand side is

$$\sum_{n=0}^{\infty} nx^{n-1}, \tag{E.6}$$

but do to so requires us to justify

$$\frac{d}{dx}\sum_{n=0}^{\infty} x^n \ = \ \sum_{n=0}^{\infty} \frac{d}{dx} x^n. \tag{E.7}$$

A standard way to justify statements like this is as follows. We note that $\sum_{n=0}^{\infty} nx^{n-1}$ converges for $|x| < 1$; if we can show that for any $\epsilon > 0$ that this is within ϵ of $(1-x)^{-2}$, then we will have justified the interchange.

To see this, fix an $\epsilon > 0$. For each N, we may write

$$\begin{aligned} \sum_{n=0}^{\infty} x^n \ &= \ \sum_{n=0}^{N} x^n + \sum_{n=N+1}^{\infty} x^n \\ &= \ \sum_{n=0}^{N} x^n + \frac{x^{N+1}}{1-x} \ = \ \frac{1}{1-x}. \end{aligned} \tag{E.8}$$

We can differentiate each side, and we can justify interchanging the differentiation and the summation because we have *finite* sums. Specifically, there are only $N+2$ terms ($N+1$ from the sum and then one more, $\frac{x^{N+1}}{1-x}$). Therefore we have

$$\begin{aligned}
\frac{d}{dx}\sum_{n=0}^{N} x^n + \frac{d}{dx}\frac{x^{N+1}}{1-x} &= \frac{d}{dx}\frac{1}{1-x} \\
\sum_{n=0}^{N} nx^{n-1} + \frac{(N+1)x^N(1-x) - x^{N+1}(-1)}{(1-x)^2} &= \frac{1}{(1-x)^2} \\
\sum_{n=0}^{N} nx^{n-1} + \frac{(N+1)(1-x)+x}{(1-x)^2}x^N &= \frac{1}{(1-x)^2} \qquad \text{(E.9)}
\end{aligned}$$

As $|x| < 1$, given any $\epsilon > 0$ we can find an N_0 such that for all $N \geq N_0$,

$$\left|\frac{(N+1)(1-x)+x}{(1-x)^2}x^N\right| \leq \frac{\epsilon}{2}. \qquad \text{(E.10)}$$

Similarly we can find an N_1 such that for all $N \geq N_1$ we have

$$\left|\sum_{n=N+1}^{\infty} nx^{n-1}\right| \leq \frac{\epsilon}{2}. \qquad \text{(E.11)}$$

Therefore we have shown that for every $\epsilon > 0$ we have

$$\left|\frac{1}{(1-x)^2} - \sum_{n=0}^{\infty} nx^{n-1}\right| \leq \epsilon, \qquad \text{(E.12)}$$

proving the claim. Instead of studying these sums for a specific x, we can consider $x \in [a,b]$ with $-1 < a \leq b < 1$, and N_0, N_1 will just depend on a, b and ϵ.

One situation where we cannot interchange differentiation and summation is when we have series that are conditionally convergent but not absolutely convergent. This means $\sum a_n$ converges but $\sum |a_n|$ does not. For example, consider

$$\sum_{n=0}^{\infty} \frac{x^n}{n}. \qquad \text{(E.13)}$$

If $x = -1$ this series conditionally converges but not absolutely; in fact, as

$$-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots = \sum_{n=1}^{\infty} \frac{x^n}{n}, \qquad \text{(E.14)}$$

then (E.13) with $x = -1$ is just $-\log 2$. What happens if we try to differentiate? We have

$$\frac{d}{dx}\left[-\log(1-x)\right] = \frac{d}{dx}\left[\sum_{n=1}^{\infty} \frac{x^n}{n}\right]. \qquad \text{(E.15)}$$

The left hand side is easy to differentiate for $x \in [-1, 0]$, giving $\frac{1}{1-x}$. But if we interchange the differentiation and summation we would have

$$\frac{d}{dx}\left[\sum_{n=1}^{\infty} \frac{x^n}{n}\right] = \sum_{n=1}^{\infty} x^{n-1}, \tag{E.16}$$

and this does not converge when $x = -1$ (aside: the sum oscillates between 1 and 0; in some sense it can be interpreted as $\frac{1}{2}$, which is what $\frac{1}{1-x}$ equals when $x = -1$!).

Sometimes, however, conditionally convergent but absolutely divergent series can be managed. Consider

$$\sum_{n=1}^{\infty} \frac{x^n}{n \log n}. \tag{E.17}$$

This series converges conditionally when $x = -1$ but diverges upon inserting absolute values. If we interchange differentiation and summation we get

$$\sum_{n=1}^{\infty} \frac{x^{n-1}}{\log n}, \tag{E.18}$$

and this sum does converge (conditionally, not absolutely) when $x = -1$.

References

[BL] P. Baxandall and H. Liebeck, *Vector Calculus*, Clarendon Press, 1986.

[Fol] G. Folland, *Real Analysis : Modern Techniques and Their Applications*, Pure and Applied Mathematics, Wiley-Interscience, second edition, 1999.

[VG] W. Voxman and R. Goetschel, Jr., *Advanced Calculus*, Mercer Dekker Inc., 1981.