BINOMIAL AND MULTINOMIAL COEFFICIENTS

In what may seem an odd turnaround, we'll start this week with some combinatorics.

We have previously seen the *binomial formula*:

$$(a+b)^k = \sum_{i=0}^{k} \binom{k}{i} a^i b^{k-i}, \qquad \binom{k}{i} = \frac{k!}{i!(k-i)!}.$$

Whenever you see a factorial, it generally means you're counting possibilities. In this case, the coefficient of the ith term in the expansion above is counting the number of ways that you could choose i factors of a and $k-i$ factors of b to make a monomial of degree k. The thing about counting is that you can usually only count things that are in order.

Given n objects, there are $n! = n(n-1)(n-2)\cdots 2\cdot 1$ ways to order them. The first thing we need to order is the factors of $(a+b)^k$:

$$(a+b)_1(a+b)_2\cdots(a+b)_{n-1}(a+b)_k.$$

By changing the indices, we have changed the order of the factors, although the factors are themselves indistiguishable. We will give them a trait that distinguishes (some of) them: from the first i that appear, we choose the a factor to multiply out, and from the last $k-i$, we choose the b factor. This gives a term that will definitely appear in the expansion of $(a+b)^k$. But if we change the order of the first i terms and the order of the last $k-i$ terms, without mixing them, we will get a "new" term, which in truth we have already counted because we chose the same term, a or b, from the same factors. So these changes in ordering *reduce* the number of terms. Thus we divide by the number of such orderings to get that there are $k!/i!(k-i)!$ terms with a appearing i times and b appearing $k-i$ times.

We're now working with polynomials of more than two variables. We need a more general formula, called the *multinomial formula*:

$$(a_1+a_2+\cdots+a_n)^k = \sum_{b_1+\cdots+b_n=k} \binom{k}{b_1,\ldots,b_n} a_1^{b_1}\cdots a_n^{b_n}, \qquad \binom{k}{b_1,\ldots,b_n} = \frac{k!}{b_1!\cdots b_n!}.$$

This formula makes an appearance in the proof of Taylor's Theorem that I've posted online. One of your homework exercises asks you a related question: "What is the number of multi-indices of total degree k, with n entries?" (Hint: it's not any number that appears in the above formula.) At this point, parts of this expression should look familiar. Specifically, if we collect all the b_is into a single object $\beta = (b_1,\ldots,b_n)$, then the denominatorof the multinomial coefficient becomes $\beta!$, and the monomial part becomes $\mathbf{a}^\beta$, where $\mathbf{a}$ is the vector with entries a_i. All we're missing is a few derivatives for this to look exactly like the Taylor polynomial of a function f.

TAYLOR POLYNOMIALS

Let's go back to the expression for the kth-order Taylor polynomial of a function f, expanded around $\mathbf{a}$ (we'll not worry here about the error term):

$$f(\mathbf{a}+\mathbf{h}) \approx \sum_{|I|\le k} \frac{D^I f(\mathbf{a})}{I!}\mathbf{h}^I.$$

If $f : \mathbb{R}^n \to \mathbb{R}$ is a polynomial of degree $\leq k$, and we expand around $\mathbf{0}$, then this should of course just equal f. Can we see why that's true? If D^I is applied to a term where some entry in I surpasses the degree of the corresponding variable in $\mathbb{R}^n$, then D^I kills that term. Conversely, if some entry of I is less than the degree of the corresponding variable, then that variable will still appear in the differentiated form, and evaluating at 0 will eliminate it. So the only time a term of f contributes to the sum is when I matches exactly the exponents of the term. For example,

$$
\begin{aligned}
D^{(2,0,1)}xyz\Big|_{\left[\begin{smallmatrix}x\\y\\z\end{smallmatrix}\right]=\left[\begin{smallmatrix}0\\0\\0\end{smallmatrix}\right]} &= 0\Big|_{\left[\begin{smallmatrix}x\\y\\z\end{smallmatrix}\right]=\left[\begin{smallmatrix}0\\0\\0\end{smallmatrix}\right]} = 0 && \text{(too hot)}\\
D^{(2,0,1)}x^2yz\Big|_{\left[\begin{smallmatrix}x\\y\\z\end{smallmatrix}\right]=\left[\begin{smallmatrix}0\\0\\0\end{smallmatrix}\right]} &= 2y\Big|_{\left[\begin{smallmatrix}x\\y\\z\end{smallmatrix}\right]=\left[\begin{smallmatrix}0\\0\\0\end{smallmatrix}\right]} = 0 && \text{(too cold)}\\
D^{(2,0,1)}x^2z\Big|_{\left[\begin{smallmatrix}x\\y\\z\end{smallmatrix}\right]=\left[\begin{smallmatrix}0\\0\\0\end{smallmatrix}\right]} &= 2\Big|_{\left[\begin{smallmatrix}x\\y\\z\end{smallmatrix}\right]=\left[\begin{smallmatrix}0\\0\\0\end{smallmatrix}\right]} = 2 && \text{(just right)}
\end{aligned}
$$

Each time we differentiate with respect to a variable, we pick up a factor equal to the old exponent of the variable; this means D^I of the term will be $I!$ times the term's original coefficient. Hence we must divide out by $I!$.

Likewise, if we have a power series for a function, then the kth Taylor polynomial for the function should just be the first k terms of the series. For example, the kth order Taylor polynomial of $x \mapsto (1-x)^{-1}$ centered at 0 is $1 + x + \cdots + x^k$, because

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots = \sum_{i=0}^{\infty} x^i \qquad \text{for } |x| < 1.$$

This last observation, along with the following (loosely stated) proposition, make computing Taylor series for certain functions quite easy.

Proposition. The power series of a composition equals the composition of the power series. (Chain rule is in effect; see second example for explanation.)

Example 1. The function $\begin{pmatrix} x\\ y\\ z\end{pmatrix} \mapsto e^{x+xy^2+z^2}$ is the composition of a function $g : \mathbb{R}^3 \to \mathbb{R}$ and $\exp : \mathbb{R} \to \mathbb{R}$, each of which is expressible by a power series. Recall that

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots + \frac{x^k}{k!} + \cdots .$$

Suppose we want to compute the degree 3 Taylor polynomial of the composition. At most, we'll have to consider the terms up to degree 3 in each of the power series, because composing will only increase the exponents. g is its own degree 3 Taylor polynomial, and $P^3_{\exp}(x) = 1 + x + x^2/2 + x^3/6$. So we compose these polynomials and take just the terms of degree 3 or lower:

$$
\begin{aligned}
P^3_{\exp\circ g}\begin{pmatrix} x\\ y\\ z\end{pmatrix} &= \text{cubic}\left[1 + (x + xy^2 + z^2) + \frac{(x+xy^2+z^2)^2}{2} + \frac{(x+xy^2+z^2)^3}{6}\right]\\
&= 1 + x + xy^2 + z^2 + \frac{1}{2}(x^2 + 2xz^2) + \frac{1}{6}x^3.
\end{aligned}
$$

Example 2. The binomial formula and the series for $(1-x)^{-1}$ are special cases of a more general formula:

$$(1+x)^p \approx 1 + px + \frac{p(p-1)}{2!}x^2 + \frac{p(p-1)(p-2)}{3!}x^3 + \cdots + \frac{p(p-1)\cdots(p-(k-1))}{k!}x^k,$$

where p is not necessarily an integer. Let's use this to compute the degree 4 Taylor polynomial of $x \mapsto \sqrt{\cos x}$ at 0. At a certain point we must be careful, because we have the Taylor polynomials for cosine and $(1+x)^{1/2}$ *computed at* 0. Once we've run x through the cosine, however, we're no longer at 0, and we have to think about what terms we should use in the composition.

$$\begin{aligned}
P^4_{\sqrt{\cos}}(x) &= \text{quartic}\left[\sqrt{1 - \frac{x^2}{2!} + \frac{x^4}{4!}}\right] \\
&= \text{quartic}\left[1 + \frac{1}{2}\left(-\frac{x^2}{2!} + \frac{x^4}{4!}\right) + \frac{1}{2!}\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{x^2}{2!} + \frac{x^4}{4!}\right)^2\right] \\
&= 1 - \frac{x^2}{4} + \frac{x^4}{48} - \frac{1}{8}\left(\frac{x^4}{4}\right) = 1 - \frac{x^2}{4} - \frac{x^4}{96}.
\end{aligned}$$

Covariant tensors

Let's return to a discussion of the information that the Taylor polynomial holds. The terms of degree 1 add up to

$$\sum_{|I|=1} D^I f(\mathbf{a}) \cdot \mathbf{h}^I = (D_1 f(\mathbf{a}))h_1 + (D_2 f(\mathbf{a}))h_2 + \cdots + (D_n f(\mathbf{a}))h_n = [Df(\mathbf{a})]\mathbf{h}.$$

Recall from long ago, when derivatives were being introduced, that this is exactly the directional derivative in the direction of $\mathbf{h}$, i.e.,

$$D_{\mathbf{h}} f(\mathbf{a}) = [Df(\mathbf{a})]\mathbf{h}.$$

Thus the degree 1 Taylor polynomial of f at $\mathbf{a}$ can be written

$$P^1_f(\mathbf{h}) = f(\mathbf{a}) + D_{\mathbf{h}} f(\mathbf{a}).$$

The directional derivative is a *covariant tensor of rank 1*.

Now suppose we've chosen k vectors $\mathbf{h}_1, \ldots, \mathbf{h}_k$ in $\mathbb{R}^n$. Then we can compute

$$[D^k f(\mathbf{a})](\mathbf{h}_1, \ldots, \mathbf{h}_k) := D_{\mathbf{h}_1}(D_{\mathbf{h}_2}(\cdots(D_{\mathbf{h}_k} f)\cdots))(\mathbf{a}).$$

As you saw in yesterday's lecture, this function $(\mathbb{R}^n)^k \to \mathbb{R}$ is linear in each of its components. Moreover, because f is C^k, its mixed partial derivatives of order $\le k$ commute, which implies that $D^k f(\mathbf{a})$ is a *symmetric* multilinear form; switching the order of the $\mathbf{h}_i$ doesn't change the resulting value. This function is a *covariant tensor of rank k*. Essentially, it has k "slots" for vectors in $\mathbb{R}^n$, and requires each of these slots to be filled before it evaluates to a real number. If we set $\mathbf{h}_1 = \cdots = \mathbf{h}_k = \mathbf{h}$, then we obtain a *homogeneous polynomial of degree k* from $\mathbb{R}^n$ to $\mathbb{R}$, defined by

$$\mathbf{h} \mapsto [D^k f(\mathbf{a})]\mathbf{h}^{\otimes k}, \qquad \text{where } \mathbf{h}^{\otimes k} := (\mathbf{h}, \ldots, \mathbf{h}) \in (\mathbb{R}^n)^k.$$

The k-linear function $[D^k f(\mathbf{a})]$ captures all of the information about the kth-order partial derivatives of f at $\mathbf{a}$, as a higher-order generalization of the way the Jacobian captures the first partials and the Hessian captures the second partials. Because of this, we can rewrite the Taylor polynomial of f in terms of these multilinear forms:

$$f(\mathbf{a}+\mathbf{h}) \approx \sum_{i=0}^{k} \frac{1}{i!}[D^i f(\mathbf{a})]\mathbf{h}^{\otimes i}.$$

The presence of the $1/i!$ coefficient shouldn't be surprising: $D^i f(\mathbf{a})$ is a rank i tensor, and we're plugging in i copies of $\mathbf{h}$. "Shuffling the order" of these i identical copies has no effect, so we should take the $i!$ possible orders into account. Let's see if we can understand this more precisely.

Simply put, this is the reappearance of the multinomial formula in the guise of differential operators. Recall that partial derivatives are precisely directional derivatives, taken in the direction of a standard basis vector; i.e.,$D_j f(\mathbf{a}) = D_{\mathbf{e}_j} f(\mathbf{a})$. Now we use the fact that $\mathbf{h}$ is a linear combination of the $\mathbf{e}_j$, and that covariant tensors are linear in each of their slots. For example,

$$\begin{aligned}
[D^2 f(\mathbf{a})] \begin{pmatrix}1\\1\\1\end{pmatrix}^{\otimes 2} =& D_{\mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3}(D_{\mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3} f)(\mathbf{a}) \\
=& D_{\mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3}(D_1 f + D_2 f + D_3 f)(\mathbf{a}) \\
=& D_1\big(D_1 f + D_2 f + D_3 f\big)(\mathbf{a}) + D_2\big(D_1 f + D_2 f + D_3 f\big)(\mathbf{a}) \\
&\quad + D_3\big(D_1 f + D_2 f + D_3 f\big)(\mathbf{a}) \\
=& D_1 D_1 f(\mathbf{a}) + D_1 D_2 f(\mathbf{a}) + D_1 D_2 f(\mathbf{a}) + D_2 D_1 f(\mathbf{a}) + D_2 D_2 f(\mathbf{a}) + D_2 D_2 f(\mathbf{a}) \\
&\quad + D_3 D_1 f(\mathbf{a}) + D_3 D_2 f(\mathbf{a}) + D_3 D_2 f(\mathbf{a}),
\end{aligned}$$

which is the sum of all of the second-order partial derivatives for a function of three variables. If we have different coordinates, as in $\mathbf{h} = h_1\mathbf{e}_1 + h_2\mathbf{e}_2 + h_3\mathbf{e}_3$, then the h_j in each term simply come out as the appropriate monomial (e.g., $D_{h_1\mathbf{e}_1} D_{h_1\mathbf{e}_1} D_{h_2\mathbf{e}_2} f(\mathbf{a}) = h_1^2 h_2 D_1 D_1 D_2 f(\mathbf{a})$). But, thanks to equality of mixed partials, some of these terms are duplicated. The multinomial formula accounts for all of this.

Thus the more general question is: how many times does each ith-order partial derivative appear in $D^i f(\mathbf{a})$? Suppose we're looking for the partial derivative $D^{(b_1,\ldots,b_n)}$, where $b_1+\cdots+b_n = i$. We evaluate this rank i tensor on i copies of $\mathbf{h}$, which now plays the rôle of the multinomial in the multinomial formula. The coefficient $D^{(b_1,\ldots,b_n)} f(\mathbf{a})$ appears whenever we choose to look at the jth element of $\mathbf{h}$ b_j times. As in the multinomial formula, we thus have that

$$[D^{(b_1,\ldots,b_n)} f(\mathbf{a})] h_1^{b_1}\cdots h_n^{b_n} \quad \text{appears} \quad \binom{i}{b_1,\ldots,b_n} = \frac{i!}{b_1!\cdots b_n!} \text{ times in } [D^i f(\mathbf{a})]\mathbf{h}^{\otimes i}.$$

This is the right extra factor in the Taylor polynomial as we saw it before, except for the $i!$ in the numerator. Once we divide out that extra factor, we're back to our original expression for Taylor polynomials.