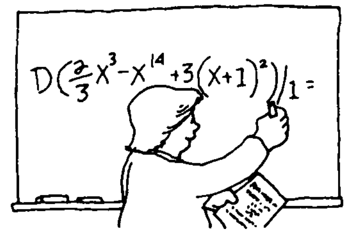
CLASSROOM CAPSULES

EDITOR

Warren Page
30 Amberson Ave.
Yonkers, NY 10705



Classroom Capsules consists primarily of short notes (1–3 pages) that convey new mathematical insights and effective teaching strategies for college mathematics instruction. Please submit manuscripts prepared according to the guidelines on the inside front cover to the Editor, Warren Page, 30 Amberson Ave., Yonkers, NY 10705-3613.

The Parable of the Lucky Student?

Vince Matsko (matskvi@quincy.edu), Quincy University, Quincy, IL 62301

Correct answers obtained by techniques that should not work are always beguiling. In [**1**], Dan Kennedy discussed such a situation encountered in grading the 1996 AB Advanced Placement exam.

> Let R be the region in the first quadrant under the graph of $y = 1/\sqrt{x}$ for $4 \leq x \leq 9$.
>
> (a) Find the area of R.
> (b) If the line $x = k$ divides the region R into two regions of equal area, what is the value of k?
> (c) Find the volume of the solid whose base is the region R and whose cross sections cut by planes perpendicular to the x-axis are squares.

When solving (b) for $0 < a < b$, the problem of finding k so that

$$\int_a^k \frac{1}{\sqrt{x}}\,dx = \int_k^b \frac{1}{\sqrt{x}}\,dx \tag{1}$$

has solution

$$k = \left(\frac{\sqrt{a}+\sqrt{b}}{2}\right)^2.$$

As Dan noted, some students found the appropriate value of k by solving

$$\frac{1}{\sqrt{k}} = \frac{1}{b-a}\int_a^b \frac{1}{\sqrt{x}}\,dx. \tag{2}$$

In other words, they found the average value of $1/\sqrt{x}$ over the interval $[a, b]$. Students were given full credit for this solution because the point at which the average value occurs *does* in fact give the correct solution. Is $1/\sqrt{x}$ the only function with this property? We will answer this question more or less in the affirmative.

We first assume that f is a positive, differentiable function and that for distinct a, b, there is exactly one number $p(a, b)$ between a and b satisfying

$$\int_a^{p(a,b)} f(x)\,dx = \int_{p(a,b)}^b f(x)\,dx. \tag{3}$$

Assume, furthermore, that p is "nice" in that all second-order partials exist, and that $p_{ab} = p_{ba}$ (we use the subscript notation here to denote partial differentiation, and omit writing the arguments (as in "$p_{ab}(a, b)$") to avoid cumbersome expressions). We also assume that p satisfies the appropriate "average-value" equation

$$(b - a) f(p) = \int_a^b f(x)\,dx. \tag{4}$$

Now taking partials of (3) with respect to a and b, respectively, yields

$$f(a) = 2f(p)p_a, \quad f(b) = 2f(p)p_b. \tag{5}$$

Note that since f is positive, p_a and p_b must also be positive.

From (5) we have

$$f(a)p_b = f(b)p_a. \tag{6}$$

Next, taking partials of (4) with respect to a and b, and then eliminating the "$f'(p)$" terms from these equations results in

$$f(p)(p_a + p_b) = f(b)p_a + f(a)p_b. \tag{7}$$

Substituting from (5) into (7) and using the fact that f is positive, we see that

$$p_a + p_b = 4p_a p_b. \tag{8}$$

Solve (8) for p_b, substitute back into (6), and solve for p_a to obtain

$$p_a = \frac{1}{4}\left(1 + \frac{f(a)}{f(b)}\right). \tag{9}$$

Solving (8) for p_a, substituting into (6), and solving for p_b results in

$$p_b = \frac{1}{4}\left(1 + \frac{f(b)}{f(a)}\right). \tag{10}$$

Finding p_{ab} from (9) and p_{ba} from (10), and using $p_{ab} = p_{ba}$, we have

$$\frac{f'(a)}{f(a)^3} = \frac{f'(b)}{f(b)^3}.$$

Since the left-hand side of this equation is a function of a and the right is a function of b, both sides must be constant. Solving by the usual means yields the existence of constants m and c such that

$$f(x) = \frac{1}{\sqrt{mx + c}}.$$

Note that when $m = 0$, the function $f(x)$ is constant and therefore, $p(a, b) = \frac{1}{2}(a + b)$. When $m \neq 0$, the graph of this function lies to the left or right of the vertical asymptote $x = -\frac{c}{m}$ depending on whether the sign of $\frac{c}{m}$ is negative or positive. So aside from constant functions, the only positive functions with the desired property are $f(x) = 1/\sqrt{mx + c}$, where $m \neq 0$. A good thing for our students!

Remark: Filling in the details of the above calculations is a great exercise for third-semeseter calculus students.

Reference

1. Dan Kennedy, Things I have learned at the AP reading, *The College Mathematics Journal* **30**:5 (1999) 346–355.

Comparing ∅, {∅}, {{∅}}, etc.

Allen J. Schwenk (schwenk@wmich.edu), Western Michigan University, Kalamazoo, MI 49008-5152

I have often had the experience of explaining the difference between the empty set ∅ and the set whose element is the empty set {∅}. And then I get to {{∅}}. By the time I get through mouthing the words "The set whose element is the set whose element is the empty set," I give up and move on, disappointed that the students haven't seen the need for such subtleties. Perhaps the following example will help convince students that there really is a difference.

In a certain class, the professor reports to the students their performance on the weekly assignments. For each assignment, say assignment i, the professor provides student x with a set of problem numbers missed by x. That is, student x receives a set of numbers $P_x^i = \{j \mid \text{Student } x \text{ missed problem } j \text{ on assignment } i\}$. If x got them all right, $P_x^i = \emptyset$.

Now, for comparative purposes, the professor (preserving confidentiality) provides student x with the sets P_x^i for each student y who missed as many or fewer problems than x did. This is reported to x as a set $A_x^i = \{P_y^i \mid y \neq x \text{ and } |P_y^i| \leq |P_x^i|\}$. If $A_x^i = \emptyset$, then no one did as well or better than x. But if $A_x^i = \{\emptyset\}$, we know that only students with perfect papers were equal to or better than x.

Now this is repeated each week for k weeks, and student x forms a set $S_x^k = \{A_x^i \mid 1 \leq i \leq k\}$. If $S_x^k = \emptyset$, there have not been any assignments yet. If $S_x^k = \{\emptyset\}$, no one has matched or beaten x for the k weeks. If $S_x^k = \{\{\emptyset\}\}$, then for each week, only those with perfect papers have equaled or beaten x.

Example. An illustration with three students and two assignments.

$$\left.\begin{array}{l} P_1^1 = \emptyset \\ P_2^1 = \{4\} \\ P_3^1 = \{5,7\} \end{array}\right\} \Rightarrow \begin{array}{l} A_1^1 = \emptyset \\ A_2^1 = \{\emptyset\} \\ A_3^1 = \{\{4\},\emptyset\} \end{array} \qquad \left.\begin{array}{l} P_1^2 = \emptyset \\ P_2^2 = \emptyset \\ P_3^2 = \{3,9\} \end{array}\right\} \Rightarrow \begin{array}{l} A_1^2 = \{\emptyset\} \\ A_2^2 = \{\emptyset\} \\ A_3^2 = \{\emptyset,\emptyset\} = \{\emptyset\} \end{array} \; \Rightarrow \begin{array}{l} S_1^1 = \left\{A_1^1\right\} = \{\emptyset\} \\ S_2^1 = \left\{A_2^1\right\} = \{\{\emptyset\}\} \\ S_3^1 = \left\{A_3^1\right\} = \{\{\{4\},\emptyset\}\} \\ S_1^2 = \left\{A_1^1, A_1^2\right\} = \{\emptyset,\{\emptyset\}\} \\ S_2^2 = \left\{A_2^1, A_2^2\right\} = \{\{\emptyset\},\{\emptyset\}\} = \{\{\emptyset\}\} \\ S_3^2 = \left\{A_3^1, A_3^2\right\} = \{\{\{4\},\emptyset\},\{\emptyset\}\} \end{array}$$

If the professor compiles the set $T^k = \{S_x^k \mid x \text{ is a student in the class}\}$, what can $T^k = \{\{\{\emptyset\}\}\}$ possibly mean? For a class of three or more students, I maintain that each week at most one student failed to have a perfect paper. Can you explain why?

Can you invent a set that might be $\{\{\{\{\emptyset\}\}\}\}$ or $\{\{\{\{\{\emptyset\}\}\}\}\}$? How about if the professor has several classes, and the department has several professors?

Divergence of Series by Rearrangement

Bernard August (mollbern@hotmail.com) and Thomas J. Osler (osler@rowan.edu), Rowan University, Glassboro, NJ 08028

In [**1**], Michael Ecker proved the divergence of the harmonic series by a novel method, which we will call *divergence by rearrangement*. The main idea was this.

> It is known that a convergent series of positive terms can be rearranged in any way, and the sum remains the same. Suppose we are given a series of positive terms and we *assume it converges*, say to the value S. If upon rearranging the terms we obtain the new sum S' and are able to deduce the contradiction $S' \neq S$, then the series must be divergent.

In this capsule, we use this method to prove the divergence of the series

$$\sum_{n=1}^{\infty} \frac{1}{(n^{r_1}+a_1)^{p_1}\,(n^{r_2}+a_2)^{p_2}\cdots(n^{r_m}+a_m)^{p_m}}, \tag{1}$$

where all the a_k are nonnegative and all the r_k and p_k are positive with

$$r_1 p_1 + r_2 p_2 + \cdots + r_m p_m \leq 1.$$

Note that special cases of (1) include the harmonic series, the p series $\sum \frac{1}{n^p}$ for $0 < p \leq 1$, and such series as $\sum \frac{1}{\sqrt{n(n+1)}}$ and $\sum \frac{1}{\sqrt[3]{n^2+1}\,\sqrt[3]{n+2}}$.

To simplify the exposition, we consider the case in which only two factors appear. The reader will have no difficulty seeing how to generalize this proof to include m factors in series (1).

Theorem. Let a and b be nonnegative, and let r, s, p, q be positive with $rp + sq \leq 1$. Then the series

$$\sum_{n=1}^{\infty} \frac{1}{(n^r + a)^p\,(n^s + b)^q} \tag{2}$$

diverges.

Proof. Assume that the series converges to the value S. Then $S = S_{\text{odd}} + S_{\text{even}}$, where

$$\begin{aligned}
S &= \frac{1}{(1^r + a)^p\,(1^s + b)^q} + \frac{1}{(2^r + a)^p\,(2^s + b)^q} + \frac{1}{(3^r + a)^p\,(3^s + b)^q} + \cdots, \\
S_{\text{odd}} &= \frac{1}{(1^r + a)^p\,(1^s + b)^q} + \frac{1}{(3^r + a)^p\,(3^s + b)^q} + \frac{1}{(5^r + a)^p\,(5^s + b)^q} + \cdots, \\
S_{\text{even}} &= \frac{1}{(2^r + a)^p\,(2^s + b)^q} + \frac{1}{(4^r + a)^p\,(4^s + b)^q} + \frac{1}{(6^r + a)^p\,(6^s + b)^q} + \cdots.
\end{aligned}$$

It is clear by comparing corresponding terms that

$$S_{\text{odd}} > S_{\text{even}}. \tag{3}$$

Observe also that

$$S_{\text{even}} = \frac{1}{2^{rp+sq}} \left[\frac{1}{\left(1^r + \frac{a}{2^r}\right)^p \left(1^s + \frac{b}{2^s}\right)^q} + \frac{1}{\left(2^r + \frac{a}{2^r}\right)^p \left(2^s + \frac{b}{2^s}\right)^q} + \frac{1}{\left(3^r + \frac{a}{2^r}\right)^p \left(3^s + \frac{b}{2^s}\right)^q} + \cdots \right]$$

Let S^* denote the series above in square brackets. Thus,

$$S_{\text{even}} = \frac{1}{2^{rp+sq}} S^*.$$

Since $rp + sq \le 1$, we have $\frac{1}{2^{rp+sq}} \ge \frac{1}{2}$. So

$$S_{\text{even}} \ge \frac{1}{2} S^*. \tag{4}$$

Comparing S^* with S term by term, we have $S^* \ge S$. Therefore,

$$S_{\text{even}} \ge \frac{1}{2} S^* \ge \frac{1}{2} S. \tag{5}$$

From (3) and (5), we now have $S = S_{\text{odd}} + S_{\text{even}} > S_{\text{even}} + S_{\text{even}} \ge S$. Thus, we arrive at the contradiction $S > S$ and thereby prove the theorem.

Reference

1. Michael W. Ecker, Divergence of the harmonic series by rearrangement, *College Mathematics Journal* **28** (1997) 209–210.

Almost-Binomial Random Variables

Peter Thompson (thompsop@wabash.edu), Wabash College, Crawfordsville, IN 47933

The parameter n in the binomial distribution is a positive integer. What happens if we allow it to be any positive number? The new distribution we get is not only easy to work with but is useful in approximation situations that beginning statistics students could encounter. In addition, applications involving the distribution provide a good source for undergraduate research problems.

We begin by setting $g(x) = \binom{n}{x} p^x (1-p)^{n-x}$, where $\binom{n}{x} = \frac{n(n-1)\cdots(n-x+1)}{x!}$. When $p < 0.5$, we observe that many of the features of the binomial distribution still hold. In particular,

$$\sum_{x=0}^{\infty} g(x) = 1, \text{ which follows from the binomial series,}$$

$$\left(1 + \frac{p}{1-p}\right)^n = \sum_{x=0}^{\infty} \binom{n}{x} \left(\frac{p}{1-p}\right)^x,$$

$$\sum_{x=0}^{\infty} x g(x) = np,$$

$$\sum_{x=0}^{\infty} (x - np)^2 g(x) = np(1-p),$$

and

$$\eta(t) = \sum_{x=0}^{\infty} t^x g(x) = (1 - p + pt)^n, \quad \text{for } |t| < 0.5/p.$$

The function $g(x)$ is not a probability density function (*pdf*) because $\binom{n}{x}$ alternates between negative and positive values for $x > [n] + 1$. With this in mind, we define the *pdf* of the *almost-binomial*(n, p) distribution to be

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, 2, \ldots, [n], \\ 1 - \sum_{x=0}^{[n]} \binom{n}{x} p^x (1-p)^{n-x}, & x = [n] + 1, \end{cases}$$

where $[n]$ denotes the greatest integer less than or equal to n.

We observe that when $p < 0.5$, the sum of the terms beyond $x = [n]$ in each series above tends to be extremely small. In fact, $|\sum_{x=[n]+1}^{\infty} g(x)| < |g([n]+1)| < p^{[n]}$. Consequently, if $p < 0.5$ and X has an *almost-binomial*(n, p) distribution, $E(X) \approx np$ and $\mathrm{Var}(X) \approx np(1-p)$, and the probability generating function (*pgf*) for X, $\eta(t)$, satisfies $\eta(t) \approx (1 - p + pt)^n$, $|t| < 0.5/p$.

Example 1. For $n = 5.5$, $p = 0.4$, we get the following *pdf* of the almost-binomial(5.5, 0.4) distribution.

x	0	1	2	3	4	5	6
$f(x)$	0.06023	0.22085	0.33128	0.25766	0.10736	0.02147	0.00115

Note that $np = 2.2$ and $E(X) = 2.2008$, and $np(1-p) = 1.32$ and $\text{Var}(X) = 1.3205$.

Almost-binomial(n, p) distributions form a rich two-parameter family of discrete distributions where the mean exceeds the variance, and they are useful in a variety of modeling situations. Next, we show how they can be used to model sums of independent Bernoulli random variables. Suppose that $X_1, X_2, \ldots, X_k$, are independent Bernoulli(p_i) random variables (referred to as *Poisson trials*). Let $X = \sum_{i=1}^k X_i$ (the distribution of X is called the *Poisson-binomial* distribution). We can approximate probabilities concerning X as follows. We set $np = \mu = \sum_{i=1}^k p_i$, $np(1-p) = \sigma^2 = \sum_{i=1}^k p_i(1-p_i)$, and get

$$n = \frac{\left(\sum_{i=1}^k p_i\right)^2}{\sum_{i=1}^k p_i^2} \quad \text{and} \quad p = \frac{\sum_{i=1}^k p_i^2}{\sum_{i=1}^k p_i}.$$

Then we use the almost-binomial(n, p) distribution to approximate probabilities involving X.

Example 2. A machine has 20 components that operate independently. Suppose the probability that the ith component fails is $p_i = \frac{i}{100}$. Let X be the total number of components that fail. For beginning students, the exact distribution of X is difficult to deal with, especially if they only have a hand calculator. However, almost-binomial approximations are straightforward. Calculating n and p as above, we get $p = 0.13667$ and $n = 15.366$. The next table gives the almost-binomial approximations to $P(X = x)$, denoted $AB(x)$, and (for comparison purposes) the true values for $P(X = x)$.

x	0	1	2	3	4	5	6	7	8
$AB(x)$	0.10455	0.25431	0.28917	0.20395	0.09981	0.03592	0.00982	0.00208	0.00034
$P(X = x)$	0.10432	0.25452	0.28946	0.20385	0.09961	0.03585	0.00985	0.00211	0.00036

A theoretical justification for almost-binomial approximations is given in [**4**]. (See also [**1**, pp. 188–191], for a similar theoretical discussion of binomial approximations.) There we can find the following result as well as a comparison of almost-binomial, binomial, and Poisson approximations.

Theorem. Suppose $X = \sum_{i=1}^k X_i$, where $X_1, X_2, \ldots, X_k$, are independent Bernoulli random variables with parameters p_i. Let $p = \frac{\sum_{i=1}^k p_i^2}{\sum_{i=1}^k p_i}$, $n = \frac{(\sum_{i=1}^k p_i)^2}{\sum_{i=1}^k p_i^2}$, and let $\tilde{P}$ be a random variable for which $P(\tilde{P} = p_i) = \frac{p_i}{\sum_{i=1}^k p_i}$. Then for $A \subset \{0, 1, \ldots, [n]\}$,

$$|P(X \in A) - P(AB(n, p) \in A)| \le \frac{4}{1-p}\text{Var}(\tilde{P}) + \frac{k - [n] - 1}{([n]+1)(1-p)} P(X \ge [n] + 2).$$

I have directed two undergraduate research projects related to the almost-binomial distribution. The first intern (320 hours), Abu Jalal, developed error bounds for almost-binomial approximations to probabilities involving sums of independent hypergeometric random variables [**3**]. The second intern (160 hours), Don Claycomb, is currently finishing work on practical (non-theoretical) improvements to the error bound given in the theorem. This theorem is rather conservative and one can generally get much smaller bounds on the errors in approximation that hold for intuitively worst-case situations, thus arguably everywhere.

There are other opportunities for undergraduate research that faculty members could craft. One interesting problem would be to work out the details of testing a null hypothesis that data is from a Poisson distribution by comparing the Poisson fit to the data with that of the best fitting almost-binomial distribution. If one suspected the data would be under-dispersed relative to the Poisson distribution (before collecting data), a one-sided test of this sort may be an interesting competitor to the Poisson dispersion test.

Other possible problems would include looking at some of the material in [**2**] that used the Poisson-binomial distribution, namely logistic regression and conditional Bernoulli models. The almost-binomial approximation to the Poisson-binomial may be of interest here.

References

1. A. D. Barbour, L. Holst, and S. Janson, *Poisson Approximation*, Oxford University Press, 1992.
2. S. Chen, and J. Liu, Statistical applications of the Poisson-binomial and conditional Bernoulli distributions, *Statistica Sinica* **7** (1997) 875–892.
3. A. Jalal, Error bounds involving almost-binomial approximations of hypergeometric probabilities, *Pi Mu Epsilon Journal*, to appear.
4. P. Thompson, Almost-binomial random variables, technical report, Wabash College, 1999.

The Roots of a Quadratic

Leonard Gillman (len@math.utexas.edu), The University of Texas, Austin, TX 78712

In a recent discussion among several math teachers, both college and precollege, someone remarked that we do students a disservice when we let them "solve" a quadratic equation by means of the formula without having them check their answers. The obvious question at this point is just how the students are expected to do the checking. Most of the group agreed that substituting into the quadratic is too hard, at least for beginning students, especially when complex numbers are involved.

Regarding this last assertion, observe that there is no need to use the i-notation in order to do the checking. Consider the equation $f(x) = 0$, where

$$\begin{aligned} f(x) &= ax^2 + bx + c \quad (a \neq 0) \\ &= 2x^2 - 5x + 7. \end{aligned} \tag{1}$$

I include enough detail to illustrate the mechanism. Introduce the symbol

$$u = \sqrt{b^2 - 4ac} = \sqrt{-31}, \text{ and note that } u^2 = -31.$$

The solutions given by the quadratic formula are $\frac{5\pm u}{4}$. For the solution $s = \frac{5+u}{4}$, say, we get

$$f(s) = 2\left(\frac{5+u}{4}\right)^2 - 5\left(\frac{5+u}{4}\right) + 7$$

$$= 2\left(\frac{25 + 10u + u^2}{16}\right) - \frac{25}{4} - \frac{5u}{4} + 7$$
$$= \frac{1}{8}\left[(25 + 10u - 31) - 50 - 10u + 56\right]$$
$$= \frac{1}{8}\left[(10u - 6) - 10u + 6\right] = 0,$$

verifying that s is a root.

But we can check for errors without substituting into the equation. A well-known (and easily proved) theorem tells us that the sum of the roots of (1) is $-b/a$ and the product is c/a. So test the sum and product; if either fails the test then you know your solution-pair is wrong. In the example, call the two solutions s_1 and s_2:

$$s_1 = \frac{5 + u}{4}, \quad s_2 = \frac{5 - u}{4}. \tag{2}$$

Then

$$s_1 + s_2 = \frac{5 + u}{4} + \frac{5 - u}{4} = \frac{5}{2} = -\frac{b}{a},$$

and

$$s_1 s_2 = \left(\frac{5 + u}{4}\right)\left(\frac{5 - u}{4}\right)\frac{25 - u^2}{16} = \frac{25 - 31}{16} = \frac{7}{2} = \frac{c}{a},$$

so we have failed to prove we made a mistake.

Moreover, these test results show that s_1 and s_2 are in fact the actual roots—that is, the converse of the theorem quoted above is a theorem. To see this, let f be as in (1) and let r_1 and r_2 be numbers satisfying

$$r_1 + r_2 = -\frac{b}{a}, \quad r_1 r_2 = \frac{c}{a}. \tag{3}$$

Then

$$ax^2 + bx + c = a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right)$$
$$= a\left[x^2 - (r_1 + r_2)x + r_1 r_2\right] = a(x - r_1)(x - r_2),$$

whose roots are precisely r_1 and r_2.

I remark that the same idea works for polynomials of arbitrary degree, though admittedly the practical usefulness of this fact appears doubtful. For instance, if r_1, r_2, r_3 are the roots of

$$g(x) = ax^3 + bx^2 + cx + d \quad (a \neq 0), \tag{4}$$

then the general theorem on sums and products of roots tells us that

$$r_1 + r_2 + r_3 = -\frac{b}{a}, \quad r_1 r_2 + r_2 r_3 + r_3 r_1 = \frac{c}{a}, \quad r_1 r_2 r_3 = -\frac{d}{a}, \tag{5}$$

and an argument analogous to the one given for the quadratic case shows that any numbers r_1, r_2, r_3 satisfying (5) are necessarily the three roots of the cubic polynomial (4).

Fermat's Little Theorem From the Multinomial Theorem

Thomas J. Osler (osler@rowan.edu), Rowan University, Glassboro, NJ 08028

Fermat's Little Theorem [**1**] states that $n^{p-1} - 1$ is divisible by p whenever p is prime and n is an integer not divisible by p. This theorem is used in many of the simpler tests for primality. The so-called multinomial theorem (described in [**2**]) gives the expansion of a multinomial to an integer power $p > 0$,

$$(a_1 + a_2 + \cdots + a_n)^p = \sum_{k_1+k_2+\cdots+k_n=p} \binom{p}{k_1, k_2, \ldots, k_n} a_1^{k_1} a_2^{k_2} \cdots a_n^{k_n}. \quad (1)$$

Here the multinomial coefficient is calculated by

$$\binom{p}{k_1, k_2, \ldots, k_n} = \frac{p!}{k_1!\, k_2! \cdots k_n!}. \quad (2)$$

This is a generalization of the familiar binomial theorem to the case where the sum of n terms $(a_1 + a_2 + \cdots + a_n)$ is raised to the power p. In (1), the sum is taken over all nonnegative integers $k_1, k_2, \ldots, k_n$ such that $k_1 + k_2 + \cdots + k_n = p$.

In this capsule, we show that Fermat's Little Theorem can be derived easily from the multinomial theorem. The following steps provide the derivation.

1. All the multinomial coefficients (2) are positive integers. This is clear from the way in which they arise by repeated multiplication by $(a_1 + a_2 + \cdots + a_n)$ in (1).
2. There are n values of the multinomial coefficient that equal 1. These occur when all but one of the indices $k_r = 0$, so that the remaining index equals p. For example, $\binom{p}{0,\ldots,0,p,0,\ldots,0} = \frac{p!}{0!\cdots 0!\, p!\, 0! \cdots 0!} = 1$.
3. With the exception of the n coefficients just listed above, all of the remaining coefficients are divisible by p if p is a prime number. This follows from the fact that (2) is an integer, so the denominator $k_1!\, k_2! \cdots k_n!$ divides the numerator $p!$. Since $k_r < p$ for $r = 1, 2, \ldots, n$, the factor p never occurs in the prime factorization of the denominator $k_1!\, k_2! \cdots k_n!$. Therefore, $k_1!\, k_2! \cdots k_n!$ must divide $(p-1)!$ and so p divides the multinomial coefficient.
4. Let each of the $a_r = 1$ for $r = 1, 2, \ldots, n$ in (1). Then from step 2 above,

$$(1 + 1 + \cdots + 1)^p = 1^p + 1^p + \cdots + 1^p + \sum \binom{p}{k_1, \cdots k_n}. \quad (3)$$

Note, from step 3, that all the multinomial coefficients in the sum are divisible by p. And since $1 + 1 + \cdots + 1 = n$ in (3), we get

$$n^p = n + \{\text{number divisible by } p\}.$$

It follows that $n^p - n = n(n^{p-1} - 1)$ is divisible by p. Finally, $n^{p-1} - 1$ is divisible by p if n is not divisible by p.

The author wishes to thank James Smoak for correspondence which motivated this capsule.

References

1. David M. Burton, *Elementary Theory of Numbers*, (4th ed.), McGraw-Hill, 1997, pp. 91–92.
2. R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, 1989, pp. 166–172.