for arbitrary a and b. A special case of Theorem 1 in S. J. Yakowitz's *Computational Probability and Simulation* [Addison-Wesley, 1977] shows that all 16 different numbers will be generated if and only if b is relatively prime to 16, and $a-1$ is a multiple of 4. With some guidance, students can discover this, as well as the fact that one number is repeated indefinitely in the sequence when $a-1$ is odd.

These student explorations can be used as the basis for discussing the "suitability" of pseudorandom number generators in real computers. In order for

$$L_i = aL_{i-1} + b \pmod{K}$$

to be "suitable," choose a and b so that K different numbers are generated. We would want the numbers in the sequence to be "uniformly distributed." For example, the number of the first n elements in the sequence less than or equal to n should approximate n/K. Moreover, the numbers in the sequence should be "independent." For example, the number of "runs up" should be comparable to a theoretical distribution. Students who explore pseudorandom number generators modulo 16 should be able to more easily understand the related ideas in later courses.

———o———

Conditional Expectations and the Correlation Function

Barthel W. Huff, Randolph-Macon College, Ashland, VA

If the random variables X and Y have joint density

$$f_{X,Y}(x,y) = x + y \quad \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1,$$

then it is easily computed that

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\cdot\operatorname{Var}(Y)}} = \frac{E(XY) - E(X)\cdot E(Y)}{\sqrt{\operatorname{Var}(X)\cdot\operatorname{Var}(Y)}} = -\frac{1}{11}.$$

This appears as Example 22 on p. 156 of Mood, Graybill, and Boes' *Introduction to the Theory of Statistics*, 3rd edition, McGraw-Hill, 1974, where the authors conclude by asking if a negative correlation seems "right." In Example 24 on p. 158 of that volume, it is calculated that

$$E(Y|X=x) = \frac{3x+2}{6x+3} \quad \text{for } 0 \le x \le 1,$$

but there is no indication that this might be related to the earlier question. Since

$$\frac{d}{dx}E(Y|X=x) = \frac{d}{dx}\left(\frac{3x+2}{6x+3}\right) = \frac{-3}{(6x+3)^2} < 0,$$

the conditional expectation is a decreasing function of x, and the observer might conclude that Y is expected to give smaller values as the value of X increases; that is, conclude that X and Y are tending in opposite directions and therefore should have a negative correlation.

In view of the above, we prove the following.

Theorem. *Let X and Y be random variables such that*

(i) $E(Y|X=x)$ *is a non-increasing function of* x, *and*

(ii) $E(Y|X=x_1) > E(Y|X=x_2)$ *for some* $x_1 < x_2$ *such that* $P[X \leq x_1] \cdot P[X \geq x_2] > 0$.

If $\rho(X,Y)$ *exists, then* $\rho(X,Y) < 0$.

Proof. Since the correlation coefficient is invariant under translation of the random variables involved, we can assume that $E(X) = 0 = E(Y)$ and need only to show that $E(XY) < 0$. Since the result holds for general distributions, we shall use the Lebesgue-Stieltges integral. (Restricted results could be obtained with discrete or continuous probability density functions.)

Condition (ii) guarantees that either

$$x_1 < 0 \text{ and } E(Y|X=x_1) > E(Y|X=0),$$

or

$$x_2 > 0 \text{ and } E(Y|X=0) > E(Y|X=x_2).$$

If the first case holds, then condition (i) implies:

$$xE(Y|X=x) \leq xE(Y|X=x_1) < xE(Y|X=0) \quad \text{for } x \leq x_1 < 0$$
$$xE(Y|X=x) \leq xE(Y|X=0) \quad \text{for } x > x_1.$$

Thus, since $X \leq x_1$ with positive probability,

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty} xE(Y|X=x)\,dF_X(x) \\
&= \int_{-\infty}^{x_1} xE(Y|X=x)\,dF_X(x) + \int_{x_1}^{\infty} xE(Y|X=x)\,dF_X(x) \\
&< \int_{-\infty}^{x_1} xE(Y|X=0)\,dF_X(x) + \int_{x_1}^{\infty} xE(Y|X=x)\,dF_X(x) \\
&\leq \int_{-\infty}^{x_1} xE(Y|X=0)\,dF_X(x) + \int_{x_1}^{\infty} xE(Y|X=0)\,dF_X(x) \\
&= E(Y|X=0)\int_{-\infty}^{\infty} x\,dF_X(x) \\
&= E(Y|X=0)\cdot E(X) \\
&= 0.
\end{aligned}$$

A similar argument works if the second case holds.

Remark 1. Condition (ii) of the theorem cannot be relaxed without allowing the possibility that $E(Y|X=x)$ is constant on the support of X. But then X and Y would be uncorrelated with $\rho(X,Y) = 0$.

Remark 2. It is well known that if $E(Y|X=x)$ is a linear function of x, then the least-squares regression line is given by that linear function. Thus, the derivative of the conditional expectation, slope of the regression line, and correlation function would have the same sign.

Finally, suppose X and X^* are independent and identically distributed random variables having mean 0 and variance 1, and let $Y = X^* - \varepsilon X$. Then

$$\rho(X, Y) = \frac{-\varepsilon}{\sqrt{1+\varepsilon^2}}$$

and

$$E(Y|X) = E(X^*|X) - \varepsilon E(X|X) = -\varepsilon X.$$

Thus, it is possible for $E(Y|X = x)$ to be a strictly decreasing function of x and yet have $\rho(X, Y)$ arbitrarily close to 0.

Another Proof of Jensen's Inequality

Norman Schaumberger and Bert Kabak, Bronx Community College, Bronx, NY

In this capsule we use the derivative to prove that if $f(x)$ is concave down on the interval $a < x < b$ and if $a < x_i < b$, $i = 1, 2, \dots, n$, then

$$\frac{1}{n}\sum_1^n f(x_i) \le f\left(\frac{\sum_1^n x_i}{n}\right). \tag{1}$$

Equality holds if and only if $x_1 = x_2 = \cdots = x_n$. This is known as Jensen's inequality and the usual proof uses an adaptation of Cauchy's proof of the arithmetic-geometric mean inequality. (See, for example, I. Niven, *Maxima and Minima Without Calculus*, MAA, 1981, p. 87.) Our argument rests on the following proposition.

If $a < x < b$ and $A = (1/n)\sum_1^n x_i$ where $a < x_i < b$, then

$$f(x) - xf'(A) \le f(A) - Af'(A), \tag{2}$$

with equality if and only if $x = A$.

Since $f(x)$ is concave down on (a, b), its second derivative is negative in this interval. So, (2) follows from the observation that $g(x) = f(x) - xf'(A)$ takes its maximum in (a, b) at $x = A$, because its derivative $g'(x) = f'(x) - f'(A)$ is monotone decreasing on this interval and thus vanishes if and only if $x = A$. Applying (2) to each x_i and adding, we obtain

$$\sum_1^n f(x_i) - f'(A)\sum_1^n x_i \le nf(A) - nAf'(A)$$

or

$$\sum_1^n f(x_i) \le nf\left(\frac{\sum_1^n x_i}{n}\right).$$

Furthermore, there is equality if and only if each x_i in (2) equals A; that is, if and only if $x_1 = x_2 = \cdots = x_n$.