An Inequality related to the Birthday Problem

Terry R. McConnell
Syracuse University

October 5, 2001

Result

In this paper we prove the theorem below and discuss an application to the Birthday Problem.

Theorem: Let $N \geq 1$ and $x_1, x_2, \ldots, x_N$ be a sequence of nonnegative real numbers. Let $k \geq 1$ be an integer. Then

$$\frac{1}{\binom{N}{k}} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq N} x_{i_1} x_{i_2} \ldots x_{i_k} \quad \leq \quad (\bar{x})^k, \tag{1}$$

where $\bar{x}$ denotes the arithmetic average of the x_j. Equality holds if $k = 1$, and if and only if all the x_j are equal when $k > 1$.

Proof: We may assume $k > 1$. Let $f(x_1, \ldots, x_N)$ be the function defined by the left hand side of (1). Then f is continuous. Thus, it attains its maximum on the simplex S defined by

$$\bar{x} \quad = \quad 1, \ x_j \geq 0, \ j = 1, 2, \ldots, N, \tag{2}$$

at some point $y = (y_1, \ldots, y_N)$. It suffices to prove that

$$y_j \quad = \quad 1, \quad j \ = \ 1, \ldots, N. \tag{3}$$

We shall prove this by contradiction.

Suppose for some fixed pair of indices $i \neq j, 1 \leq i, j \leq N$, we have

$$y_i \neq y_j. \tag{4}$$

Consider the point z with the same components as y except for its i-th and j-th components. The i-th and j-th components of z are both equal to the average $\frac{y_i + y_j}{2}$.

Let A be $\sum y_{i_1} \ldots y_{i_k}$, where the sum extends over k-tuple indices satisfying $1 \leq i_1 < i_2 < \cdots < i_k \leq N$, with $i, j \notin \{i_1, \ldots, i_k\}$. Let B and C be defined

similarly, but with k-1-tuples and k-2-tuples respectively. (C is an empty sum if $k = 2$. We interpret it then as $C = 1$.) Then

$$f(z) \;=\; A \;+\; B(z_i + z_j) \;+\; Cz_iz_j \;=\; A \;+\; B(y_i + y_j) \;+\; C(\frac{y_i + y_j}{2})^2.$$

The last expression is strictly greater than $A + B(y_i + y_j) + Cy_iy_j$, by the Arithmetic-Geometric mean inequality.

Thus, $f(z) > f(y)$, contradicting the fact that the maximum of f on S occurs at y. This contradiction shows that $y_i = y_j$ for all pairs of distinct indices i and j, i.e., the y_j are constant. Thus (3) holds, since y satisfies (2).

Finally, the same averaging argument shows that strict inequality holds whenever the y_j are not constant.

Discussion

Inequality (1) has an amusing application to the Birthday Problem. As it is commonly stated in probability textbooks, the Birthday Problem asks for the probability that two or more people in a group of k people share a birthday. Equivalently, what is the complementary probability, $P(k)$, that all k people have distinct birthdays?

One assumes that there are exactly $N = 365$ possible birthdays, i.e., leap years are ignored. Moreover, if Y_i denotes the i-th person's birthday, then the Y_i are assumed to be independent and uniformly distributed on $\{1, 2, \ldots, N\}$. Under these assumptions one can show that $P(23) < \frac{1}{2}$. Thus, there are better than even odds that some pair of people in a group of 23 or more share a birthday.

It follows from the theorem, however, that uniform distribution is unnecessary for this conclusion. Suppose, more generally, that N is not necessarily 365, and that $Y_1, Y_2, \ldots, Y_k$ are independent, indentically distributed, but not necessarily uniformly distributed. Let $x_j = P(Y_1 = j)$. Then by identical distribution and independence,

$$P(k) = k! \sum_{1 \le i_1 < \cdots < i_k \le N} P(Y_1 = i_1, \ldots, Y_k = i_k) = k! \sum_{1 \le i_1 < \cdots < i_k \le N} x_{i_1} x_{i_2} \ldots x_{i_k}.$$

By (1), the latter is bounded above by

$$k! \binom{N}{k} (\bar{x})^k \;=\; \frac{N!}{(N-k)!N^k}.$$

This is the familiar expression from the uniform case [1]. Thus, the probability of having distinct "birthdays" is maximized if birthdays are uniformly distributed. See links at [2] for more on the Birthday Problem and its variants.

Case $k = 2$ of (1) can be deduced easily from the Cauchy-Schwartz inequality in the form

$$\left(\sum_{j=1}^{N} x_j \right)^2 \;\le\; N \sum_{j=1}^{N} x_j^2. \tag{5}$$

Indeed, case $k = 2$ of (1) is equivalent to (5). Thus (1) can also be viewed as a generalization of (5).

There is no reverse inequality. The right side of (1) cannot be bounded by any multiple of the left side when $k \geq 2$. Take $x_1 = 1$, and the remaining x_j very small. There also does not seem to be a noncommutative extension of the inequality. An appropriate conjecture would be

$$\frac{(N-k)!}{N!} \sum_{1 \leq i_1 \neq i_2 \neq \cdots \neq i_k \leq N} x_{i_1} x_{i_2} \ldots x_{i_k} \quad \leq \quad \left(\frac{1}{N} \sum_{j=1}^{N} x_j \right)^k ,$$

where $x_1, \ldots, x_N$ are, e.g, nonnegative definite $m \times m$ matrices, and $B \leq A$ for $m \times m$ matrices means that $A - B$ is nonnegative definite. It is not difficult, however, to find counterexamples with 2×2 matrices and $N = k = 3$. See [3].

Finally, we remark that the following probabilistic consequence holds:

Corollary: Let $x_1, x_2, \ldots, x_N$ be independent, mean zero random variables with finite variances $\sigma_1^2, \ldots, \sigma_N^2$. Then

$$\frac{1}{\binom{N}{k}} E \left(\sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq N} x_{i_1} x_{i_2} \ldots x_{i_k} \right)^2 \quad \leq \quad \left(\frac{1}{N} \sum_{j=1}^{N} \sigma_j^2 \right)^k .$$

If $k > 1$, equality holds if and only if all variances are equal.

This follows from (1) using orthogonality of the terms on the left hand side.

Acknowlegement

The Author thanks Vincent Fatica for asking whether the uniform distribution is extremal in the Birthday Problem, and Philip Griffin and Tadeusz Iwaniec for some very interesting conversations.

References

1. William Feller, *An Introduction to Probability Theory and Its Applications*, Vol I, 3rd Edition, Wiley, New York, 1968, p. 33.

2. http://www.mathsoft.com/mathcad/library/puzzle/soln28/soln28.html

3. http://barnyard.syr.edu/mat_ineq.ma (maple commands.)