The Friendship Paradox and A Friendship Model

Sheldon M. Ross
Department of Industrial and Systems Engineering
University of Southern California

Friendship Paradox

- n people
- $x_{i,j} = x_{j,i}$ is indicator of whether i and j are friends, $i \neq j$
- $f(i) = \sum\limits_{j \neq i} x_{ij}$
- $f = \sum_i f(i)$
- $P(X = i) = 1/n, \qquad E[f(X)] = f/n$
- each of the n writes, on separate sheets of paper, the names of each of their friends. (total of f sheets)
- Y is name on randomly chosen sheet

Friendship Paradox: (Feld, S. L., Why Your Friends Have More Friends Than You Do, *American Journal of Sociology*, 1991)

$$E[f(Y)] \geq E[f(X)]$$

Proof: $P(Y = i) = f(i)/f$

$$\begin{aligned} E[f(Y)] &= \sum f^2(i)/f \\ &= \frac{E[f^2(X)]}{E[f(X)]} \\ &\geq E[f(X)] \qquad \blacksquare \end{aligned}$$

Stochastic Orders

Say that $X \geq_{st} Y$ if for all c

$$P(X > c) \geq P(Y > c)$$

$$X \geq_{st} Y \Leftrightarrow E[h(X)] \geq E[h(Y)] \text{ for all } h \uparrow$$

Suppose X and Y have densities (or mass functions) f and g. Say that $X \geq_{lr} Y$ if

$$\frac{f(x)}{g(x)} \uparrow x$$

$X \geq_{lr} Y$ is equivalent to

$$X|a < X < b \ \geq_{st} \ Y|a < Y < b \text{ for all } a, b$$

Proposition 1:

$$f(Y) \geq_{lr} f(X)$$

Proof: n_j = number of people with j friends

$$P(f(Y) = j) = \frac{jn_j}{f}, \quad P(f(X) = j) = \frac{n_j}{n} \qquad \blacksquare$$

Let Z be a randomly chosen friend of X.

Proposition 2: $f(Z) \geq_{st} f(X)$.

Remarks

- Not necessarily true that $f(Z) \geq_{lr} f(X)$.
- If each pair is independently friendly with probability p, then

$$f(Z)|f(X) > 0 \geq_{lr} f(X)|f(X) > 0$$

- No inequality between $E[f(Y)]$ and $E[f(Z)]$.

- Let $R(i)$ denote a randomly chosen friend of i. Then

$$\begin{aligned} R(Y) &=_{st} Y \\ R(Z) &\neq_{st} Z \end{aligned}$$

Cao, Y, and S. M. Ross, **The Friendship Paradox,** *The Mathematical Scientist*, 41, 1, 2016

Friendship Model: joint with Rebecca Dizon-Ross

- $\mathbf{X} = (X_1, \ldots, X_n)$ iid distribution G.
- $X_{i,j} = X_{j,i}$ is indicator of whether i and j are friends
- $X_{i,j}, i < j$, conditionally independent given $\mathbf{X}$, and

$$P(X_{i,j} = 1 | \mathbf{X} = (x_1, \ldots, x_n)) = p(x_i, x_j), i < j$$

$p(x, y) = p(y, x)$ called *friendship function*.

- When $P(0 \leq X_i \leq 1) = 1$.

  $p(x, y) = xy$

  $p(x, y) = x + y - xy$
- finite case: X_i gives type of person i.
  $p(r, t)$ is prob type r and type t are friends.

Some Preliminaries

Let X, Y be iid G. Let

$$p(x) = E[p(x, Y)]$$

Also, let

$$p = P(X_{i,j} = 1) = E[p(X, Y)] = E[p(X)]$$

Is

$$P(X_{i,j} = 1 | X_{i,k} = 1) \geq P(X_{i,j} = 1) \ \ ??$$

$$\begin{aligned} P(X_{i,j} = 1, X_{i,k} = 1) &= E[P(X_{i,j} = 1, X_{i,k} = 1 | X_i)] \\ &= E[P(X_{i,j} = 1 | X_i) P(X_{i,k} = 1 | X_i)] \\ &= E[p^2(X)] \\ &\geq E^2[p(X)] = p^2 \end{aligned}$$

Is

$$P(X_{i,j} = 1 | X_{i,k} = 1, X_{j,k} = 1) \geq P(X_{i,j} = 1) \, ?$$

Associated Random Variables

Definition Random variables $W_1, \ldots, W_m$ are said to be *associated* if for all increasing functions h and t

$$\mathrm{Cov}(h(W_1, \ldots, W_m),\ t(W_1, \ldots, W_m)) \geq 0.$$

Some results:

- Independent random variables are associated.
- Increasing functions of associated random variables are associated.

Proposition: If $p(x, y)$ is increasing function, then $X_{i,j}, i \neq j$ are associated.

Proof: Let $U_{i,j}, i < j$, be iid $U(0, 1)$, independent of $X_1, \ldots, X_n$. For $i < j$, define

$$X_{i,j} = X_{j,i} = \begin{cases} 1, & \text{if } U_{i,j} > 1 - p(X_i, X_j) \\ 0, & \text{if otherwise,} \end{cases}$$

Because $X_{i,j}$ are all increasing functions of the independent random variables $X_1, \ldots, X_n$, $U_{i,j}$, $i < j$, they are associated. ■

Corollary: If $p(x, y)$ is an increasing function, then

$$P(X_{i,j} = 1 | X_{i,k} = 1, X_{j,k} = 1) \geq p$$

Proof: Need show

$$P(X_{i,j} = 1, X_{i,k} = 1, X_{j,k} = 1) \geq P(X_{i,j} = 1)P(X_{i,k} = 1, X_{j,k} = 1)$$

To show this, let

$$h = I\{X_{i,j} = 1\}, \quad t = I\{X_{i,k} = 1, X_{j,k} = 1)\}.$$

Remark: Even when $p(x, y)$ is increasing, it is not necessarily true that

$$P(X_{i,j} = 1 | X_{i,k} = 1, X_{j,k} = 1) \geq P(X_{i,j} = 1 | X_{i,k} = 1)$$

Let $p(x, y) = x + y - xy$, and

$$P(X_i = 1) = \epsilon = 1 - P(X_i = 0)$$

Then

$$\begin{aligned} P(X_{1,2} = 1 | X_{1,3} = 1) &\approx 1/2 \\ P(X_{1,2} = 1 | X_{1,3} = 1, X_{2,3} = 1) &\approx 0 \end{aligned}$$

Number of Friends

For any specified set of people S, let

$$F_i(S) = \sum_{j \in S} X_{i,j}$$

denote the number of friends i has in S.

Theorem: For any sets S and T and any r

$$P(F_i(T) \geq r | F_i(S) = k) \uparrow k.$$

Inspection Paradox

Population is $0, 1, \ldots, n$. By symmetry can let 0 be randomly chosen person.

$$E[F_0] = np$$

To determine $E[F_R]$, where R is a randomly chosen friend of 0

- Condition on X_0. Say $X_0 = x$.

- Let Y be number of friends of 0.

- Let

$$g_1(y) = \frac{g(y)p(x,y)}{p(x)}$$

$$g_2(y) = \frac{g(y)q(x,y)}{q(x)}$$

- Conditional on $X_0 = x, Y,$ the values X_i are independent,

  X_R and $Y - 1$ others have density g_1

  $n - Y$ have density g_2

- Compute $E[F_R | X_0 = x]$