Human Behavior and Stochastic Models

Gustav N. Kristensen
University of Southern Denmark
Centre of Health Economics Research (COHERE).
Campusvej 55, DK-5230, Odense M, Denmark

Abstract

In a random system "groups" will form accidentally, and their distribution functions can be described as mathematical idealizations. Nature, including social relations, deviates from the idealized model. This article thus discusses empirical group formation as a deviation from random group formation. The article can thereby in principle reveal "false" grouping. The applied case is group formation in a canteen. The group formation is investigated based on nine sets of observations of students dining in the canteen, in total 526 individuals. The article combines the use of a transformed geometric distribution as well as the binomial distribution as a benchmark for measuring the empirical group formation. An index for measuring the degree of gender segregation is developed.

JEL Codes: C15, C18, C51.

Keywords: Segregation, geometric distribution, gender, group formation, canteen.

Introduction

The research question of this article is: Is a given group formation random or human?

Without a solid theory, human behavior is to a large extent an empirical question. How can we know that an outcome is not just a consequence of random laws?

As an example, Horvath (1966 p 516) writes: "One would expect that human behavior has such purposeful roots that random processes would be the exception rather than the rule." Horvath supports the scientific value of such models by quoting an article from 1944 by E. Schrödinger, who writes: "The working of the statistical mechanism itself is what we are really interested in."

This article underlines the need to clarify what is random (or mathematics) and to define the deviation from that as human. The title of Horvath's article, "Stochastic models of behavior", is in this context seen as a self-contradicting title.

A critical discussion of using models for making clinical predictions was given by Kleinmuntz (1990), who like this article supports the idea that mathematics and statistics are by their clear and unambiguous definitions sharp instruments in the discussion of human behavior in an economic and sociological context, but that they are only instruments and never the real world.

This philosophy shall here be illustrated by the group formation in a canteen.

Segregation, group formation, and group dynamics are important for the life of humans and have many facets as e.g. shown by Forsyth (2010). Therefore, this study is limited to selected aspects of discussing what is random and what is human.

The specific topic is the spontaneous group formation in our daily life. Latané and Bidwell (1977: 574) thus focus on group behavior in college cafeterias. In their study they find that women have a stronger preference for eating together than men.

Like Clack et al. (2005) this article underlines the importance of investigating segregation as a micro process by observing patterns of seating in a canteen. Clack et al. focus on a multi-ethnic population and underline that the literature has neglected the micro-level processes that may sustain racial barriers 'on the ground'.

In their study Clack et al. uses the P index as a measure of inter group exposure (segregation). The index is based on conditional distributions

This article focuses on gender. The ethnic aspect could easily be included, but due to the low rate of people of other origin than Danish, ethnicity was finally dropped.

In his section: 'Men, Women, and Groups', (Forsyth 2010: 91-92) reminds us that until recently in the Western World women were legally and administratively separated from men. Consequently he suggests, "As sexist attitudes decline, differences in membership of various types of groups may also abate."

Related to this discussion are Fisher and Byrne (1975) who discusses the feelings of males and females towards invaders of their personal space when sitting in the university library.

As underlined by Latané and Bidwell (1977), the group effect is different for men and women, and thereby could indicate an aspect of more efficient group formation as well as give advice in evaluating group dynamics.

Going to a cafeteria or canteen can be seen as a data generating process. This can e.g. be regarded as a random process. Of course, when human beings are involved, such a "random" process will never exist. However, the difference between what is human and what is random should be traced as the (human) group effect. The question is treated theoretically by Diaconis (1988), and empirically by Bryman (2008).

Efficiency in integration could be measured by controlling whether a certain initiative for integration has improved an otherwise random outcome.

As women are risk averse we could ask if a higher number of women in the boards of banks would have mitigated the housing bubble. A random model including risk averse women on the boards would certainly say "Yes", but how would reality look? An interesting and highly relevant model building on optimal forecasting groups (Lamberson and Page, 2012) is based on idealized statistical models, which, however, are only sporadically held up against empirical human behavior.

The empirical paper on improved judgments by using multiple sources (Yaniv and Milyavsky, 2007) could be held up against the results from one or more random processes.

Comparing actual data with theoretical statistical distributions has likewise been an important way of revealing cheating in the data creation (Benford 1938, Angoff 1991, Jacob and Levitt 2003, Diaconis and Graham 2012, Kristensen 2012). In principle the article shows how it is possible to reveal "false" groups.

The Random Model (RAN)

This article applies a transformed geometric distribution and the binomial distribution to describe what is random. The deviation seen in the empirical data is the human effect - the group effect.

The random model is based on the following case: A number of people shall dine in a dining room of 232 seats. Let us say that on a given day 116 guests enter the dining room. For a given place (seat) to become occupied the probability (in a random system) is 116/232 = .5 = p and likewise for the neighbor. The probability of a *sequence* of k seats to be occupied follows a geometric distribution, where an occupied seat is "success".

The (random) probability that k persons will sit together in a group (sequence) is consequently described by the geometric distribution:

$$k = p^{k} (1-p) \qquad (1)$$

A sequence is a group of people separated from others by empty seats. To be a member of the group persons can sit face to face or side by side with another member of the group. The order is disregarded. Two persons sitting diagonally across from each other are not members of the same group (unless they obviously - to the observer - interact).

Equation (1) can be transformed into the probability of being a member of a group of size k.

The theoretic relationship is

$$RAN = k*p^{k}(1-p)^{2} \quad (2)$$

$$\sum_{k} RAN = p \quad (3)$$

For a given trail (distribution on a given day) we have

RAN	the random probability of being in a group of size k
H	total number of persons in the room
N	total number of seats in the room (232)
k	group size
NRAN	number of members in (all) groups of size k
H/N = p	share of occupied places, and the probability of a *given* seat to be occupied

The Actual Data (ACT)

The data are collected from the university canteen.

ACT	the actual probability of being in a group of size k Thus ACT is the empirical counterpart to RAN
NACT	actual number of members in groups of size k

In a sheet we counted how many (total on a day for each sheet) were sitting in groups of size 1, size 2, size 3, etc. and registered the result as shown in Table 1.

Table 1 reports the data from nine collected sheets as NACT.

Taking e.g. sheet no 8 where N = 232; p = H/N = 83/232 = .357.

In sheet 8 groups size 2 we have NACT = 34, and NACT = N*ACT = 34, and ACT = NACT/N = .147.

Below in Table 1 we see the aggregate distribution on women (W) and men (M).

Table 1. Actual Distribution (NACT)

	k	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
Sheet	**p**											**H**
1	**.129**	6	12	3	4	5						**30**
2	**.194**	7	10	9	8	5	6					**45**
3	**.203**	3	22	12	0	10						**47**
4	**.228**	11	30	12								**53**
5	**.245**	8	18	15	16							**57**
6	**.245**	7	32	9	4	5						**57**
7	**.254**	9	12	12	8	0	0	0	8	0	10	**59**
8	**.357**	10	34	15	4	5	6	0	0	9		**83**
9	**.418**	13	32	12	12	10	0	7	0	9		**95**
	SUM	**74**	**202**	**99**	**56**	**40**	**12**	**7**	**8**	**18**	**10**	**526**
	W	39	**128**	39	29	30	5	3	1	6	3	283
	M	35	**74**	**60**	27	10	7	4	7	12	7	243

For each counting we can now compare the theoretical random relationship with the actual.
Based on the random model we can write

$$ACT = k*p^{k}(1-p)^{2} \qquad (4)$$

$$\sum_{k} ACT = p \qquad (5)$$

This can also be seen as H0 the null-hypothesis stating that group formation among humans is random, however, to be estimated on the functional form we use:

$$ACT = a_0 k^{a_1} *p^{k}(1-p)^{2} \qquad (6)$$

where H0: $a_0 \approx 1$, $a_1 \approx 1$, as when estimated on a random dataset.

Equation (6) can be estimated in pilot studies on e.g. three different collected sheets.

$ACT_{30} = .28*k^{3.66}$ $\quad p = .129$

$ACT_{59} = .45*k^{2.55}$ $\quad p = .254$

$ACT_{83} = .48*k^{1.96}$ $\quad p = .357$

It is seen that the coefficients (numerically) clearly deviate from 1, which indicates group effect. It is also seen that the coefficients move with the probability of occupied seat "p".

The coefficients can therefore be seen as functions of p, and accordingly the entire dataset can be joined together in one equation by the expansion method (Casetti, 1972, Kristensen, 1997). Here "i" indicate sheet number and "k" group size.

A special concentration was found around group size 2, and was caught up by a dummy.

Consequently, equation (6) is transformed to

$$ACT_{ik} = a_{00}*D2*p_i^{\,2} + (a_{10} + a_{11}*p_i)*k^{(a_{20}+a_{21}*p_i)}*p_i^{\,k}(1-p_i)^{2} \qquad (7)$$

where D2 is a dummy for group size 2. D2 is equal to 1 for k = 2, otherwise 0.

Estimated by Weighted Least Squares (WLS) where weight = 1/k we get

$$ACT_{ik} = .47*D2*p_i^{\,2} + (.11 + .59*p_i)*k^{(4.88-7.78*p_i)}*p_i^{\,k}(1-p_i)^{2}$$

t (5.44) (1.78) (2.52) (11.61) (-5.26)

$R^2 = .86$ Obs. = 56 DW = 2.10

As the random distribution can be described according to equation (2), we now have the random (RAN) and the human group formation (ACT).

We can hereby calculate the group effect as

$$ACT - RAN = \text{Group effect} \qquad (8)$$

Figure 1 shows the "movement" away from a "random" distribution and show that dining alone is not popular. Figure 1 also shows how the group effect is separated out by deducting the random distribution. RAN is the calculated random distribution while ACT is calculated with the estimated model on actual data, equation (7). Figure 1 gives the net group formation effect compared to group size k. Groups of 2-3 are popular. "Groups of one" are evaded as p increases (here from p = .15 to p = .25), and group size gets bigger.

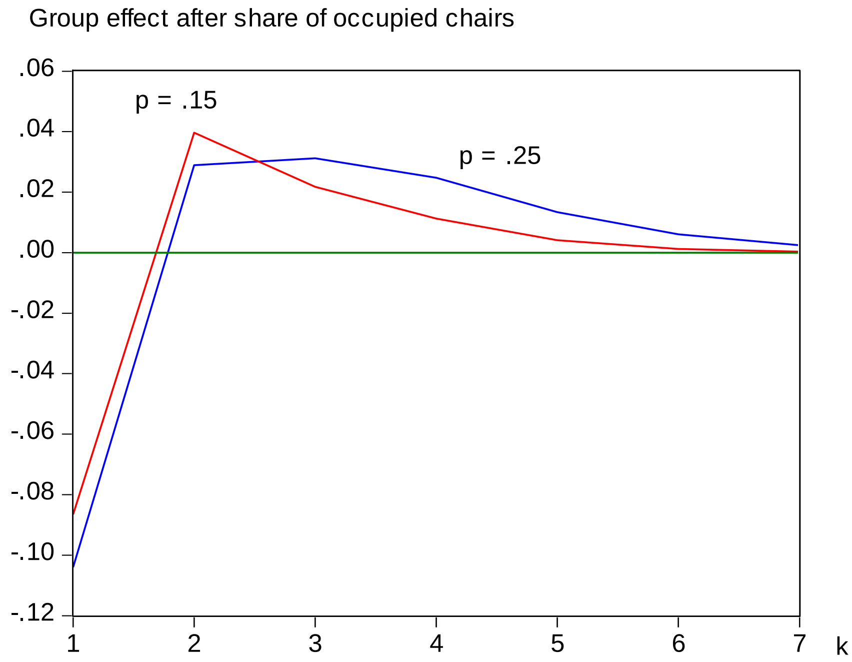

Figure 1. Net group effect, as group formation that is beyond random. Less single, more double, and more triple.

Men and Women

The group formation curves can be estimated for men and women again by WLS.

ACTW the actual probability for a woman of being in a group of size k
ACTM the actual probability for a man of being in a group of size k

For women

$$ACTW_{ik} = .42{*}D2{*}\,p_i^{\,2} + (.040 + .39{*}p_i){*}\,k^{(4.70-7.67{*}p_i)}\,{*}p_i^{\,k}\,(1\text{-}\,p_i)^2$$

t (6.31) (0.83) (2.20) (7.21) (-3.83)

$R^2 = .82$ Obs = 56 DW = 1.48

And for men

$$ACTM_{ik} = .05{*}D2{*}p_i^{\,2} + (.071 + .20{*}p_i){*}k^{(5.03-7.80{*}p_i)}\,{*}p_i^{\,k}\,(1\text{-}\,p_i)^2$$

t (1.01) (2.14) (1.64) (11.49) (-5.02)

$R^2 = .77$ Obs = 56 DW = 2.26

The dummy D2 shows how people cling together in couples, women much stronger than men, and in “spontaneous” groups where men tend to prefer bigger groups than women.

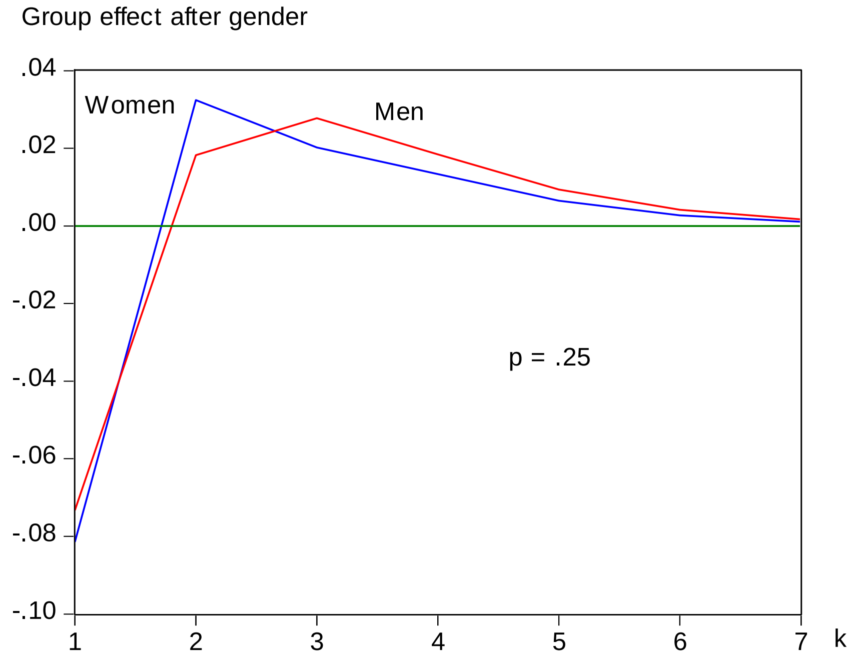

Figure 2. Net group effect for men and women, as group formation that is beyond random. Women prefer groups of 2, while men prefer groups of 3 or larger.

Figure 2 shows the distribution on groups of different size for men and women calculated for p = .25.

The theory that men dine more alone than women was not (significantly) confirmed. The most significant feature was (beyond that people do not want to dine alone) that women were significantly attracted by two person groups, while men preferred two- and three- person groups or bigger.

Gender segregation

The binomial distribution for a given group size gives the random distribution on men and women within a group.

p_m – share of people in the total population who are men

$$f(x) = \frac{(k)!}{(x)!(k-x)!} p_m^{\,x} (1 - p_m)^{k-x} \qquad (9)$$

In this case p_m = 243/526 = .462 (from Table1.)

Thus, the binomial distribution expresses the random distribution *within* each group of men and women. Any deviation from the random distribution is a segregation effect based on gender.

The Binomial Paradigm for Gender Grouping

As the data for groups above the size of 6 are scattered, the data in this section is only presented up to group size 6. Now for each group size k we calculate the percentage distribution on men and women and compare this actual distribution (ACT) with a random distribution which follows the binomial distribution (RAN).

For k = 1, we have a random distribution when p_w = .538 and p_m =.462. The *deviation* from random is

$$\text{ACT-RAN} = \text{DIF} \qquad (10)$$

where

$$\sum_k ACT = 1 \qquad (11)$$

$$\sum_k RAN = 1 \tag{12}$$

The *segregation* for each group size k is *the numerical sum of the differences*. As both ACT and RAN (in this case of gender segregation) sum up to 1 the maximum segregation is 2. (See the details in Table 4).

It is seen from Table 2 that the actual (ACT) distribution does not follow the binomial distribution and we have a segregation effect based on gender.

Table 2. Actual (ACT) and binomial random (RAN) percentage distribution of men and women for each group size

k								
		M						**W**
1	ACT	.473						.527
	RAN	.462						.538
	DIF	**.011**						**-.011**
		MM			**MW**			**WW**
2	ACT	.343			.250			.406
	RAN	.214			.496			.289
	DIF	**.129**			**-.246**			**.117**
		MMM	**MMW**			**MWW**		**WWW**
3	ACT	.333	.364			.152		.152
	RAN	.010	.344			.400		.156
	DIF	**.323**	**.020**			**-.248**		**-.004**
		MMMM	**MMMW**		**MMWW**		**MWWW**	**WWWW**
4	ACT	.167	.250		.167		.250	.167
	RAN	.046	.212		.369		.287	.084
	DIF	**.121**	**.038**		**-.202**		**-.037**	**.083**
		MMMMM	**MMMMW**	**MMMWW**		**MMWWW**	**MWWWW**	**WWWWW**
5	ACT	.125	.125	.000		.000	.125	.625
	RAN	.022	.123	.285		.330	.194	.045
	DIF	**.103**	**.002**	**-.285**		**-.330**	**-.069**	**.580**
		MMMMMM	**MMMMMW**	**MMMMWW**	**MMMWWW**	**MMWWWW**	**MWWWWW**	**WWWWWW**
6	ACT	.000	.500	.000	.0	.500	.0	.0
	RAN	.010	.069	.198	.305	.267	.125	.024
	DIF	**-.010**	**.431**	**-.198**	**-.305**	**.233**	**-.125**	**-.024**

The maximum segregation occurs when men and women are totally separated in different groups, e.g. here shown as an example in Table 3 for k = 4.

Table 3. Example of calculating maximum segregation of group size 4

k		MMMM	MMMW	MMWW	MWWW	WWWW	Sum
4	Act	.500				.500	1.000
	Ran	.063	.250	.375	.250	.063	1.000
	Dif	.437	-.250	-.375	-.250	.437	**1.749**

Under “Act” there is no man sitting together with a woman, - the maximum segregation. Sum is the sum of the numerical values.

Index for segregation

Based on the pattern in Table 2 we can now make an index (B index, for binomial based) for segregation in the canteen according to gender, as shown in Table 4.

Table 4. An index for Gender Segregation in a Canteen

k	Max. segregation MAXS	Actual segregation ACTS	B Index: ACTS/MAXS
1			
2	1.000	0.492	.492
3	1.500	0.595	.396
4	1.749	0.481	**.275**
5	1.876	1.369	.729
6	1.937	1.326	.684
7			
∞	2.000	2.000	

MAXS (in the B Index) represents complete segregation and zero represents complete integration. Table 4 shows that two-person groups are quite gender segregated. While groups on 4 are the most gender integrated.

Discussion

Men and women segregate in a canteen as they tend to seek same sex company. It is not possible in this study to say if women from the university administration prefer the company of women or just their closest colleagues, who are women. Still, e.g. the secretarial jobs are dominated by women.

However, among the students groups of young girls are regularly seen sitting together in the canteen (see Table 4 for k = 5). Therefore, it is doubtful whether Forsyth’s prediction that woman in a society without institutional separation will mix totally with men.

The absence of a (significant) difference in single dining by men and women can be attributed to the fact that Danish women are among the most independent in the world, and thus the female labor participation rate is the highest in the world.

The study opens for some research question: Are men better at solving problems in three-person groups than women? Are the optimal group sizes two for women and three for men?
According to this study groups of four are the most gender integrated groups. This could be interpreted in the way that groups of four opens for subgroups of e.g. two men and two women.

Important is also that if we suppose that the group distribution in the canteen followed equation (2) then we could see groups of different size but there was no proven (human created) group formation.

Conclusion

This article discusses group formation as the deviation from what it would be if all distributions were random. Thus, it excludes the random effect of an empirical distribution, leaving the rest as "group effect". The models can thus reveal "false" grouping in a human behavior context.

There is a considerable group effect. People tend to cling together in groups of two, women much more than men. Otherwise men prefer bigger groups than women.

There is strong gender segregation among groups. According to the applied indexes women's preference for closed groups is above that of men. It is not likely that woman in a society without institutional separation will mix totally with men.
Group size plays a critical role in integration. Groups of four are the most gender integrated groups.

Without a solid theory, human behavior is described as the difference between actual and random behavior.

References

Angoff, W. H. (1991). "The Development of Statistical Indices for Detecting Cheaters," *Journal of the American Statistical Association*, Vol. 69, No. 345, 44-49.

Benford, F. (1938). "The law of Anomalous Numbers," *Proceedings of the American Philosophical Society*, Vol. 78, No. 4, 551-572.

Diaconis, P. and Graham, R. (2012). "*Magical Mathematics. The Mathematical Ideas that Animate Great Magic Tricks,"* Princeton University Press, Princeton and Oxford, 2012.

Jacob, B. A. and Levitt, S. D. (2003). "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics*, Vol. 118 (3), 2003, 843-877.

Bryman, A. (2008). *Social Research Methods*. (Third ed. Oxford University Press)

Casetti, E. (1972). Generating Models by the Expansion Method; Application to Geographical Research, *Geographical Analyses*, 4: 81-91.

Clack, B., Dixon J. & Tredoux C. (2005). Eating Together and Apart: Pattern of segregation in Multiethnic Cafeteria. *Journal of Community & Applied Social Psychology*, 15, 1-16.

Diaconis, P. (1988). *Group representation in probability and statistics.* Institute of Mathematical statistics. ISBN 0-940600-14-5.

Duncan, O.D. & Duncan B. (1955). A methodological analysis of segregation indices. *American Sociological Review*, 20, 210-217.

Fisher, J.D. & Byrne D. (1975). Too Close for Comfort: Sex Differences in Response to Invasions of Personal Space. *Journal of Personality and Social Psychology*. Vol. 32, No. 1, 15-21.

Forsyth, D. R. (2010). *Group Dynamics, 5th Edition*. Belmont, CA: Thomson Wadsworth.

Horvath, W. J. (1966). Stochastic Models of Behavior. *Management Science* 12.12, 513-518.

Kleinmuntz, B. (1990). Why we Still Use Our Heads Instead of Formulas: Towards an integrative Approach. *Psychological Bulletin*. 107(3), 296-310.

Kristensen, G. (1997). Women's Economic Progress and the Demand for Housing: Theory and Empirical Analyses. *Urban Studies*, 34 (3), 403-418.

Kristensen, G. (2012). Disclosure of cheating by statistical methods. *Journal of Statistical and Econometric Methods*. (Forthcoming)

Lamberson, P.J. and Page, S.E. (2012). Optimal Forecasting Groups. *Management Science*. Vol.58 (4) 805-810.

Latané, B. & Bidwell L.D. (1977). Sex and Affiliation in College Cafeterias. *Personality and Social Psychology Bulletin*, 3, 571-574.

Ransom, M. R. (2000). Sampling distribution of segregation indices. *Sociological Methods and Research*, 28, 454-475.

Yaniv I. and Milyavsky M. 2007. Using advice from multiple sources to revise and improve judgements. *Organizational Behavior and Human Decision Processes*. Vol 103,1, pp104-120.