International Game Theory Review, Vol. 7, No. 4 (2005) 1–13

UNCERTAINTY OF THE SHAPLEY VALUE

VLADISLAV KARGIN

Cornerstone Research, 599 Lexington Avenue
New York, NY 10022, USA
Skarguine@cornerstone.com
slava@bu.edu

This paper defines a measure of bargaining uncertainty that quantifies Roth's concept of strategic risk. It shows how this measure can be used for checking reliability of the Shapley value in cost allocation problems and in the theory of competitive equilibrium. Salient properties of the new measure are investigated and illustrated by examples of majority voting and market games and by a cost allocation problem from epidemiology.

Keywords: Shapley value; strategic risk; bargaining uncertainty.

Subject classification: 91A12, 91A30, 91B50.

The dread of beatings! Dread of being late!
And, greatest dread of all, the dread of games!
Sir John Betjeman, "Summoned by Bells", ch.7

1. Introduction

An important problem in epidemiology is to assign weights to the factors that contribute to an increase in the incidence rate of a disease. For example, a higher incidence rate of a heart disease may be attributable to smoking, bad eating habits, or lack of physical exercise, and epidemiologists are interested in quantifying the relative degree of importance of these factors. There are different methods to perform the attribution, all of which must deal with the fact that the contribution of a factor to the incidence rate depends on the presence of the other factors. For instance, smoking may hypothetically have a relatively small impact on health unless it is accompanied by other factors. This difficulty is similar to the problems that arise in cooperative game theory, and in a recent paper Gefeller and Land (1997) have proposed an attribution according to the Shapley value, the standard solution concept from game theory.

Specifically, they have studied the influence of smoking and three types of cholesterol level, LDL, HDL, and VLDL, on a certain heart disease, myocardial infarction. When all four factors are absent, the incidence rate of the disease is 77 percent lower than the incidence rate in the total population. The Shapley attribution for this reduction is 21 percent for smoking, and 41, 8, and 8 percent for the cholesterol levels respectively. A natural question arises about the statistical significance of these attributions. What is needed is a measure of uncertainty of the Shapley value solution. The present paper proposes such a measure by utilizing the probabilistic interpretation of the Shapley value.

The usefulness of this uncertainty measure is not restricted to practical applications similar to the epidemiological example. Quantifying uncertainty of the Shapley value is also important in economic theory. For example, following a line of research initiated by Shapley, Aumann (1975) has shown that as the number of participants in a market economy grows, the Shapley value of a game associated with this economy converges to the competitive equilibrium allocation. This convergence puts the concept of the competitive equilibrium on firm game-theoretical grounds, but only if it is presumed that the Shapley value accurately reflects the bargaining outcome in the economy. In cases when the Shapley value describes the bargaining outcome approximately, the natural question is: What if the uncertainty of the Shapley value grows with the increase in the number of participants? Wouldn't it imply that the bargaining outcome does not necessarily converge to competitive equilibrium? And if the uncertainty does decrease, then how is the rate of the decrease related to the details of the market organization? Answering these questions is likely to deepen our understanding of how a market economy achieves the competitive equilibrium.

From another perspective, the Shapley value is used not only as a device to compute a fair allocation of resources among players but also as a tool to evaluate the prospect of playing a game. A measure of bargaining uncertainty would provide an additional dimension for evaluating this prospect. Indeed, both experimental evidence and personal introspection suggest that participation in a bargaining situation implies undergoing the vagaries of the bargaining process, an experience that may be a source of pleasure or displeasure depending on the personality of the participant. A measure of bargaining uncertainty would serve as a useful tool to investigate this aspect of a game. Characterizing a game by both its value and uncertainty is similar to characterizing a weapon by its power and precision, or a financial stock by its expected return and risk. In all these situations the complexity of a real-world object cannot be compressed into just one dimension; it needs at least two for proper description.

So, how should the game uncertainty be measured? To answer this question, let us first examine the different frameworks that are used to define the Shapley value.

The shortest way to define the Shapley value is axiomatic. However, this approach is not suitable for measuring the Shapley value uncertainty. This is because the Shapley value is the unique solution satisfying all the axioms, and relaxation of any of them leads to an infinity of new solutions. In contrast, the probabilistic interpretation of the Shapley value suits the problem of defining uncertainty very well: the

Shapley value is defined as the expectation of the marginal contribution to a random coalition, and it is straightforward to define the uncertainty as the standard statistical deviation of the marginal contribution. This is the way of defining uncertainty that the present paper propounds.

The probabilistic interpretation of the Shapley value was sketched by Shapley in his seminal paper (Shapley (1953)). Later it was set on firm ground by Gul (1989), Evans (1996), and Hart and Mas-Colell (1996) in their work on non-cooperative foundations of the Shapley value, work that was inspired by Rubinstein's paper on the bargaining problem (Rubinstein (1982)). Essentially, in all these probabilistic interpretations, the randomness of the model is the randomness of bargaining order and coalition formation.

To understand how the randomness in the order of bargaining introduces uncertainty about the outcome, it is useful to recall Rubinstein's model of the bargaining game. In this game two players aim to divide a pie that has unit value. The players are impatient and make offers in turn until one of them agrees to the other player's offer. The player who makes the first offer is determined by chance. Rubinstein has shown that there is a unique subgame-perfect equilibrium in this game. The ex-ante expected payoff of the game is one half, but the first player has an advantage and gets a greater part of the pie. The uncertainty here arises because the players initially do not know who will make the first offer, and a natural measure of uncertainty is the expectation of the squared difference between actual and expected payoffs. The present paper generalizes this measure to other games, illustrates it with examples, and derives some of its properties.

The concept of uncertainty of a game outcome is not entirely new. Roth (1977)[a] has pointed out that the Shapley value is insufficient for decision-making purposes unless the players are risk-neutral to a specific kind of uncertainty in the strategic interactions, which he calls "strategic risk". He distinguishes this kind of risk from the "ordinary risk" that typically arises in lotteries, and he characterizes the class of utility functions that may depend on this risk. Roth, however, does not attempt to introduce a quantitative measure of strategic risk, and that is the main contribution of the present paper.

The introduced measure provides a yardstick for both measuring the reliability of the Shapley value and quantifying the strategic risk. The main advantage of this measure is its simplicity. Future work will perhaps provide alternative measures of uncertainty based on more accurate analysis of bargaining randomness. The purpose of this paper is simply to bring attention to the importance of measuring this randomness and to suggest a possible solution.

The rest of the paper is organized as follows. Section 2 provides notation and basic definitions. Section 3 lists examples. Section 4 shows certain properties of the uncertainty concept, and Sec. 5 concludes.

[a]See also Roth (1977b) and Roth (1988).

2. Notation and Definitions

Let A be a set of players: $A = \{1, 2, \ldots, N\}$. A game G is a map of a set of all subsets of A to real numbers: $2^A \to \mathbb{R}$. The interpretation of $G(S)$ is the payoff that the coalition of players in subset S can achieve on their own.

Let (X, Σ, μ) be a probability space and $f_i : X \to 2^{A-\{i\}}$ be a random variable that takes its values in the set of all subsets of $A - \{i\}$, and has the following distribution function:

$$\mu\{f_i(X) = S\} = \frac{|S|!(N - |S| - 1)!}{N!}. \tag{1}$$

Random variable f_i can be interpreted as the random coalition that player i joins. The particular choice of its probability distribution is motivated by both the axiomatic approach and non-cooperative bargaining games that support the Shapley value.

Define the marginal contribution of player i as a function d_iG that maps coalitions from $A - \{i\}$ to $\mathbb{R}$:

$$d_iG(S) = G(S + i) - G(S). \tag{2}$$

Then the Shapley value of game G for player i is the expectation of his marginal contribution when the coalition is chosen randomly.

$$V_i(G) = \mathrm{E}\{d_iG \circ f_i\}. \tag{3}$$

The Shapley uncertainty of game G for player i is defined as the standard statistical deviation of his marginal contribution:

$$R_i(G) = \mathrm{Var}^{1/2}\{d_iG \circ f_i\} \tag{4}$$

This definition easily generalizes to other, non-Shapley, values. Indeed, almost all reasonable values are probabilistic values (Weber (1988)), so each of them can be represented as the expectation of the player's marginal contribution with respect to a certain probability distribution imposed on coalitions. Of course, for different bargaining mechanisms the probability distributions may be different. For example, another popular value, the Banzhaf value, corresponds to the following probability distribution:

$$\mu\{f_i(X) = S\} = \frac{1}{2^{N-1}}. \tag{5}$$

For probabilistic values, it is natural to define the corresponding uncertainty as the standard deviation of the marginal contribution. Most of the properties that will be derived below for the uncertainty of the Shapley value will also hold for the uncertainty of other probabilistic values. This paper, however, sticks with the Shapley value as the most popular in practical applications.

3. Interpretation and Examples

In many non-cooperative interpretations of the Shapley value, the actual payoff of a player is a weighted average of his marginal contribution to a random coalition and his expected payoff:

$$\pi_i = \alpha(d_i G(S)) + (1-\alpha)V_i(G). \tag{6}$$

For instance, in the Rubinstein-Gul-type bargaining models the parameter α is related to player impatience. Then, the standard deviation of player payoff is proportional to the Shapley uncertainty:

$$\mathrm{Var}^{1/2}\{\pi_i\} = \alpha R_i(G). \tag{7}$$

According to the Chebyshev inequality, it follows that

$$\Pr\{|\pi_i - V_i(G)| \geq c\alpha R_i(G)\} \leq c^{-2}, \tag{8}$$

where c is an arbitrary positive constant. That is, the Shapley uncertainty helps in estimating the probability of the actual payoff deviation from the expected payoff. Moreover, if the marginal contributions are distributed according to the Gaussian law, which is likely to hold in large games at least approximately, then the following inequality holds:

$$\Pr\{|\pi_i - V_i(G)| \geq 1.96\alpha R_i(G)\} \leq 0.05. \tag{9}$$

So, the Shapley uncertainty can be interpreted as measuring the variability of the actual payoff.

Let me now present several examples.

Example 1. Additive game.

In the additive game the value of a coalition is simply the sum of the values of individual players. The Shapley value for a player coincides with his individual value and, consequently, the risk is zero: Each player can guarantee his own value.

Example 2. Uncertainty may be different for different players.

Let a three-player game be defined as follows. $G(\{1\}) = 0$, $G(\{2\}) = 3$, $G(\{3\}) = 6$, $G(\{1,2\}) = 24$, $G(\{2,3\}) = 18$, $G(\{1,3\}) = 21$, $G(\{1,2,3\}) = 60$. Then,

$$V_1 = V_2 = V_3 = 20,$$

and

$$R_1 = \sqrt{299} \approx 17.2;$$
$$R_2 = \sqrt{230} \approx 15.2;$$
$$R_3 = \sqrt{155} \approx 12.4.$$

This example illustrates that players can have the same Shapley values but different Shapley uncertainties.

Example 3. Majority game.

In the majority game the value of a coalition is N if the size of the coalition is greater than half of the number of players, and 0 otherwise. For simplicity, let the number of players N be odd. Then the Shapley value of the game for player i is

$$V_i(G) = 1,$$

and the Shapley uncertainty is

$$R_i(G) = \sqrt{N}.$$

So the uncertainty of the majority game increases with an increase in the number of players. Intuitively the marginal contribution of a player is non-zero only if his vote is pivotal in a random ordering of players, but this situation is very rare in the majority game with a large number of players.

Example 4. Large production economy.[b]

An economy consists of N agents, of which $K = kN$ are capitalists and $L = (1-k)N$ are workers. Each capitalist owns a factory and each worker owns one unit of labor power. K capitalists and L workers can produce $\sqrt{KL}$ units of output. What is the value and the uncertainty of the game for the participants?

Proposition 1. *The value and the uncertainty of the game for a worker are given by the following asymptotic formulas*:

$$V_w(G) \sim \frac{1}{2}\sqrt{\frac{k}{1-k}} - \frac{1}{16}\frac{1-2k}{k^{1/2}(1-k)^{3/2}}\frac{\ln N}{N}, \tag{10}$$

$$R_w(G) \sim \frac{1}{4(1-k)}\sqrt{\frac{\ln N}{N}}, \tag{11}$$

where $N \to \infty$. *Similarly, for a capitalist the value and the uncertainty are*

$$V_c(G) = \frac{1}{2}\sqrt{\frac{1-k}{k}} + \frac{1}{16}\frac{1-2k}{k^{3/2}(1-k)^{1/2}}\frac{\ln N}{N}, \tag{12}$$

$$R_c(G) \sim \frac{1}{4k}\sqrt{\frac{\ln N}{N}}, \tag{13}$$

where $N \to \infty$.

Proof is in the Appendix.

In this simple example of production economy, the uncertainty of the game decreases as the size of the game grows. The intuition behind this is simple: As the number of participants grows, almost all coalitions have approximately the same structure as the total population. For all these typical coalitions an agent has almost the same marginal contribution, so his uncertainty is low.

[b]This example is based on an example of a production economy in Shapley and Shubik (1967) and Osborne and Rubinstein (1994).

Example 5. Large market economy.

Suppose there are just two commodities in the economy, apples and bread. Assume there are kN apple traders and $(1-k)N$ bread traders. Suppose also that initially each trader has one unit of commodity of his own type, and that the utility function of every trader is the same:

$$u(a,b) = \sqrt{ab}, \tag{14}$$

where a is the number of apples and b is the quantity of bread. Assume that the utility is transferable.

Then a coalition of A apple and B bread traders has worth

$$G(S) = \sqrt{AB}, \tag{15}$$

and the problem is completely analogous to the problem of Example 4. Thus, for the apple trader the value and the uncertainty are

$$V_a(G) = \frac{1}{2}\sqrt{\frac{1-k}{k}} + \frac{1}{16}\frac{1-2k}{k^{3/2}(1-k)^{1/2}}\frac{\ln N}{N}, \tag{16}$$

$$R_a(G) \sim \frac{1}{4k}\sqrt{\frac{\ln N}{N}}, \tag{17}$$

where $N \to \infty$.

The expressions for the value and uncertainty of the bread trader are similar. Therefore, this example suggests that in the market economy, as well as in the production economy, the uncertainty approaches zero when the size of the economy increases.

Example 6. Risk factor attribution.

The Shapley values for the epidemiological example from Gefeller and Land (1997) are 0.405, 0.08, 0074 respectively for abnormal LDL, VLDL, and HDL cholesterol levels and 0.211 for smoking. The corresponding Shapley uncertainties can be computed as 0.118, 0.056, 0.057, and 0.11. The standard statistical test of significance suggests that only the effects of abnormal LDL and smoking are significant at a 5 percent significance level.

4. Properties of the Shapley Uncertainty

Several of the properties of the Shapley uncertainty are similar to the properties of the Shapley value and follow directly from the definition. They are scaling, "zero risk for dummy", and symmetry properties:

Proposition 2 (Scaling). $R_i(tG) = tR_i(G)$.

Proposition 3 (Dummy). *If $G(X+i) = G(X) + G(i)$ for any X that does not contain i, then $R_i(G) = 0$.*

Proposition 4 (Symmetry). *If $G(X) = G(X - i + j)$ for any X that contains i, then $R_i(G) = R_j(G)$.*

For two-player games a stronger symmetry property holds. It states that in two-player games both players are always equally exposed to the Shapley uncertainty even if their values are different.

Proposition 5 (Strong Symmetry). *For any 2-player game $R_i(G) = R_j(G)$.*

For games with more than two players this property is not true but still the risks of different players cannot be too different, as the following proposition shows.

Proposition 6. *For every N-player game G*

$$R_i(G) \leq \sqrt{(N-1)\sum_{j \neq i} R_j^2(G)}. \tag{18}$$

Proof. First, the Shapley uncertainty can be defined in an equivalent way. Let $\mathfrak{B}$ be the space of all orderings of elements in A. Define a map $p_i : \mathfrak{B} \rightarrow 2^A$ that takes an ordering to the subset of elements of A that precede element i in the ordering, and let $\widetilde{d_i G} = d_i G \circ p_i$. Take a random variable $g : X \rightarrow \mathfrak{B}$ that has a uniform distribution over orderings. Then it is easy to check that:

$$f_i = p_i \circ g, \tag{19}$$

$$V_i(G) = \mathrm{E}(\widetilde{d_i G} \circ g), \tag{20}$$

$$R_i(G) = \mathrm{Var}^{1/2}(\widetilde{d_i G} \circ g). \tag{21}$$

Second, for any ordering $b \in \mathfrak{B}$, the sum of marginal contributions of all players is equal to the value of the game, which is not random:

$$\sum_i \widetilde{d_i G}(b) = G(A). \tag{22}$$

It follows

$$\mathrm{Var}(\widetilde{d_i G} \circ g) = \mathrm{Var}\left(\sum_{j \neq i} \widetilde{d_j G} \circ g\right) \leq (N-1)\sum_{i \neq j} \mathrm{Var}(\widetilde{d_j G} \circ g), \tag{23}$$

where the last inequality is a well-known inequality for the variance of the sum of random variables. So,

$$R_i^2(G) \leq (N-1)\sum_{j \neq i} R_j^2(G). \tag{24}$$

□

The properties above compare the Shapley uncertainties of the same game for different players. In the following proposition we state a property of the uncertainty relative to the addition of games.

Proposition 7. *For each pair of games G and H,*

$$R_i(\alpha G + (1-\alpha)H) \le \sqrt{\alpha R_i^2(G) + (1-\alpha)R_i^2(H)}. \tag{25}$$

The equality is only possible if the games are proportionate to each other.

Proof. The square of the uncertainty of a game is the variance of a random variable. The assertion of the proposition follows from the well-known property of the variance of a weighted sum of two random variables. □

This property implies the following upper bound for the Shapley uncertainty of the sum of two games:

Corollary 1. *For each pair of games G and H,*

$$R_i(G+H) \le \sqrt{2}\sqrt{\left[R_i^2(G) + R_i^2(H)\right]}. \tag{26}$$

The equality is only possible if the games are proportionate to each other.

Corollary 1 gives an upper bound for the uncertainty of the sum of two games. What can be said about the lower bound? The question boils down to whether two games with large Shapley uncertainty can add up to a game with low uncertainty. Numerical experiments suggest the following conjecture:

Conjecture 1. *There exists a function $d(N) > 0$ such that for each pair of superadditive N-player games G and H*

$$R_i(G+H) \ge d(N)\sqrt{\left(R_i^2(G) + R_i^2(H)\right)}. \tag{27}$$

This conjecture is true for symmetric convex games. By definition, in these games the worth of a coalition depends only on its size, and the dependence is non-decreasing and convex:

$$G(S) = g(|S|), \tag{28}$$

where $g'(x) \ge 0$, and $g''(x) \ge 0$. For these games, the following proposition holds:

Proposition 8. *For each pair of symmetric convex games G and H,*

$$R_i(G+H) \ge \sqrt{R_i^2(G) + R_i^2(H)}. \tag{29}$$

Proof. Clearly the proposition holds if player i's marginal contributions have positive covariance for each pair of symmetric convex games G and H:

$$\mathrm{Cov}\{d_iG(S), d_iH(S)\} \ge 0, \tag{30}$$

where S is the random coalition. In other words, it is sufficient to prove that

$$\mathrm{E}\{d_iG(S)d_iH(S)\} - \mathrm{E}\{d_iG(S)\}\mathrm{E}\{d_iH(S)\} \ge 0. \tag{31}$$

Marginal contributions $d_iG(S)$ and $d_iH(S)$ are non-negative and non-decreasing in argument S, and therefore inequality (31) holds because of the following lemma:

Lemma 1. *If $f_1(x)$ and $f_2(x)$ are two non-negative, non-decreasing functions, and $\mathcal{F}(x)$ is the cumulative distribution function of a random variable, then*

$$\int_{-\infty}^{\infty} f_1(x)f_2(x)d\mathcal{F}(x) \geq \left(\int_{-\infty}^{\infty} f_1(x)d\mathcal{F}(x)\right)\left(\int_{-\infty}^{\infty} f_2(x)d\mathcal{F}(x)\right). \tag{32}$$

Proof of Lemma. Note that if the lemma holds for $\{f_1', f_2\}$ and $\{f_1'', f_2\}$, then it also holds for $\{\alpha' f_1' + \alpha'' f_1'', f_2\}$, where α' and α'' are non-negative coefficients. Similarly, if the lemma holds for $\{f_1, f_2'\}$ and $\{f_1, f_2''\}$, then it also holds for $\{f_1, \beta' f_2' + \beta'' f_2''\}$, where β' and β'' are non-negative. It follows that the lemma needs only to be proved for a set of elementary functions that can approximate all non-negative, non-decreasing functions by linear combinations with non-negative coefficients. It is enough to take the set of the following elementary functions ϕ_T :

$$\phi_T(x) = \begin{cases} 1 & \text{if } x \geq T, \\ 0 & \text{if } x < T, \end{cases} \tag{33}$$

which are characteristic functions of sets $[T; \infty)$.

Define measure $\mu_{\mathcal{F}}$ as follows:

$$\mu_{\mathcal{F}}\{X\} = \int_{-\infty}^{\infty} \chi_X(x)d\mathcal{F}(x), \tag{34}$$

where $\chi_X(x)$ denotes the characteristic function of set X.

In terms of this measure, the inequality for elementary functions becomes:

$$\mu_{\mathcal{F}}\{[T_1, \infty) \cap [T_2, \infty)\} \geq \mu_{\mathcal{F}}\{[T_1, \infty)\}\mu_{\mathcal{F}}\{[T_2, \infty)\}, \tag{35}$$

which is evidently true because $[T_1, \infty) \cap [T_2, \infty)$ is either $[T_1, \infty)$ or $[T_2, \infty)$, and $\mu_{\mathcal{F}}\{X\} \leq 1$ for any X. □

5. Conclusion

In the course of this paper, the concept of the Shapley uncertainty has been rigorously defined and illustrated by means of several examples. Salient properties of this concept have been derived, properties that show how the uncertainty behaves under addition of games, scaling in the size of the game, and substitution in the role of players. These examples and properties illustrate that the Shapley uncertainty is an interesting object. The ultimate question, however, is: What is this concept good for?

Actually, it is good for several important purposes. First, it measures the reliability of the Shapley value in solving various practical problems. Second, it is useful in checking the robustness of theoretical arguments that employ the Shapley value as the embodiment of the outcome in market games. Third, the Shapley uncertainty is an additional dimension that a player should take into account if he evaluates the prospect of playing a game. Finally, the Shapley uncertainty may be helpful in the design of games as implementation mechanisms, where the designer is interested in ensuring the stability of the outcome and making the mechanism attractive to risk-sensitive participants.

Appendix A. Proof of Proposition 1

For a coalition of size X with Y capitalists, the contribution of an additional worker is

$$c_{X,Y} = \sqrt{Y(X-Y+1)} - \sqrt{Y(X-Y)}. \tag{A.1}$$

From (1), the probability that a random coalition has size X is N^{-1}, where X can take values $0, \ldots, N-1$. The probability that a given random coalition of size X has Y capitalists is

$$P_{Y|X} = \frac{C_Y^K C_{X-Y}^{N-K-1}}{C_X^{N-1}}, \tag{A.2}$$

where C_k^n is the notation for binomial coefficients. Indeed, C_X^{N-1} is the total number of different coalitions of size X that exclude worker i. Among those coalitions, the total number of coalitions that have Y capitalists and $X-Y$ workers is $C_Y^K C_{X-Y}^{N-K-1}$.

Consequently, computing the first and the second moments of the marginal worker contribution means computing the following sums:

$$Ec_{X,Y} = \sum_{X=0}^{N-1} \sum_{Y=0}^{X} c_{X,Y} \frac{1}{N} \frac{C_Y^K C_{X-Y}^{N-K-1}}{C_X^{N-1}}, \tag{A.3}$$

$$Ec_{X,Y}^2 = \sum_{X=0}^{N-1} \sum_{Y=0}^{X} c_{X,Y}^2 \frac{1}{N} \frac{C_Y^K C_{X-Y}^{N-K-1}}{C_X^{N-1}}, \tag{A.4}$$

where we use the convention that $C_k^n = 0$ if $k < 0$ or $k > n$.

We will derive asymptotic approximations for these sums as N approaches infinity. The derivation is non-rigourous and proceeds by a sequence of non-controlled approximations. However, we will give some heuristic reasons for why these approximations are likely to lead to the correct answer.

First thing to note is that $c_{X,Y} \leq \sqrt{X}$. Consequently, the contribution of coalitions with $X \leq C$ to (A.4) is

$$\sum_{X=0}^{C} \sum_{Y=0}^{X} c_{X,Y}^2 \frac{1}{N} P_{Y|X} \leq C^2 \frac{1}{N}. \tag{A.5}$$

It will be clear from the following that as N grows, this contribution becomes negligible relative to the contribution of the terms with $X > C$.

Let us define the frequency of capitalists in a coalition, $f = Y/X$. The joint probability of f and X follows from (A.2):

$$P_{f,X} = \frac{1}{N} \frac{C_{fX}^K C_{X-fX}^{N-K-1}}{C_X^{N-1}}. \tag{A.6}$$

Here it is understood that f can only take such values that fX is an integer. We can use the Stirling formula to approximate $P_{f,X}$ for large values of N, X, K, and $N - K$. The density of the resulting distribution is:

$$P_{f,X} \sim \frac{\sqrt{NX}}{\sqrt{2\pi}\sqrt{K(N-K)(N-X)}} \exp\left[-\frac{1}{2}\left(f - \frac{K}{N}\right)^2 \frac{N^3 X}{K(N-K)(N-X)}\right] \tag{A.7}$$

In other words, conditional on X, the frequency of capitalists in a coalition is distributed approximately according to the Gaussian law with expectation $E(f) = K/N = k$, and variance

$$V(f) = \frac{K(N-K)(N-X)}{N^3}\frac{1}{X} = k(1-k)\frac{N-X}{N}\frac{1}{X}. \tag{A.8}$$

The key fact for the validity of the proposition is that as the coalition size grows the variance of the frequency of capitalists declines as X^{-1}. Up to the terms that are $o(X^{-1})$ (that is, up to the terms that approach 0 with probability 1 after multiplying by X as X grows), we can approximate the worker's contribution as

$$c_{X,fX} = X\left[\sqrt{f(1-f+X^{-1})} - \sqrt{f(1-f)}\right] \tag{A.9}$$

$$\sim \frac{1}{2}\sqrt{\frac{f}{1-f}} - \frac{1}{8}\frac{f^{1/2}}{(1-f)^{3/2}}\frac{1}{X} \tag{A.10}$$

$$\sim \frac{1}{2}\sqrt{\frac{k}{1-k}} - \frac{1}{8}\frac{k^{1/2}}{(1-k)^{3/2}}\frac{1}{X} + \frac{1}{4}\frac{1}{k^{1/2}(1-k)^{3/2}}(f-k) \tag{A.11}$$

$$-\frac{1}{16}\frac{1-4k}{k^{3/2}(1-k)^{5/2}}(f-k)^2. \tag{A.12}$$

Similarly for the square of the contribution we have the following approximation:

$$c^2_{X,fX} \sim \frac{1}{4}\frac{f}{1-f} - \frac{1}{8}\frac{f}{(1-f)^2}\frac{1}{X} \tag{A.13}$$

$$\sim \frac{1}{4}\frac{k}{1-k} - \frac{1}{8}\frac{k}{(1-k)^2}\frac{1}{X} + \frac{1}{4}\frac{1}{(1-k)^2}(f-k) \tag{A.14}$$

$$+\frac{1}{4}\frac{1}{(1-k)^3}(f-k)^2. \tag{A.15}$$

Computing the expectation over the distibution of f (conditional on X) we get

$$E(c_{X,fX}|X) \sim \frac{1}{2}\sqrt{\frac{k}{1-k}} - \frac{1}{8}\frac{k^{1/2}}{(1-k)^{3/2}}\frac{1}{X} \tag{A.16}$$

$$-\frac{1}{16}\frac{1-4k}{k^{1/2}(1-k)^{3/2}}\frac{N-X}{N}\frac{1}{X}, \tag{A.17}$$

$$E\left(c^2_{X,fX}|X\right) \sim \frac{1}{4}\frac{k}{1-k} - \frac{1}{8}\frac{k}{(1-k)^2}\frac{1}{X} + \frac{1}{4}\frac{k}{(1-k)^2}\frac{N-X}{N}\frac{1}{X}. \tag{A.18}$$

The next step is to compute the expectation over the distribution of X. This is easy to do using the following approximations:

$$\sum_{X=1}^{N}\frac{1}{X}\sim\sum_{X=1}^{N}\frac{N-X}{N}\frac{1}{X}\sim\ln N. \tag{A.19}$$

We get:

$$Ec_{X,fX}\sim\frac{1}{2}\sqrt{\frac{k}{1-k}}-\frac{1}{16}\frac{1-2k}{k^{1/2}(1-k)^{3/2}}\frac{\ln N}{N}, \tag{A.20}$$

$$Ec^2_{X,fX}\sim\frac{1}{4}\frac{k}{1-k}+\frac{1}{8}\frac{k}{(1-k)^2}\frac{\ln N}{N}. \tag{A.21}$$

Consequently

$$\mathrm{Var}^{1/2}(c_{X,fX})=\sqrt{E_f c^2_{X,fX}-(Ec_{X,fX})^2}\sim\frac{1}{4(1-k)}\sqrt{\frac{\ln N}{N}}. \tag{A.22}$$

□

References

Aumann, R. J. [1975] "Value of markets with a continuum of traders," *Econometrica* **43**, 611–646.

Evans, R. A. [1996] "Value, consistency, and random coalition formation," *Games and Economic Behavior* **12**, 68–80.

Gefeller, O. and Land, M. [1997] "A game-theoretic approach to partitioning attributable risks in epidemiology," *Biometrical Journal* **39**, 777–792.

Gul, F. [1989] "Bargaining foundations of Shapley value," *Econometrica* **57**, 81–95.

Hart, S. and Mas-Colell, A. [1996] "Bargaining and value," *Econometrica* **64**, 357–380.

Osborne, M. J. and Rubinstein, A. [1994] *A Course in Game Theory*, (MIT Press).

Roth, A. E. [1977] "The Shapley value as a von Neumann-Morgenstern utility," *Econometrica* **45**, 657–664.

Roth, A. E. [1977b] "Bargaining ability, the utility of playing a game, and models of coalition formation," *J. Mathematical Psychology* **16**, 153–160.

Roth, A. E. [1988] "The expected utility of playing a game," in *The Shapley Value*, ed. Roth, A. E. (Cambridge University Press), pp. 51–70.

Rubinstein, A. [1982] "Perfect equilibrium in a bargaining model," *Econometrica* **50**, 97–110.

Shapley, L. S. [1953] "A value for n-Person Games," in *Contributions to the Theory of Games, Vol. 2*, eds. Kuhn, H. W. and Tucker, A. W. (Princeton, Princeton University Press), pp. 307–317.

Shapley, L. S. and Shubik, M. [1967] "Ownership and the production function," *Quarterly J. Economics* **81**, 88–111.

Weber, R. J. [1988] "Probabilistic values for games," in *The Shapley Value*, ed. Roth, A. E. (Cambridge University Press), pp. 101–119.