

## ОЦЕНКА РЕЗУЛЬТАТОВ ГОЛОСОВАНИЯ В УСЛОВИЯХ НАЛИЧИЯ ЗАВИСИМОСТИ МНЕНИЙ ГОЛОСУЮЩИХ.

И.В. Заляпин, И.В. Кузнецов, В.Ф. Писаренко,  
e-mail: zaliapin@mitp.rssi.ru

Международный Институт Теории Прогноза Землетрясений и Математической Геофизики РАН,  
Москва, Россия.

### ВВЕДЕНИЕ.

Общественно-социальные науки традиционно используют в своей работе опросы общественного мнения. При этом основное внимание уделяется способам построения репрезентативных выборок опрашиваемых, формированию опросника, формулировке вопросов и т.п. С эффективным решением таких проблем напрямую связывается надежность получаемых результатов. Это в первую очередь связано с тем, что проведение опроса является гораздо более трудоемким и дорогостоящим этапом исследования по сравнению с последующей обработкой собранного материала. Однако, надежность выводов может существенно зависеть также от статистических свойств моделей, используемых для анализа имеющихся данных (специфике обработки результатов социального эксперимента посвящена монография [1]). Ниже мы хотим продемонстрировать наличие такой зависимости, рассмотрев задачу оценки исхода голосования по результатам предварительного опроса.

### ПОСТАНОВКА ЗАДАЧИ.

Рассмотрим следующую задачу, часто возникающую на практике. Требуется оценить отношение людей из некоторой популяции (генеральной совокупности) к интересующему нас аспекту жизни (социальному, бытовому, экономическому, политическому и т.п.). Отметим, что это может интересовать нас как при мониторинговых исследованиях состояния общественного мнения, так и при оценке возможных исходов предстоящего всенародного голосования по тому или иному вопросу. Для простоты будем считать, что вся популяция делится на “оптимистов” и “пессимистов”, то есть отношение произвольно выбранного человека к рассматриваемому вопросу может быть либо строго положительным либо строго отрицательным. Говоря более определенно, на некоторый вопрос, предложенный каждому из представителей популяции, мы можем получить только два ответа: “да” (если отношение респондента положительно) или “нет” (если отношение отрицательно). Такой подход становится естественным при рассмотрении, например,

выборов президента или парламента в стране с двухпартийной системой. Кроме того, известным является факт увеличения надежности оценивания при таком бинарном делении общества в социальных исследованиях (см., например [2], [3]). Задача оценки общественного мнения или предсказания исхода выборов сводится в данном случае к определению доли “оптимистов”. Проводя опрос, мы обычно не имеем возможности опросить всех представителей генеральной совокупности, и, поэтому, оцениваем долю “оптимистов” по ответам представителей некоторой выборки, объем которой значительно меньше объема реальной популяции. Точность нашей оценки будет зависеть от множества факторов, таких как репрезентативность выборки, количество опрошенных и т.п. Традиционный подход сводит задачу подобного оценивания к определению параметров некоторых классических вероятностных моделей. Регулярная ошибка таких оценок убывает с ростом выборки как  $1/\sqrt{N}$ , где  $N$  – количество опрошенных. Таким образом, качественный вывод, который можно сделать рассматривая классические модели, состоит в том, что регулярная ошибка получаемых нами оценок устранима с помощью увеличения количества опрашиваемых. Эффективность увеличения объема выборки пропорциональна корню из его значения. Однако, в таких моделях не учитывается влияние опрашиваемых друг на друга, что делает надежность результатов неудовлетворительной. Необходимым становится рассмотрение моделей, учитывающих взаимозависимость взглядов при принятии групповых решений.

Изучению межличностной зависимости в теории принятия решений уделяется в последнее время большое внимание. Некоторые вероятностные и комбинаторные подходы к решению данной проблемы можно найти, например, в [4], [5], [6].

Цель предлагаемой нами модели – продемонстрировать возможность возникновения неустраняемых регулярных ошибок при оценке коллективного решения. Мы опишем возникновение ошибок оценивания вследствие существования зависимости между взглядами респондентов и оценим эффективность увеличения числа опрашиваемых.

## МОДЕЛЬ.

Рассмотрим общество, разделенное на группы по принципу общности некоторых интересов (например, культурных или социально-политических). В формировании отношения члена такого общества к произвольной проблеме играют роль как интересы группы, к которой он принадлежит, так и его личные предпочтения. Эти предпочтения могут в большей или меньшей степени совпадать или не совпадать с

групповыми интересами в зависимости от степени консолидации общества. Мы хотим оценить результаты всенародного опроса в этом обществе на основании ответов некоторой группы респондентов.

Итак, предположим, что популяция объема  $V$  состоит из  $m$  непересекающихся групп  $G_1, G_2, \dots, G_m$  таких, что мнения респондентов зависимы внутри каждой группы и независимы у респондентов из разных групп. Группы имеют различную численность, так что каждая группа является набором индивидуумов  $I_i^k$ :

$$G_k = \{I_1^{(k)}, I_2^{(k)}, \dots, I_{V_k}^{(k)}\}, \quad k = 1, \dots, m; \quad \sum_k V_k = V. \quad (1)$$

Каждому представителю популяции предлагается вопрос, на который можно ответить “да” или “нет”. Ответы респондентов будем обозначать буквой  $\xi$ . Тогда  $\xi_i^{(k)}$  будет обозначать ответ респондента  $I_i^{(k)}$  ( $i$ -го представителя группы  $G_k$ ):

$$\xi_i^{(k)} = \begin{cases} 0 = 'HET' \\ 1 = 'DA' \end{cases} \quad k = 1, \dots, m; \quad i = 1, \dots, V_k. \quad (2)$$

Определим случайные величины (*определяющие факторы*)  $\eta_k$ , задающие настроение в группе  $G_k$ :

$$\eta_k = \begin{cases} 1; & \text{Pr.} = \pi_k \\ 0; & \text{Pr.} = 1 - \pi_k \end{cases} \quad k = 1, \dots, m. \quad (3)$$

Пусть существует некоторая *вероятность согласия*  $p$ , с которой ответ респондента из группы  $G_k$  совпадает со значением соответствующего определяющего фактора:

$$P\{\xi_i^{(k)} = L \mid \eta_k = L\} = p; \quad P\{\xi_i^{(k)} = L \mid \eta_k = \bar{L}\} = 1 - p; \quad k = 1, \dots, m; \quad i = 1, \dots, V_k; \quad (4)$$

$$L = \begin{bmatrix} 0 \\ 1 \end{bmatrix};$$

и  $\bar{L}$  обозначает ответ, противоположный ответу  $L$ .

Внутри каждой группы ответы предполагаются условно независимыми:

$$P\{\xi_{i_1}^{(k)}, \dots, \xi_{i_l}^{(k)} \mid \eta_k\} = P\{\xi_{i_1}^{(k)} \mid \eta_k\} \cdot K \cdot P\{\xi_{i_l}^{(k)} \mid \eta_k\} \quad \forall (i_1, \dots, i_l) \in (1, \dots, V_k). \quad (5)$$

Будем считать, что распределение представителей популяции по группам подчиняется полиномиальному закону с параметрами  $(\alpha_1, \dots, \alpha_m)$ , то есть вероятность попадания в группу  $G_k$  для произвольно взятого респондента есть  $\alpha_k$ . (Отметим, что, таким образом,  $\alpha_k$  есть доля, составляемая членами группы  $G_k$  по отношению ко всей популяции.)

Вероятность получить положительный ответ от произвольного представителя группы  $G_k$  обозначим через  $p_k$ . Необходимо отметить, что параметры  $p, p_k, \pi_k$  связаны соотношением

$$p_k = p\pi_k + (1-p)(1-\pi_k) \quad (6a)$$

$$\text{или} \quad \pi_k = \frac{p + p_k - 1}{2p - 1} \quad (6б)$$

в силу формулы полной вероятности. Избыточность обозначений объясняется желанием иметь легко интерпретируемые параметры модели, описывающие различные ее свойства.

Таким образом, наша модель описывается следующими параметрами:

$\pi_k$  – вероятности “оптимизма” определяющих факторов,

$p$  – вероятность согласия с определяющим фактором ,

$\alpha_k$  – вероятность попадания в группу  $G_k$ ,

$p_k$  – вероятность “оптимизма” представителя группы  $G_k$ .

Наша цель состоит в выяснении свойств оценки параметра  $J = \sum_{k=1}^m p_k \alpha_k$  , то есть оценки реальной доли “оптимистов” в популяции по результатам опроса  $N$  респондентов. В качестве такой оценки рассмотрим случайную величину

$$\bar{J} = \frac{1}{N} \sum_{k=1}^m \sum_{i=1}^{v_k} \xi_i^{(k)}, \quad N = \sum_{k=1}^m v_k, \quad (7)$$

где  $v_k$  обозначает количество представителей группы  $G_k$ , принявших участие в опросе.

Можно показать, что

$$E\bar{J} = J \equiv \sum_{k=1}^m p_k \alpha_k; \quad (8)$$

$$D\bar{J} \equiv \frac{1}{N} D_1 + \frac{1}{N} D_2 + \left(1 - \frac{1}{N}\right) L, \quad (9)$$

где

$$D_1 = \sum_{k=1}^m \alpha_k p_k (1 - p_k),$$

$$D_2 = \sum_{k=1}^m \alpha_k p_k^2 - \left( \sum_{k=1}^m \alpha_k p_k \right)^2,$$

$$L = \sum_{k=1}^m \alpha_k^2 [p_k (1 - p_k) - p(1 - p)].$$

Таким образом, оценка является несмещенной, а дисперсия ее разлагается на три составляющих:  $\frac{1}{N} D_I$  – дисперсия, возникающая в модели без взаимозависимости ответов респондентов и таким же разбиением популяции на группы;  $\frac{1}{N} D_2$  – дисперсия полиномиального распределения вероятностей оптимизма по группам – мера “неоднородности” взглядов в популяции;  $L$  – член, определяющий степень зависимости мнений респондентов.

В невырожденных случаях, то есть при  $\pi_k \neq 0; 1$  имеем

$$\lim_{N \rightarrow \infty} DJ = L = \sum_{k=1}^m \alpha_k^2 [p_k(1 - p_k) - p(1 - p)] \neq 0. \quad (10)$$

Мы видим, что с некоторого момента точность оценки практически перестает улучшаться с ростом числа опрошенных  $N$ ; она стремится к предельному значению  $L$ , зависящему от соотношения параметров  $p$  и  $p_k$  (а точнее, от степени взаимозависимости мнений респондентов). Таким образом, получено ограничение на достижимую точность оценки.

## ОБСУЖДЕНИЕ

Предложенная выше модель рассмотрена с целью описания проблем, возникающих при оценке надежности результатов статистической обработки социологических данных, таких как результаты массовых опросов общественного мнения. Причем описания исключительно на качественном уровне. Реальные процессы, приводящие к появлению взаимозависимости мнений у членов общества, несомненно, настолько сложны, что не могут быть описаны простой вероятностной моделью и должны изучаться в рамках общественно-социальных наук. Тем не менее, основные выводы о надежности получаемых при обработке данных заключений нам представляется возможным сделать с помощью простейших теоретико-вероятностных методов.

Очевидная интерпретация параметров предложенной модели позволяет исследовать возможность ее применимости к широкому спектру реальных ситуаций.

*Модель парламентского голосования* предполагает естественное деление генеральной совокупности-парламента на группы-фракции. При этом известным является мнение лидера фракции по вопросу, выносимому на голосование ( $\pi_k$ ); уровень “партийной дисциплины” ( $p$ ) и состав парламента ( $\alpha_k$ ).

*Модель выборов президента* предполагает, что нам известно деление общества на социально-политические группы. Настроение каждой такой группы перед выборами определяется по ситуации в обществе (например, можно предположить, что работники коммерческих банков не поддержат коммунистов), а вероятность согласия  $p$

определяется по степени предполагаемого изменения условий существования той или иной группы при изменении политического руководства страны.

Можно привести ряд других ситуаций, для описания которых пригодна предложенная модель. Во всех случаях основные выводы вытекают из поведения дисперсии (9) оценки (7), отвечающей за надежность результатов. Взаимозависимость взглядов членов общества приводит к невозможности сколь угодно точно оценить мнение популяции по тому или иному вопросу. Увеличение числа опрашиваемых в опросах общественного мнения выше некоторого порога может оказаться неэффективным для повышения надежности выводов. При статистической обработке результатов социологических исследований особую важность приобретает вопрос о степени раздробленности общества и факторах, влияющих на консолидацию составляющих общество групп. Без учета такой информации надежность полученных результатов может оказаться недостаточной для их практического использования.

## ЗАКЛЮЧЕНИЕ

В теории принятия коллективных решений важную роль играют межличностные и групповые отношения участников. Методы обработки социологических данных должны опираться на строгий статистический аппарат, который учитывал бы наличие таких отношений при оценке надежности результатов. Классический подход, при котором важную роль играет предположение о независимости экспериментов, становится неприемлемым. Важной в этой ситуации является разработка адекватных математических моделей и их проверка на реальных социологических данных. Возможность качественного описания эффектов межличностных взаимодействий с помощью моделей, подобных описанной в данной работе, представляется реальной.

## СПИСОК ЛИТЕРАТУРЫ

1. Кемпбелл Д. Модели экспериментов в социальной психологии и прикладных исследованиях, СПб. 1996. 391с.
2. Lichtman A., Keilis-Borok V.I. Pattern recognition applied to presidential elections in the United States 1860-1980; Role of integral social, economic and political traits// Proc. Nat. Acad. Sci. USA, 1981. Vol.78. P.7230-7234.
3. Lichtman A., Keilis-Borok V.I. Aggregate-level analysis and prediction of midterm senatorial elections in the United States, 1974-1986// Ibid. 1989. Vol.86. P.1076-1080.
4. Hattori Yuichi Impossibility for social choice function with subordinate relation// Mem. Konan Univ., Sci. Ser. 1996. 43№1. P.67-72.

5. Hattori Yuichi Social choice function with subordinate relations and its subordinate degree// Mem. Konan Univ., Sci. Ser. 1996. 43№1. P.73-80.
6. Nida-Rumelin Julian Interpersonal dependency of preferences// Theory and Decis. 1996. 41№3. P.257-280.