Statistical analysis of the presidential elections in Belarus in 2020.

S.L. Cherkas

A Monte Carlo simulation is performed to calculate a probability of the difference between the average value in some random sample and the average over the total set. Method of analysis of the nature of the peculiarities in the probability distribution functions is suggested. The method consists of a comparison of the probability distribution functions for the percentage and the number of voters for Mr. Lukashenko in each polling station.

Keywords: random sample, average value, form of the distribution function, elections in Belarus, mathematical modeling

In any election process, as a rule, complaints to the election procedure arise from the losing party. The election of the president in Belarus is not an exception and has led to a serious split in society due to different opinions on this issue. Statistical analysis is an additional tool for analyzing the reliability of elections, although it cannot provide a completely definite answer about the fairness of the election result because statistical methods operate only with the probabilities of various statements. Currently, 1310 of the 5767 protocols of precinct election commissions (PECs) are available in the public domain [1]. It seems noteworthy, how reliable the election result declared as the victory of Mr. Lukashenko with 80.23% of the vote is. Below we will focus on two arrays of numbers: the number N_i of voters at some polling station and the number of voters for Mr. Lukashenko M_i at the same polling station [2]. These numbers could be considered as the random variables that enable to calculate the following average values

$$\bar{N} = \frac{1}{1310}\sum_{i}^{1310} N_i, \qquad \bar{M} = \frac{1}{1310}\sum_{i}^{1310} M_i, \tag{1}$$

where the first value is the average number of voters in some polling station and the second value is the average number of voters for Mr. Lukashenko. The percentage of those who voted for Mr. Lukashenko is defined as

$$\bar{\lambda} = \bar{M} / \bar{N} = \sum_{i}^{1310} M_i \Big/ \sum_{i}^{1310} N_i \approx 0.62, \tag{2}$$

i.e. it turns out to be at a 60% level. According to the official result from the analysis of the total number of 5767 PEC protocols, the percentage of voters for Mr. Lukashenko is about 80%. Thus, the result obtained from a random sample of approximately ¼ of all PEC protocols differs more than 15% from the official value. We have estimated the probability of such event by the Monte Carlo method. For this purpose, we randomly select ¼ protocols from the existing sample of 1310 protocols and calculate the average. This procedure has been repeated until appearing of a result that differs from $\bar{\lambda}$ more than ± 0.15. Computer simulation shows that with more than a million random samples, such a result does not appear. This indicates that the probability of such an event is less than 10^{-6}. In fact, this probability is even less, because a fourth part of the 5767 protocols has to been chosen, rather than that from 1310. This tiny probability simply indicates the fact, that a sufficiently large random sample must reflect the full picture accurately.

Let us now perform a more detailed analysis of the sets N_i and M_i. One could introduce the following random variable

$$\lambda_i = M_i / N_i, \tag{3}$$

representing the percentage of those who voted for Mr. Lukashenko at each polling station and consider the distribution of the probability density [3] of this value. The distribution smoothed with the Gaussian kernel is shown in Fig. 1. As is well known, the probability density [3] characterizes the probability of obtaining λ in a certain interval $\lambda \in [a,b]$ as $P = \int_a^b \rho(\lambda) d\lambda$. For example, the probability of obtaining a value at some polling station $0.6 < \lambda < 0.8$ equals the area under the graph curve from 0.6 to 0.8 and takes the numerical value $P = 0.43$.

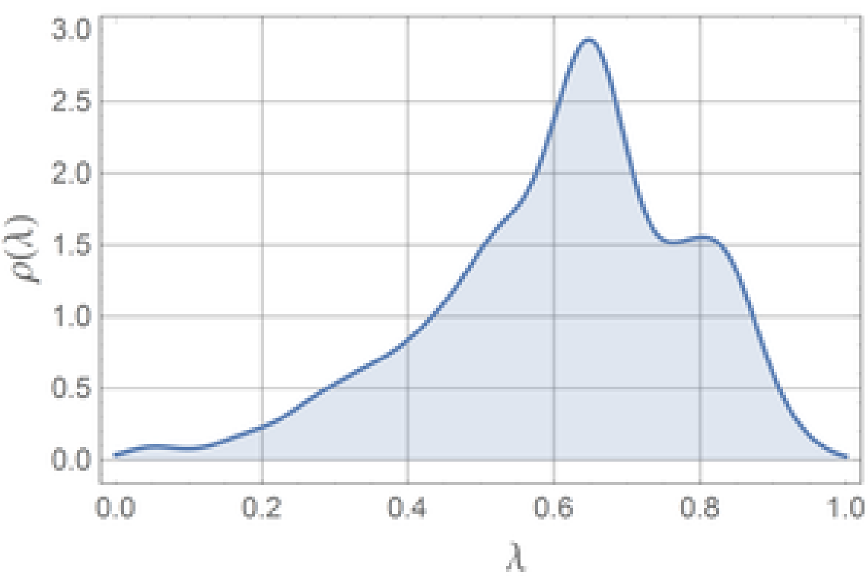

Fig. 1 The probability density function $\rho(\lambda)$ of the percentage of the voters for Mr. Lukashenko.

As one could see from Fig.1, the probability density differs considerably from the normal (Gaussian) distribution which is usually observed in the elections. Of course, this fact itself does not tell anything, since there are elections with strong deviations from the normal distribution. However, it is notable that besides the expected maximum near 60%, an additional maximum at 80% has appeared. To analyze this phenomenon, let us to construct a distribution of another quantity, namely, the number of voters who voted for Mr. Lukashenko at each polling station, by considering the normalized random quantity

$$\nu_i = M_i / \bar{N} . \tag{4}$$

Since the number of voters is different in the different polling stations, the distribution $\rho(\nu)$ should be broader than $\rho(\lambda)$, but still retains its key points. The distribution $\rho(\nu)$ is shown in Fig.2

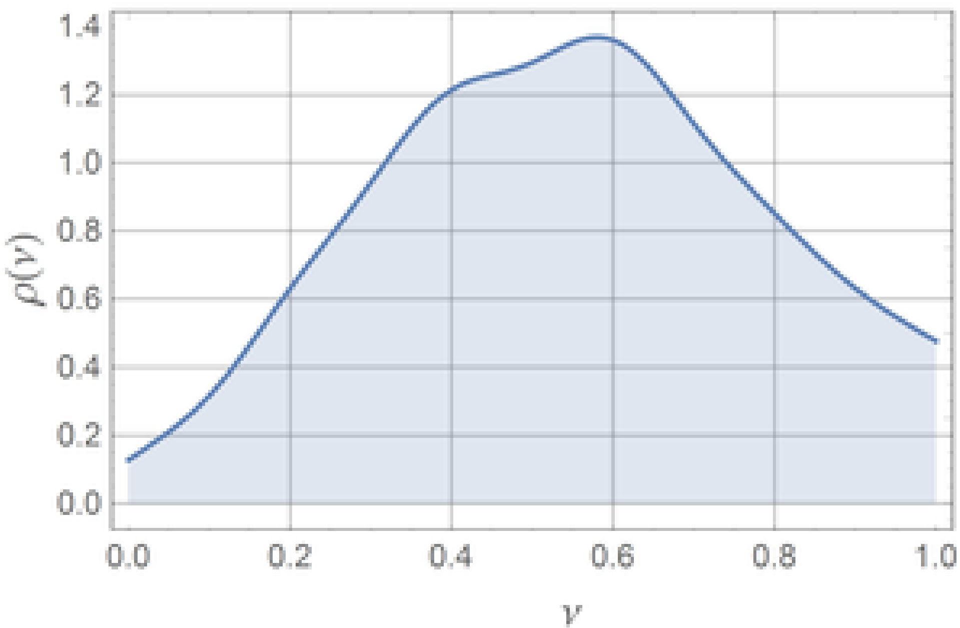

Fig. 2 The probability density function $\rho(\nu)$ of the number of the voters for Mr. Lukashenko normalized by the mean number of voters according to (4).

As can be seen from Fig. 2, the peak at 0.6 remains, although somewhat shifted, while the peak at 0.8 has disappeared traceless. One of the possible hypothesis is that the correlation in distribution $\rho(\lambda)$ arose due to artificial trimming of the percentage of those who voted for Mr. Lukashenko to 80% in some polling stations. Of course, it is only one of the possible explanations why the peak at 80% is in the distribution $\rho(\lambda)$ of the probability density of percentage, but not in the probability density $\rho(\nu)$ of the number of voters who voted for Mr. Lukashenko.

As a result, we note the following features of the statistical analysis:

1) the average percentage of those who voted for Mr. Lukashenko in the available sample of 1310 protocols is about 60%, which differs greatly from the official 80% percentage. The probability of such an event is tiny.

2) The distribution of the percentage of those who voted for Mr. Lukashenko in some polling station is very different from the normal (Gaussian distribution). It has a maximum of about 60% and an additional peak of about 80%.

3) The additional peak at 80% disappears completely if one considers a distribution not of the percentage, but of the number of people who voted for Mr. Lukashenko at each polling station. One hypotheses of this phenomenon may be the artificial equalization of the percentage of those who voted for Mr. Lukashenko to 80% in some polling stations, which, however, turned out to be insufficient to bring the average from 60% to the official value 80%.

As for the official result, of course, it could be considered that Mr. Lukashenko earned 80.23% of the votes, while in a random sample of 1310 there were about 60%, but the probability of such an event is less than a millionth.

References

1. Photographs of the 1310 protocol of precinct election commissions https://docs.google.com/spreadsheets/d/17aK3JxBTGtzULB0-YZGOF0hJwhuViHO3/edit#gid=84585767

2. *S.L. Cherkas*, (2020), Election data and computer codes for "Statistical analysis of the presidential elections in Belarus in 2020", Dryad, Dataset, https://doi.org/10.5061/dryad.d7wm37q0d

3. *J. Mathews, R. L. Walker*. Mathematical Methods of Physics, Benjamin, New York, 1964.

Статистический анализ выборов президента Беларуси в 2020 г.

С.Л. Черкас

Выполнено моделирование методом Монте-Карло вероятности получить отличие среднего значения в некоторой случайной выборке от полного среднего. Предложен метод анализа причин возникновения особенностей в функции плотности распределения, состоящий в сравнении функций распределений процента и числа проголосовавших на каждом участке за определенного кандидата.

Ключевые слова: случайная выборка, среднее значение, форма функций распределения, выборы в Беларуси, математическое моделирование

...и мгновенно от сонма
Калхас восстал Фесторид, верховный птицегадатель.
Мудрый, ведал он все, что минуло, что есть и что будет,

Гомер, "Илиада"

В процессе любых выборов, как правило, возникают те или иные претензии к выборной процедуре со стороны проигравшей стороны. Выборы президента в Беларуси в этом смысле не являются исключением и привели к серьезному расколу общества из-за различных мнений по этому поводу. Статистический анализ является дополнительным инструментом анализа достоверности выборов, хотя и не может дать полностью определенного ответа на вопрос о справедливости результата выборов, поскольку статистические методы оперируют с только вероятностями различных утверждений. В настоящее время доступны в открытой печати 1310 из 5767 общего количества протоколов участковых выборных комиссий (УИК) [1]. Представляется интересным, насколько достоверен результат выборов, объявленный как победа А.Г. Лукашенко с результатом порядка 80%. Далее мы сконцентрируем внимание на двух массивах чисел: количество голосовавших на каждом i-том участке N_i и количество голосовавших M_i за А.Г. Лукашенко на этом же участке [2]. Как M_i так N_i можно рассматривать как случайные величины и вычислить следующие средние значения

$$\bar{N} = \frac{1}{1310}\sum_{i}^{1310} N_i, \qquad \bar{M} = \frac{1}{1310}\sum_{i}^{1310} M_i, \qquad (1)$$

где первая величина представляет собой среднее число голосовавших на некотором участке, а вторая величина представляет собой среднее число голосовавших за г-на Лукашенко. Процент голосовавших за А.Г. Лукашенко определяется как

$$\bar{\lambda} = \bar{M} / \bar{N} = \sum_{i}^{1310} M_i / \sum_{i}^{1310} N_i \approx 0.62, \qquad (2)$$

и оказывается на уровне 60%. Согласно официальными результатам из анализа полного числа 5767 протоколов УИК количество голосовавших за А.Г. Лукашенко составляет 80.23%. Таким образом, получается, что случайная выборка примерно ¼ всех протоколов УИК дает результат отличающийся более чем на 15%. Оценим вероятность такого события методом Монте-Карло. Для этой цели случайным образом выберем ¼ протоколов из имеющейся выборки в 1310 протоколов и вычислим среднее. Данную процедуру будем повторять до появления результата, отличающегося от $\bar{\lambda} = 0.6$ на плюс минус 0.15. Компьютерное моделирование показывает, что при проведении более чем миллиона выборок такой результат не появляется. Это указывает, что вероятность такого события меньше 10^{-6}. На самом деле вероятность еще меньше, поскольку следовало бы выбирать четвертую часть из 5767 протоколов а не из 1310. Полученная малая вероятность просто указывает на тот факт, что достаточно большая случайная выборка должна объективно отражать полную картину.

Выполним теперь более детальный анализ множеств M_i и N_i. Можно ввести следующую случайную величину

$$\lambda_i = M_i / N_i, \qquad (3)$$

представляющую собой процент проголосовавших за А.Г. Лукашенко на каждом участке и рассмотреть распределение плотности вероятности $\rho(\lambda)$ данной величины [3]. Сглаженное с помощью Гауссова ядра распределение показано на Рис. 1. Как известно, плотность вероятности [3] характеризует вероятность P получить значение величины в некотором интервале $[a,b]$ как $P = \int_a^b \rho(\lambda) d\lambda$. Например, вероятность получить на каком-нибудь одном избирательном участке значение $0.6 < \lambda < 0.8$ равно площади под кривой графика от 0.6 до 0.8 и численно равна $P = 0.43$.

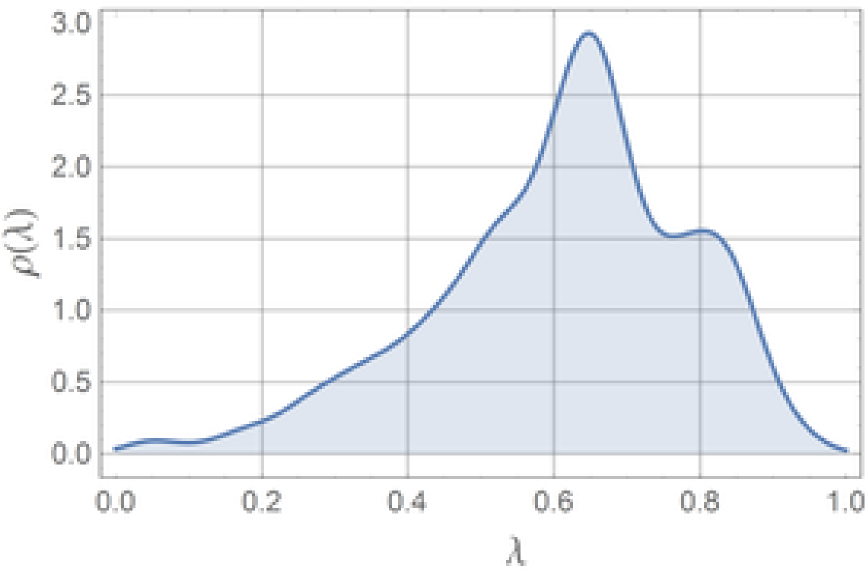

Рис. 1 Функция $\rho(\lambda)$ распределения плотности вероятности процента голосовавших за г-на Лукашенко.

Как видно из Рис.1, плотность вероятности $\rho(\lambda)$ далека от нормального (Гауссова) распределения, которое обычно наблюдается на выборах. Разумеется, сам по себе этот факт ни о чем не говорит, поскольку встречаются выборы с сильными отклонениями от нормального распределения. Интересным здесь является то, что кроме ожидаемого максимума на $\lambda \approx 0.6$, имеется также максимум на $\lambda \approx 0.8$. Для анализа этого явления построим распределение другой случайной величины: числа избирателей M_i голосовавших за А.Г. Лукашенко на каждом участке рассматривая нормированную случайную величину

$$\nu_i = M_i / \bar{N} \,. \tag{4}$$

Поскольку на разных участках число проголосовавших различно, то распределение $\rho(\nu)$ должно быть более размытым, чем $\rho(\lambda)$ однако сохранять его ключевые моменты. Распределение $\rho(\nu)$ приведено на Рис. 2

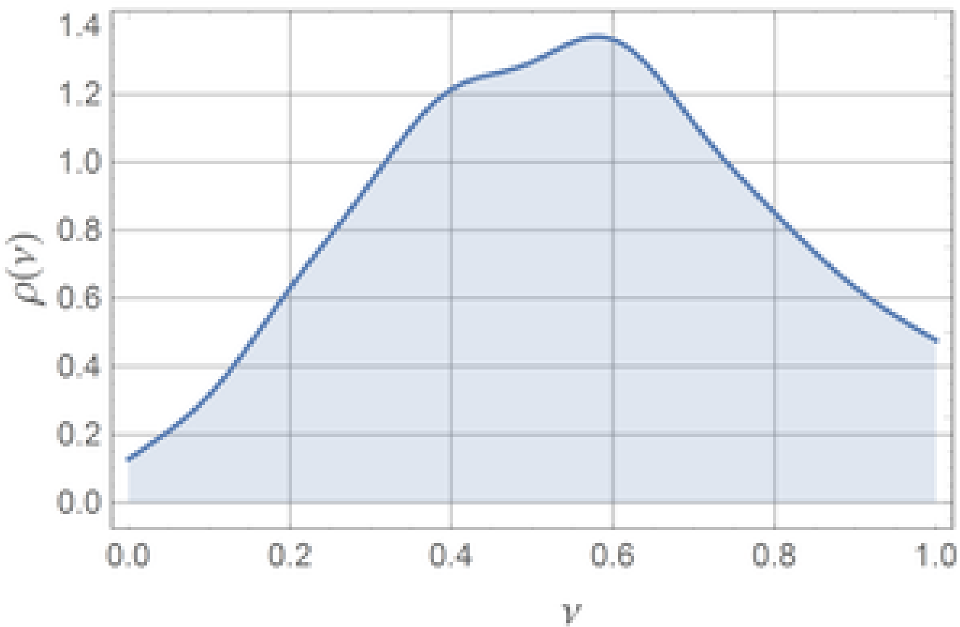

Рис. 2 Функция $\rho(\nu)$ распределения плотности вероятности числа голосовавших за г-на Лукашенко нормированного на среднее число голосовавших на участке, см. (4).

Как видно из Рис. 2 , пик на 0.6 остался, хотя несколько сместился, в то время как пик на 0.8 бесследно исчез. Одна из возможных гипотез состоит в том, что корреляция в распределении $\rho(\lambda)$ возникла из-за искусственного подравнивания на некоторых участках именно процента проголосовавших за А.Г. Лукашенко к 80. Разумеется, можно придумать и другие объяснения, почему пик на 80 есть в распределении $\rho(\lambda)$ плотности вероятности процента, но отсутствует в плотности вероятности $\rho(\nu)$ числа избирателей, голосовавших за А.Г. Лукашенко.

В результате отметим следующие особенности проведенного статистического анализа:

1) средний процент проголосовавших за А.Г. Лукашенко в выборке доступных 1310 протоколов порядка 60%, что сильно отличается от официальных 80% процентов. Вероятность такого события составляет меньше 10^{-6}

2) Распределение процента проголосовавших за А.Г. Лукашенко по участкам сильно отличается от нормального (Гауссова распределения), имеет максимум примерно на 60% и дополнительный пик примерно на 80%.

3) Дополнительный пик полностью исчезает, если рассматривать распределение не процента, а числа проголосовавших за А.Г. Лукашенко на каждом участке. Одной из гипотез данного явления может быть

искусственное подравнивание процента проголосовавших за А.Г. Лукашенко к 80% на некоторых участках, которое, тем не менее, оказалось недостаточным, чтобы приблизить среднее значение (т.е. около 60%) по исследуемой выборке к официально объявленному значению 80%.

Что касается официального результата, то, разумеется, можно считать, что А.Г. Лукашенко действительно набрал 80.23% голосов, в то время, как в случайной выборке из 1310 оказалось около 60%, но вероятность такого события меньше одной миллионной.

СПИСОК ЛИТЕРАТУРЫ

1. Фотографии 1310 протоколов участковых избирательных комиссий https://docs.google.com/spreadsheets/d/17aK3JxBTGtzULB0-YZGOF0hJwhuViHO3/edit#gid=84585767

2. *S.L. Cherkas*, (2020), Election data and computer codes for "Statistical analysis of the presidential elections in Belarus in 2020", Dryad, Dataset, https://doi.org/10.5061/dryad.d7wm37q0d

3. *Дж. Мэтьюз, Р. Уокер.* Математические методы физики, Мир, 1977.