

Statistical analysis of the presidential elections in Belarus in 2020.

SL Cherkas

ARTICLE INFO

Keywords:

Belarus elections 2020
statistics
precinct level data analysis

ABSTRACT

Usually, one wants to have a simple picture of the trustworthiness of the main elections result. However, in some situations only partial information about the elections is available. Here we suggest some criterion of comparing of the available information with the official results. One of the criterions consists in comparison of the mean value over available sample with the official mean value. 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. Another method is an analysis of the nature of the peculiarities in the probability distribution functions consisting in comparison of the probability distribution functions for the percentage and the number of voters for Mr. Lukashenko in each polling station. The last criterion is rather esthetic than exposing. It could be applied to arbitrary elections systems such as United Kingdom or United States if one wants to extract the main result in a few pictures.

1. Introduction

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.

Belarus has direct one level presidential election system. At the same time, some peculiarities exist. One of such features is that the election result over every polling station must be posted up on the wall of the station after the elections in Belarus. So, in principle, everybody could catch this result. However, duration of the result exposition is not specified by the Belarus law. That duration could be even a few minutes. This explains why not all the precinct level data were available after the Belarus elections. Consequently, one falls into the situation where an analysis of the incomplete information is needed. Although the unofficial data exist accumulated by Golos organization and analyzed by Gongalsky (2020) in addition to sociological polls (Zahorski, 2020; Charter97, 2020), here we use only official data: from the one hand, officially posted on the wall of polling stations precinct protocols (ZUBR, 2020), and, from the other hand, officially declared result of the elections for every region of Belarus and city of Minsk.

2. Results for Belarus as a whole.

Statistical analysis is an additional tool for analyzing the reliability of elections, although it cannot provide a completely definite answer about fairness of the election result, because statistical methods operate only with the probabilities of various statements. Currently, 1527 of the 5767 protocols of precinct election commissions (PECs) are available in the open access (ZUBR, 2020). It seems interesting, how reliable the election result declared as the victory of Mr. Lukashenko with a result about 80% is. Below we will focus an attention 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 Cherkas (2020). These numbers could be considered as the random variables giving possibility to calculate the following average values

$$\bar{N} = \frac{1}{1527} \sum_i^{1527} N_i, \quad \bar{M} = \frac{1}{1527} \sum_i^{1527} M_i, \quad (1)$$

where the first value is the average number of voters in some polling station and the second value is the average number

✉ cherkas@inp.bsu.by (S. Cherkas)
ORCID(s): 0000-0002-6132-3052 (S. Cherkas)

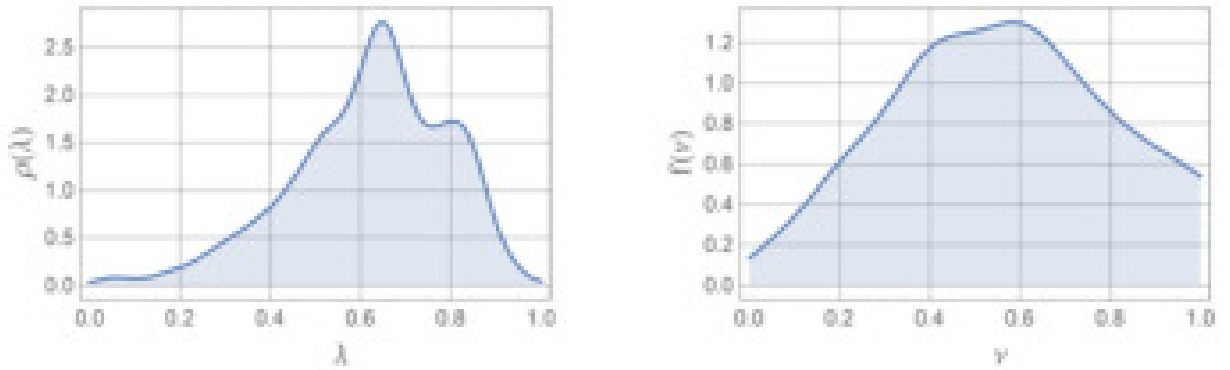


Figure 1: The probability density function $\rho(\lambda)$ of the percentage of the voters for Mr. Lukashenko. The probability density function $f(v)$ of the number of the voters for Mr. Lukashenko normalized by the mean number of voters according to (4). A random polling station is implied.

of voters for Mr. Lukashenko. The percentage of those who voted for Mr. Lukashenko is defined as

$$\bar{\lambda} = \bar{M} / \bar{N} = \sum_i^{1527} M_i / \sum_i^{1527} N_i \approx 0.63, \quad (2)$$

i.e. it turns out to be at a 63% level. According to 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 $1/4$ 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 $1/4$ protocols from the existing sample of 1527 protocols and calculate the average. This procedure has been repeated k times to obtain p results that differs from $\bar{\lambda}$ more than ± 0.15 . A probability of such event could be estimated as $P = p/k$. However, if this probability is very small a large number of simulations k and computer time could be needed. In such a case one could obtain estimation by taking, for instance, a million random samples. If such a result does not appear, this indicates that the probability P of such an event is less than $P < 10^{-6}$. Such estimation has been performed for the set of data available. In fact, real probability is even less, because a fourth part of the 5767 protocols has to be chosen, rather than that from available 1527. 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

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

representing the percentage of those, who voted for Mr.

Lukashenko at each polling station and consider the distribution of the probability density Feller (1966) of this value. The distribution smoothed with the Gaussian kernel is shown in Fig. 1. As it is known, the probability density Feller (1966); Mathews and Walker (1964) characterizes the probability of obtaining a 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 $P = 0.43$. Consequently $\int_0^1 \rho(\lambda) d\lambda = 1$.

As one could see from Fig.1 (left panel), 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 interesting that an additional maximum 80% has been appeared besides to the expected maximum near 60%. To analyze this phenomenon, let us to construct

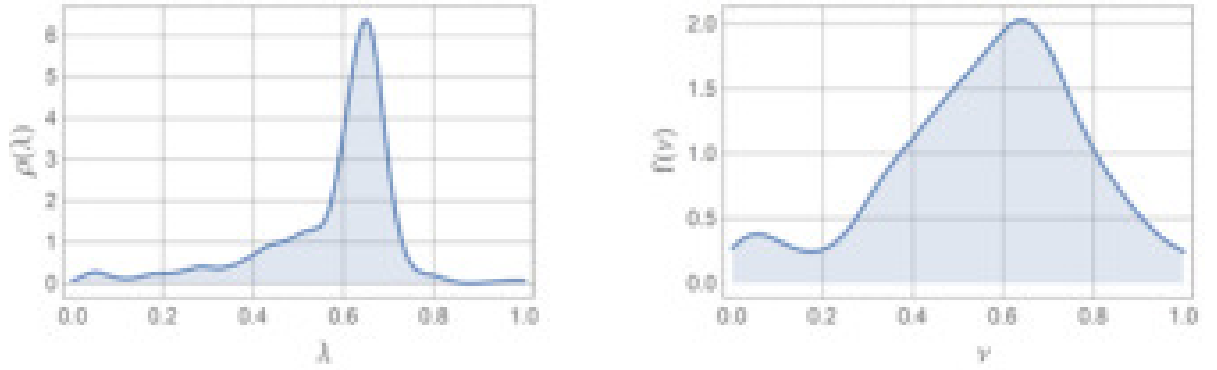


Figure 2: The probability density functions $\rho(\lambda)$ and $f(v)$ for city of Minsk.

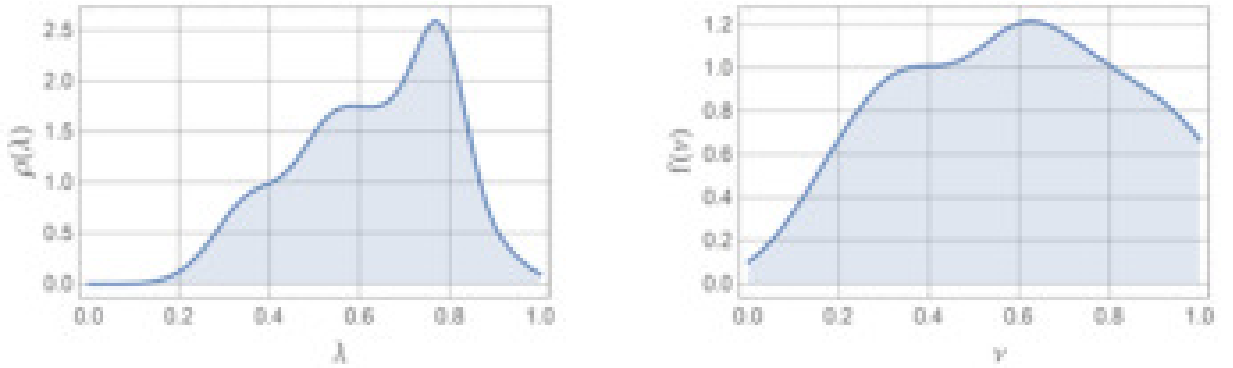


Figure 3: The probability density functions $\rho(\lambda)$ and $f(v)$ for Brest region.

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

$$v_i = M_i / \bar{N}. \quad (4)$$

Since the number of voters is different in the different polling stations, the distribution $f(v)$ should be broader than $\rho(\lambda)$, but still retains its key points. The distribution $f(v)$ is shown in right panel of Fig.1. In principle, v given by Eq. (4) could be greater than unity, because for large polling station a number of voters for Mr. Lukashenko could be greater than the total mean number \bar{N} of the voters coming to the polling station. That is why a plot of the function $f(v)$ ranges further than unity. However, we do not interest in the borders of the distribution but in the position of the peaks.

As one could see from Fig. 1, the peak at 0.6 remains, although somewhat shifted, while the peak at 0.8 has been disappeared traceless in the plot of the function $f(v)$. One the possible hypothesis is that the correlation in distribution $\rho(\lambda)$ arose due to artificial trimming of the percentage of those who voted for Lukashenko to 80% in some polling stations. An argumentation is as follow. Percentage λ_i given by Eq.(3) is evaluated quantity, whereas M_i and N_i are primary (base) quantities. If some structures (in a sense of maximums and minimums) exist in M_i and N_i than one could expect that the analogous structures (possibly smeared and shifted) would be in λ_i . However, if some corrections were performed in data by humans, some new structure in λ_i could appear, because humans usually think namely in terms of percentages. 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 $f(v)$ of the number of voters who voted

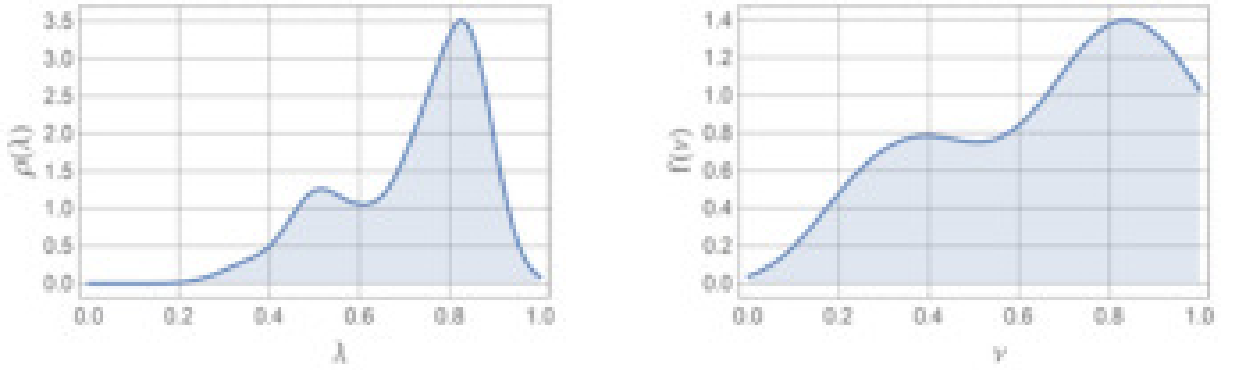


Figure 4: The probability density functions $\rho(\lambda)$ and $f(v)$ for Vitebsk region.

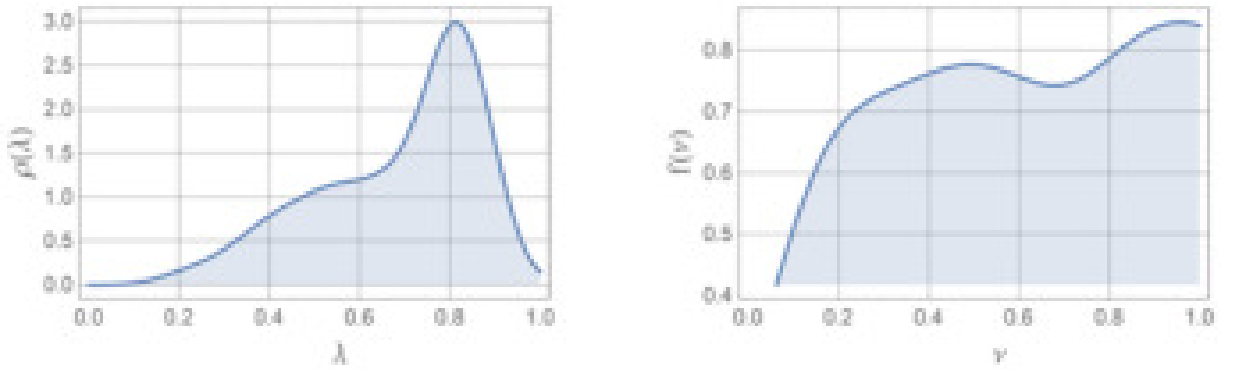


Figure 5: The probability density functions $\rho(\lambda)$ and $f(v)$ for Gomel region.

for Mr. Lukashenko. It seems, that comparison of the system of the peaks of distribution functions $\rho(\lambda)$ and $f(v)$ could be applied also for other elections systems such as United Kingdom or United States.

3. Analysis for Minsk and every of 6 regions of Belarus.

Belarus has six regions and capital in the city of Minsk. It is interesting to perform the previous analysis for every region and Minsk separately. Here we again will analyze the quantities N_i and M_i but they are ascribed to city of Minsk and every of the six Belarus regions.

3.1. City of Minsk

For Minsk more than half of the total amount of 731 protocols are available, namely, 454 protocols. Officially declared value over Minsk is $\bar{\lambda}_{official} = 0.645$. Calculated mean value is $\bar{\lambda} = 0.58$. Again, using Monte-Carlo simulation a probability was estimated to obtain deviation $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.065$ over a set of 273 protocols from 454. The number 273 was taken following the proportion $\frac{273}{454} \approx \frac{454}{731}$. An estimation for the probability of such deflection was obtained at a level of $P_0 < 10^{-6}$, where we denote Minsk city by the zero index and the regions will be denoted as $1, 2 \dots$.

The result of calculations of the probability distributions are shown in Fig. 2.

One could see rather well picture, because the peak in $\rho(\lambda)$ resembles a Gaussian one. Usually, when physicists measure some quantity they expect that it will have Gaussian distribution Mathews and Walker (1964). For elections

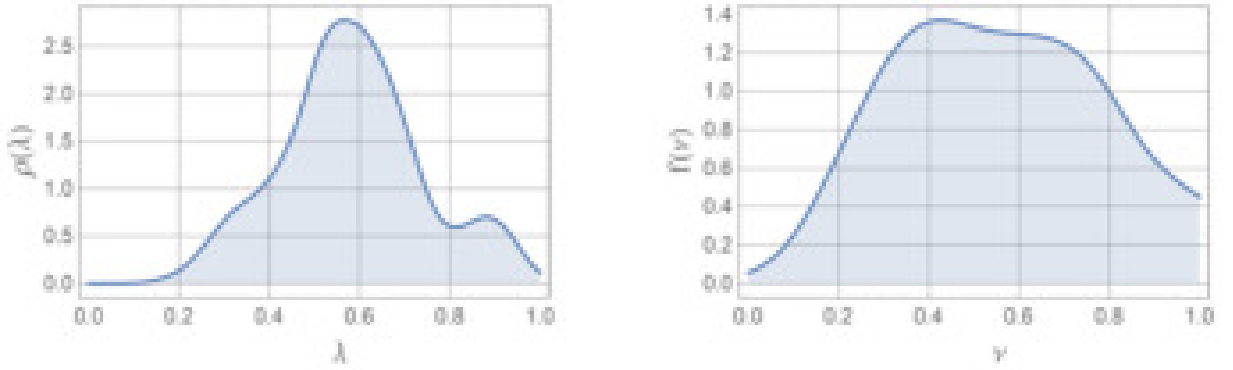


Figure 6: The probability density functions $\rho(\lambda)$ and $f(v)$ for Grodno region.

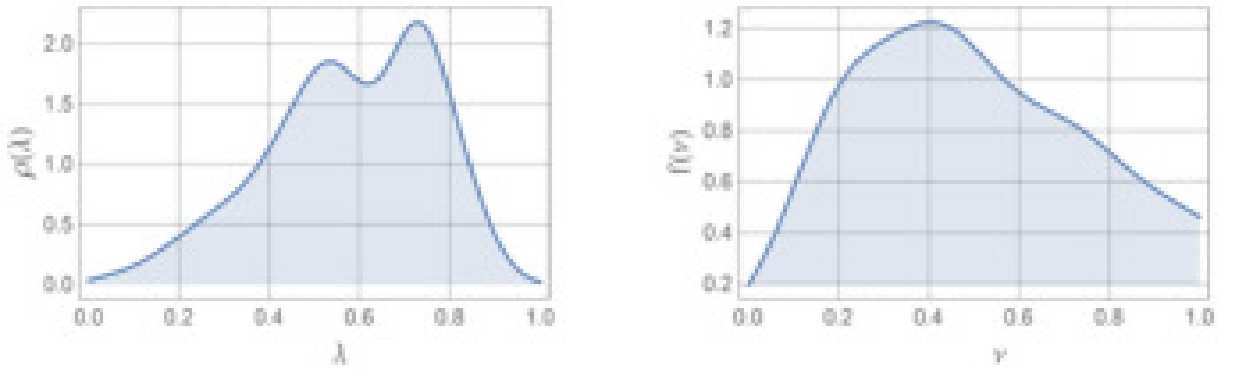


Figure 7: The probability density functions $\rho(\lambda)$ and $f(v)$ for Minsk region.

result one could refer to an ideal elector which with the probability p vote for Mr. Lukashenko and votes against with the probability $1 - p$. The probability distribution of the mean value over polling station consisting over voters n will have binomial distribution which for large n has to be close to Gaussian one Mathews and Walker (1964); Feller (1966). Certainly in reality there is no an ideal elector and one could not hope for this idealized picture Katz and King (1999). A generalization was considered by Rosenman and Viswanathan (2018) that there exist m groups having their own probability p_m to vote for some candidate and $1 - p_m$ against, but in this case distribution has also to be Gaussian if the number of groups m is large.

The distribution $f(v)$ has broadened peak with the approximately the same maximum position. It seems a reasonable, and if the official result had been closed to $\bar{\lambda} = 0.58$ one could hardly said something against the Minsk elections using official data. Nevertheless, it was noted that the Minsk results N_i and M_i have a number of suspicion peculiarities Gongalsky (2020); Zipunnikov (2020), but here we do not discuss them, because consider smothered distributions by the Gaussian kernel Cherkas (2020).

3.2. Brest region

For Brest region 196 protocols from the total 910 are available, i.e. approximately one fifth. Deflection from the official result is $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.8 - 0.63 = 0.19$. We have calculated probability to have such a deflection taking 42 protocols from 196. Samples of 42 protocols were used according to the proportion $\frac{42}{196} \approx \frac{196}{910}$. Estimated probability to obtain such a result turns out to be at $P_1 < 10^{-7}$ level.

Form of the distribution function is shown in Fig. 3. One could see a peak at 0.75 in $\rho(\lambda)$, which has disappeared

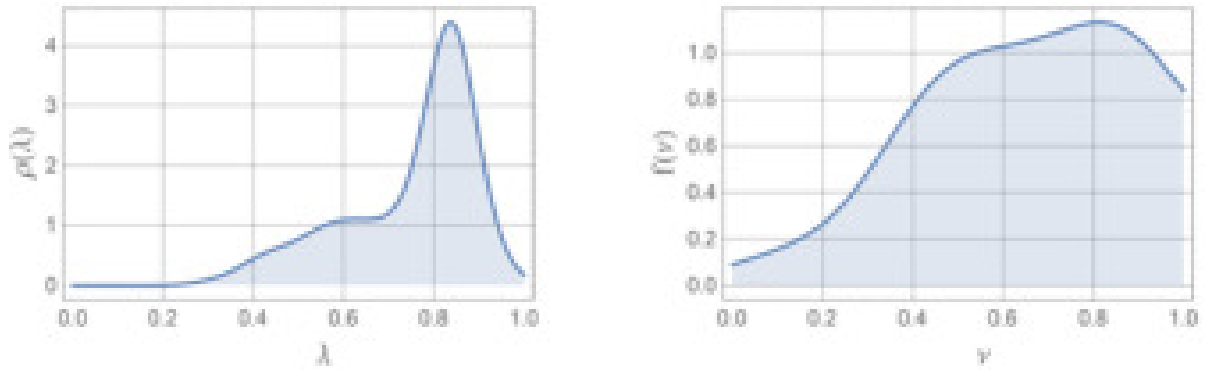


Figure 8: The probability density functions $\rho(\lambda)$ and $f(v)$ for Mogilev region.

in $f(v)$ distribution. Thus, it is artificial probably.

3.3. Vitebsk region

For this region, 145 protocols from 750 are available. Deflection of the official value from the calculated one is $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.83 - 0.72 = 0.11$. An estimation performed in the same manner as in the previous sections gives probability of this event as $P_2 \approx 3 \times 10^{-6}$.

Probability distribution pictures look rather good. Although the distribution $\rho(\lambda)$ has additional peak, its main peak looks nice. These peaks are conserved also at the function $\rho(v)$.

3.4. Gomel region

For this region 133 of 997 protocols are available. Deflection of the official mean value from the calculated one is $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.85 - 0.7 = 0.15$. Probability to obtain such a deflection is $P_3 \approx 8 \times 10^{-4}$. Although the deflection $\Delta\lambda = 0.15$ is considerable, the probability $P_3 \approx 8 \times 10^{-4}$ is not very small. It is because only few protocols available: approximately 1/7 of the total amount.

Probability distribution functions are shown in Fig.5. Very large smearing of the peak $\lambda \approx 0.8$ occurs when one transit from $\rho(\lambda)$ to $f(v)$.

3.5. Grodno region

For Grodno region we have 193 protocols from 656, i.e. approximately one third. The difference of the calculated and official mean values is $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.79 - 0.59 = 0.2$. Estimation of a probability to have such value is $P_4 < 10^{-6}$.

Form of the distribution function shown in Fig. 6 is relatively good, namely, the main peak of $\rho(\lambda)$ is smeared and shifted in $f(v)$. Second peak of $\rho(\lambda)$ at $\lambda \approx 0.9$ is absent in $f(v)$ and, probably, is of the artificial character.

3.6. Minsk region (excluding city of Minsk)

For Minsk region 294 protocols are available from the total amount of 992, that is, near 1/3. The difference between official and calculated mean values is $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.79 - 0.58 = 0.21$. An estimate of the probability of such deflection gives $P_5 < 10^{-6}$.

Probability distribution function $\rho(\lambda)$ shown in Fig. 7 has two peaks. One peak is at $\lambda \approx 0.7$. Possibly, this peak is artificial, because it has disappeared in function $f(v)$ shown in the right panel of Fig. 7.

3.7. Mogilev region

Only 112 protocols from 731 is available for Mogilev region, which is native land for Mr. Lukashenko. As an official result, so calculated value are high, but the difference still exists $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.88 - 0.765 = 0.115$. An estimation of the probability to obtain such as difference is $P_6 \approx 5 \times 10^{-4}$.

The peak at $\lambda \sim 0.9$ of the distribution function $\rho(\lambda)$ shown in Fig. 8 is conserved when one transits to function $f(v)$.

4. Conclusion

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

For Belarus generally

1) the average percentage of those who voted for Mr. Lukashenko in the available sample of 1527 protocols is about 63%, which is very different from the official 80% percentage. The probability of such an event is about $P \leq 10^{-6}$. Despite of the this small value this event could, in principle, take place. For instance, population of Belarus is 9.4 millions. If to give a lottery ticket with the probability of winning 10^{-6} to every Belarusian people, then there will be nine of lucky men.

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 of the 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%.

For Belarus regions.

4) The data for the Belarus regions gives more restrictive picture. For obtaining a possibility that the official results for the regions took a place one must multiply the probabilities for every region and city of Minsk. This gives a tiny value

$$P = P_0 P_1 P_2 P_3 P_4 P_5 P_6 < 10^{-36}. \quad (5)$$

In the Belarus elections one has in fact two stages i) picking and counting of the votes at the precinct stations ii) transferring them to the central electoral commission and declaring the result. Form of the distribution functions characterizes the first stage, whereas probability of the final result (5) refers to the second stage.

Form of the probability distribution functions looks good for Minsk city and relatively good for Vitebsk and Mogilev regions. Strange and suspicious look the curves for Brest, Gomel, Grodno and Minsk regions. This implies serious anomalies of these data indicating that they could not be used even for the statement that Mr. Lukashenko has received 60% of the votes. However, it is rather aesthetic than objecting criterium. As it has been mentioned, qualitative analysis of the voters level data has been performed by Gongalsky (2020); Zahorski (2020); Zipunnikov (2020).

As for juridical side, one could not say about a falsification of the elections, because in Belarus a falsification is a crime, whereas there are no criminal cases concerning the elections to the present time. Nevertheless, everybody has a rule to formulate his own opinion about these elections. Above numerical estimates and form of the distribution functions could be useful from this point of view.

References

- Charter97, 2020. Independent exit poll data: Tsikhanouskaya - 71.1%, lukashenka - 15.7%. <https://charter97.org/en/news/2020/8/9/388815/>.
- Cherkas, S.L., 2020. Election data and computer codes for "Statistical analysis of the presidential elections in Belarus in 2020". <https://doi.org/10.5061/dryad.d7wm37q0d>.
- Feller, W., 1966. An introduction to probability theory and its applications. volume I. John Wiley and Sons, Inc., New York-London-Sydney.
- Gongalsky, M., 2020. We have counted votes in belarus according to the protocols of 1 million voters. <https://twitter.com/MaximGongalsky/status/1293305204658536450>.
- Katz, J.N., King, G., 1999. A statistical model for multiparty electoral data. *American Political Science Review* 93, 15–32.
- Mathews, J., Walker, R.L., 1964. *Mathematical Methods of Physics*. Benjamin, New York.
- Rosenman, E., Viswanathan, N., 2018. Using poisson binomial glms to reveal voter preferences. [arXiv:1802.01053](https://arxiv.org/abs/1802.01053).
- Zahorski, A., 2020. Multilevel regression with poststratification for the national level viber/street poll on the 2020 presidential election in belarus. [arXiv:2009.06615](https://arxiv.org/abs/2009.06615).
- Zipunnikov, V., 2020. Statistical anomalies of the protocols from precinct commissions. <https://42.tut.by/697128>.
- ZUBR, 2020. Photographs of the 1527 protocol of precinct election commissions. <https://docs.google.com/spreadsheets/d/17aK3JxBTGTzULB0-YZG0F0hJwhuViH03/>.

Статистический анализ выборов президента Беларуси в 2020 г.

С.Л. Черкас

Как правило, избирателям хотелось бы иметь простую картину достоверности выборов. Тем не менее, бывают ситуации, когда доступна только частичная информация. Ниже предложены два метода сопоставления имеющейся информации с официальными результатами. Один из критериев заключается в сравнении среднего значения по имеющейся выборке с официальным средним значением. Выполнено моделирование методом Монте-Карло вероятности получить отличие среднего значения в некоторой случайной выборке от полного (официально) среднего. Другой метод представляет анализ характера особенностей функций распределения вероятностей, и заключается в сравнении формы функций распределения вероятностей для процента и количества избирателей голосовавших за А. Лукашенко на каждом участке. Последний метод скорее эстетический, чем количественный. Однако, он может быть применен к произвольным избирательным системам, например, Великобритании или Соединенных Штатов, позволяя иметь графическую картину результатов выборов.

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"

Введение

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 [1] 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 Беларуси существует система прямых одноуровневых президентских выборов. При этом есть некоторые особенности, одной из которых является то, что результат выборов на каждом избирательном участке должен быть вывешен на стене участка после выборов. В принципе, каждый может, например, сфотографировать вывешенный протокол.

Особенность состоит в том, что продолжительность экспозиции результата не оговаривается законодательством Беларуси. Эта продолжительность может быть ограниченной, вплоть до нескольких минут. В результате, не все участковые протоколы стали доступны после выборов в Беларуси, так что возникла необходимость анализа неполной информации.

В то же время, существуют неофициальные данные, состоящие в собранном платформой “Голос” [2] массиве бюллетеней индивидуальных избирателей, а также результаты опросов [1,3] которые указывают, что цифры в официальных участковых протоколах противоречат собранным “Голосом” бюллетеням и социальным опросам (сильно завышены в пользу А. Лукашенко). Тем не менее, ниже будут использованы только официальные данные: с одной стороны, официально вывешенные на стене избирательных участков участковые протоколы [4], с другой стороны, официально объявленный результат выборов по каждой области Беларуси и городу Минску.

Анализ результатов для Беларуси в целом.

В настоящее время доступны в открытой печати 1527 из 5767 общего количества протоколов участковых выборных комиссий (УИК) [4]. Представляется интересным, насколько достоверен результат выборов, объявленный как победа А. 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 г-на Лукашенко на этом же участке [5]. Как M_i так N_i можно рассматривать как случайные величины и вычислить следующие средние значения

$$\bar{N} = \frac{1}{1527} \sum_i^{1527} N_i, \quad \bar{M} = \frac{1}{1527} \sum_i^{1527} M_i, \quad (1)$$

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 г-на Лукашенко определяется как

$$\bar{\lambda} = \bar{M} / \bar{N} = \sum_i^{1527} M_i / \sum_i^{1527} N_i \approx 0.63, \quad (2)$$

и оказывается на уровне 63%. Согласно официальными результатам из анализа полного числа 5767 протоколов УИК количество голосовавших за г-на Лукашенко составляет около 80%. Таким образом, получается, что

случайная выборка примерно $\frac{1}{4}$ всех протоколов УИК дает результат отличающийся более чем на 15%. Оценим вероятность такого события методом Монте-Карло. Для этой цели случайным образом выберем $\frac{1}{4}$ протоколов из имеющейся выборки в 1527 протоколов и вычислим среднее. Данную процедуру будем повторять k раз для, пока не будет получено p результатов, которые отличаются от $\bar{\lambda}$ более чем на ± 0.15 . Вероятность такого события может быть оценена как $P = p/k$. Однако, если эта вероятность очень мала, то может потребоваться большое количество вычислений, и, соответственно, компьютерного времени. В таком случае можно получить оценку, взяв, например, миллион случайных выборок. Если такой результат не появляется, это указывает, что вероятность такого события меньше, чем $P < 10^{-6}$. На самом деле вероятность еще меньше, поскольку следовало бы выбирать четвертую часть из 5767 протоколов, а не из 1527. 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

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

представляющую собой процент проголосовавших за г-на Лукашенко на каждом участке, и рассмотреть распределение плотности вероятности $\rho(\lambda)$ данной величины [6,7]. Сглаженное с помощью Гауссова ядра распределение показано на Рис. 1. Как известно, плотность вероятности [6,7] 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 $P = 0.43$. Соответственно, $\int_0^1 \rho(\lambda) d\lambda = 1$.

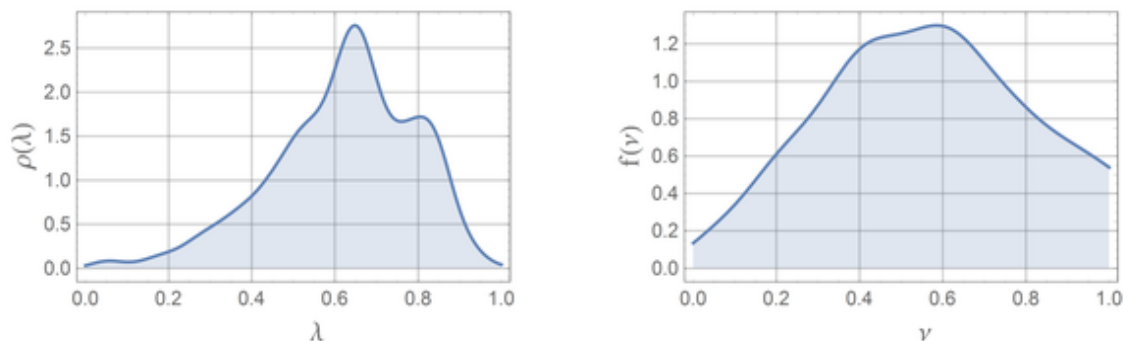


Рис. 1 Функция $\rho(\lambda)$ распределения плотности вероятности процента голосовавших за г-на Лукашенко. Функция $f(v)$ распределения плотности вероятности числа голосовавших за г-на Лукашенко нормированного на среднее число голосовавших на участке, см. (4). Подразумевается некоторый случайный избирательный участок.

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 г-на Лукашенко на каждом участке рассматривая нормированную случайную величину

$$v_i = M_i / \bar{N}. \quad (4)$$

Поскольку на разных участках число проголосовавших различно, то распределение $f(v)$ должно быть более размытым, чем $\rho(\lambda)$, однако, сохранять его ключевые моменты. Распределение $f(v)$ приведено на правой панели Рис. 1. В принципе, v_i может быть больше единицы, поскольку на больших участках число голосовавших за А. Лукашенко M_i может быть больше числа \bar{N} пришедших на участок в среднем, поэтому график $f(v)$ простирается дальше $v=1$. Однако нас интересуют не края распределения $f(v)$, а расположение пиков.

Как видно из Рис. 1 (правая панель), 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 процента проголосовавших за Лукашенко к 80. Дело в том, что процент λ_i (3) является

вычисляемой величиной, в то время как исходными (первичными) являются M_i и N_i . Если какие-либо структуры (т.е. максимумы) имеются в функциях распределения для M_i и N_i , то, можно ожидать, что примерно такая же структура (возможно размазанная или смещенная) будет в вычисляемой величине λ_i . Однако, если в данные вносились какие-то искусственные коррективы, причем в терминах процента, то в функции распределения вероятности для процента могут появиться некоторые новые структуры, поскольку, обычно, люди мыслят именно в терминах процента.

Разумеется, можно придумать и другие объяснения, почему пик на 80% есть в распределении $\rho(\lambda)$ плотности вероятности процента, но отсутствует в плотности вероятности $f(v)$ числа избирателей, голосовавших за А. Лукашенко. Ниже мы повторим проведенный анализ отдельно для каждой области из шести областей Беларуси и Минска.

Минск

Для Минска доступно 454 из протоколов участков из общего числа 731, т.е. больше половины. Официальное значение $\bar{\lambda}_{official} = 0.645$. Вычисленное по доступным протоколам $\bar{\lambda} = 0.58$. Описанным выше методом можно оценить вероятность возникновения отличия $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.065$ в выборке 273 протоколов из 454. Число 273 бралось на основании пропорции $\frac{273}{454} \approx \frac{454}{731}$. В результате приходим к следующей оценке $P_0 < 10^{-6}$, где далее город Минск будет обозначаться индексом 0, а области Беларуси индексами 1,2... и так далее.

Рассчитанные функции распределения плотности вероятности приведены на Рис. 3.

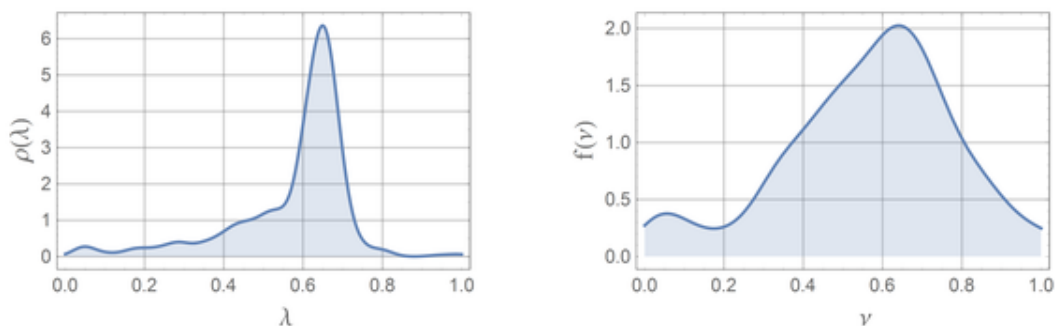


Рис. 2 Функции распределения плотности вероятности $\rho(\lambda)$ и $f(v)$ для Минска.

Представленная на Рис. 2 картина, выглядит замечательно, поскольку пик функции $\rho(\lambda)$ напоминает гауссов. Обычно, когда физики измеряют

какую-либо величину, они ожидают, что она будет иметь гауссово распределение [7]. Для результата выборов можно было бы сослаться на идеального выборщика, который с вероятностью p голосует за господина Лукашенко и с вероятностью $1-p$ голосует против. Тогда, распределение вероятности среднего значения по избирательному участку, состоящему из n избирателей, будет иметь биномиальное распределение, которое, для большого n , должно быть близко к Гауссову. Конечно, в действительности нельзя надеяться на такую идеализированную картину [8]. Например, более сложная картина состоит в том, что существует m групп избирателей, в которых избиратели голосуют за А. Лукашенко с вероятностью p_m и против с вероятностью $1-p_m$ [9]. Однако, и в такой сложной модели распределение должно быть близко к Гауссову если число групп велико m .

Распределение $f(\nu)$ имеет уширенный пик с примерно тем же положением максимума. Это представляется разумным. Таким образом, если бы официальный результат по Минску $\bar{\lambda}_{official}$ был бы близок к вычисленному значению $\bar{\lambda}$, то вряд ли что-нибудь можно было бы возразить против минских выборов, используя только официальные данные. Тем не менее, $\Delta\lambda$ велико и вероятность P_0 мала.

Кроме этого, ранее было отмечено [2,10], что минские результаты имеют ряд странностей, которые здесь не обсуждаются, поскольку рассматриваются усредненные по гауссову ядру данные [5], что позволяет проанализировать самый общий результат, а не частности (или особенности), которые могут иметь место в любых выборах.

Брестская область

Доступно 196 протоколов из 910 т.е. порядка $1/5$. Отклонение официального значения от вычисленного среднего $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.8 - 0.63 = 0.19$. Мы вычислили вероятность получить такое отклонение взяв 42 протокола из 196 согласно пропорции $\frac{42}{196} \approx \frac{196}{910}$. Найденная $P_1 < 10^{-7}$.

Функции плотности распределения вероятности показаны на Рис. 3. Пик на 0.8 исчез (или, возможно, сместился на 0.6) – скорее всего он искусственный

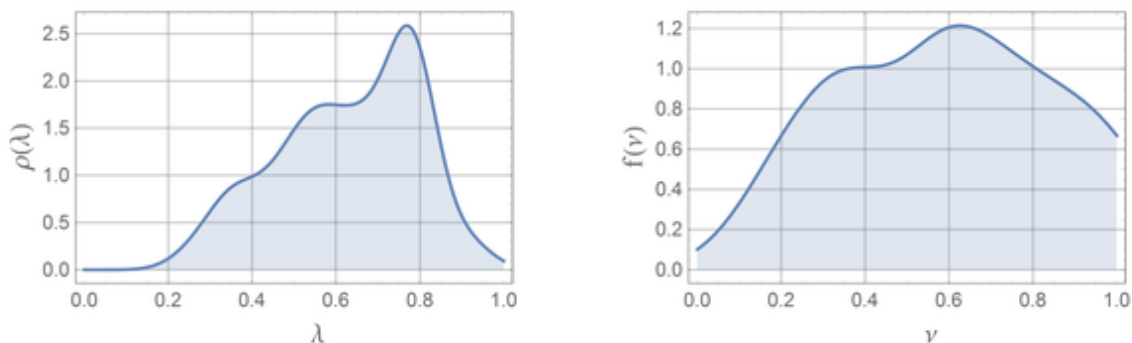


Рис. 3 Функции распределения плотности вероятности $\rho(\lambda)$ и $f(v)$ для Брестской области.

Витебская область

Имеется 145 протоколов из 750. Отклонение от официального значения $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.83 - 0.72 = 0.11$. Вероятность такого отклонения $P_2 \approx 3 \times 10^{-6}$.

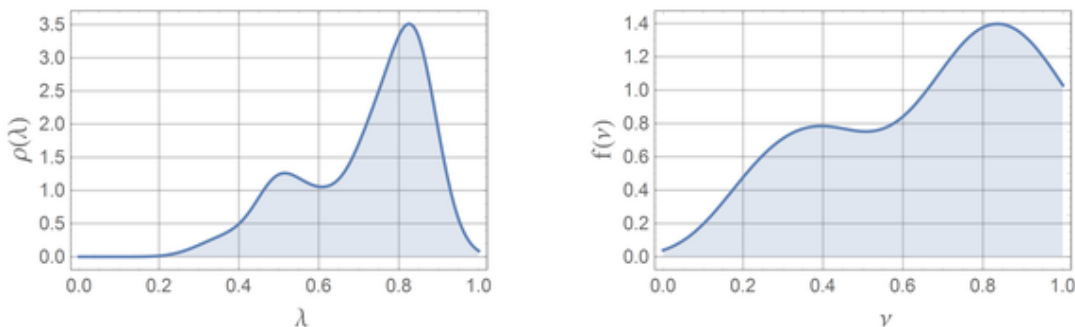


Рис. 4 Функции распределения плотности вероятности $\rho(\lambda)$ и $f(v)$ для Витебской области.

Функции распределения плотности вероятности, показанные на Рис. 4, смотрятся, довольно хорошо, поскольку, хотя распределение $f(v)$ расплылось, но структура пиков сохранена.

Гомельская область

Имеется 133 протокола из 997. Разница официального и вычисленного средних $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.85 - 0.7 = 0.15$. Вероятность получить такое $\Delta\lambda = 0.15$ оценивается как $P_3 \approx 8 \times 10^{-4}$.

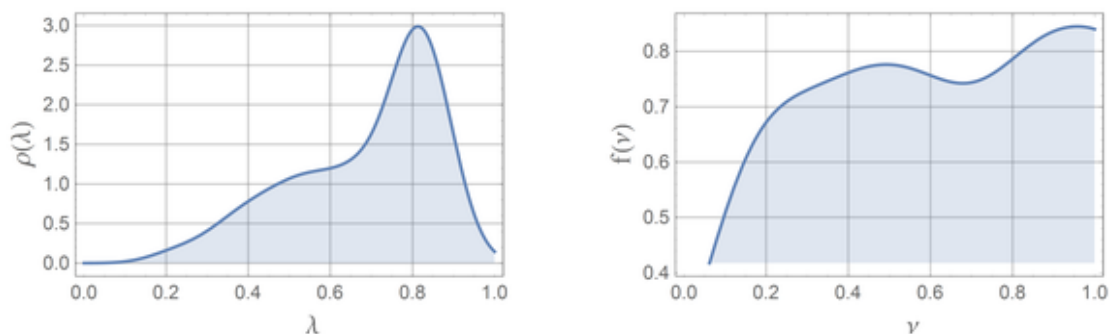


Рис. 5 Функции распределения плотности вероятности $\rho(\lambda)$ и $f(v)$ для Гомельской области.

Функции распределения вероятности изображены на Рис. 5. Можно заметить, что пик при $\lambda \approx 0.8$ слишком сильно расплылся и сместился при переходе от $\rho(\lambda)$ к $f(v)$.

Гродненская область

Для Гродненской области имеем 193 из 656 протоколов, и разницу средних $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.79 - 0.59 = 0.2$. Вероятность получить такое $\Delta\lambda$ оценивается как $P_4 < 10^{-6}$.

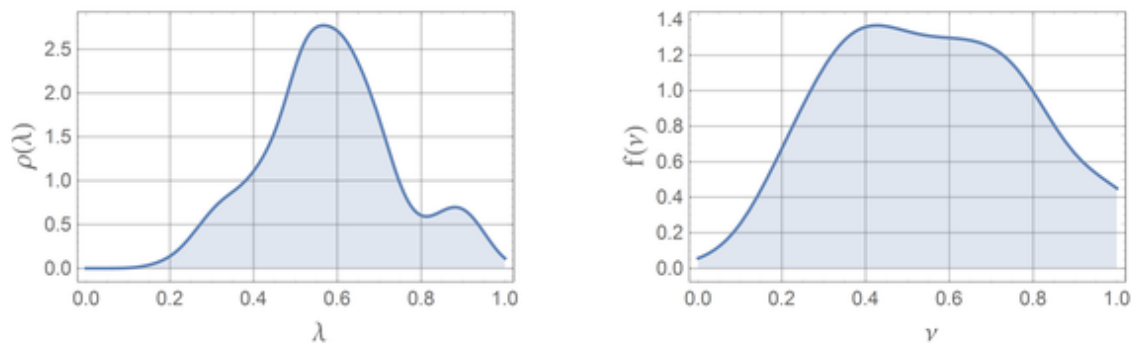


Рис. 6 Функции распределения плотности вероятности $\rho(\lambda)$ и $f(v)$ для Гродненской области.

Второй пик в $\rho(\lambda)$ при $\lambda \approx 0.9$ отсутствует в $f(v)$, т.е. скорее всего искусственный.

Минская область

Доступно 294 из 992 протоколов - около 1/3. Различие средних значений $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.79 - 0.58 = 0.21$. Оценка вероятности $P_5 < 10^{-6}$.

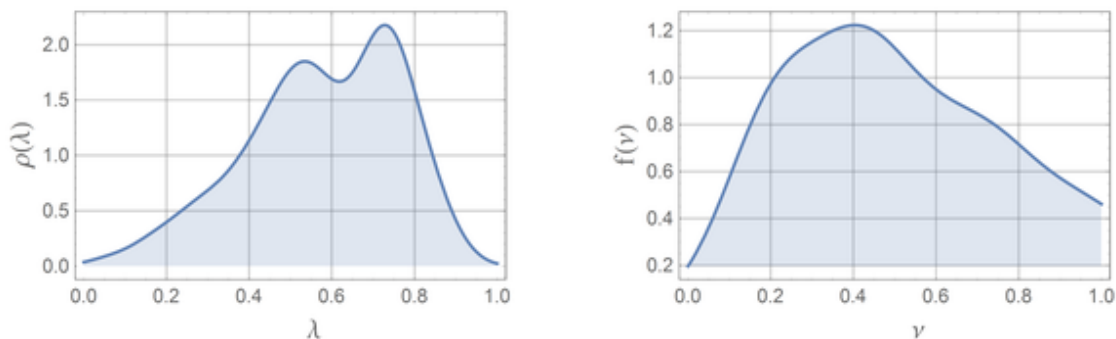


Рис. 7 Функции распределения плотности вероятности $\rho(\lambda)$ и $f(v)$ для Минской области.

Функция распределения вероятностей $\rho(\lambda)$ имеет два пика. Скорее всего, пик в $\rho(\lambda)$ на $\lambda \approx 0.7$ искусственный, поскольку исчез в $f(v)$.

Могилевская область

Имеется 112 протоколов из 731. Разница результатов $\Delta\lambda = \bar{\lambda}_{official} - \bar{\lambda} = 0.88 - 0.765 = 0.115$. Оценка вероятности такой разницы $P_g \approx 5 \times 10^{-4}$.

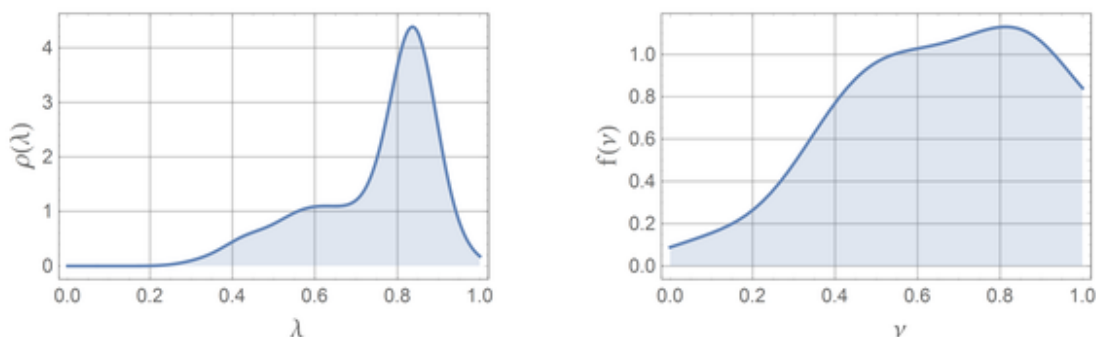


Рис. 8 Функции распределения плотности вероятности $\rho(\lambda)$ и $f(v)$ для Могилевской области.

Интересно, что функция распределения довольно хорошая, так как пик $\lambda \sim 0.85$ сохраняется примерно на том же месте при переходе к $f(v)$.

Заключение

В результате отметим следующие особенности проведенного статистического анализа для расчетов **по Беларуси в целом**:

1) средний процент проголосовавших за г-на Лукашенко в выборке доступных 1527 протоколов порядка 63%, что сильно отличается от

официальных 80% процентов. Вероятность такого события составляет меньше 10^{-6}

2) Распределение процента проголосовавших за г-на 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 г-на Лукашенко на каждом участке. Одной из гипотез данного явления может быть искусственное подравнивание процента проголосовавших за г-на 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%.

Таким образом, из анализа Беларуси в целом, для вероятности, что официальные данные действительно имели место, получается оценка $P < 10^{-6}$. Несмотря на малую величину, такое событие вполне могло бы произойти. Например, население Беларуси составляет 9.4 миллионов. Если каждому белорусу раздать по лотерейному билету с вероятностью выигрыша 10^{-6} , то среди жителей Беларуси оказалось бы около 9 счастливых. Разумеется, было бы интересно убедиться, что это действительно так. Для этого необходимо иметь недостающие протоколы участков, которые в принципе не являются секретной информацией и на основании закона о праве на информацию должны были бы быть предоставлены гражданам. Тем не менее, избирательная комиссия не нашла возможности это сделать.

4) Из анализа **по областям** следует более жесткий результат – для получения достоверности официальных результатов выборов вероятности по областям и Минску должны перемножаться, что приводит к

$$P = P_0 P_1 P_2 P_3 P_4 P_5 P_6 < 10^{-36} \quad (5)$$

т.е. к практически нулевой вероятности достоверности официальных данных. В выборах мы имеем фактически два этапа: i) сбор и подсчет голосов на участках ii) передача этих данных в избирательную комиссию и объявление официального результата. Рассмотренная вероятность достоверности окончательных результатов выборов (5) характеризуют второй этапа. Что касается первого этапа, то, как метод анализа можно рассматривать форму приведенных выше кривых распределения вероятности процента проголосовавших за А. Лукашенко $\rho(\lambda)$ и нормированного числа избирателей, проголосовавших за А. Лукашенко $f(v)$. В этом смысле довольно странно и подозрительно выглядят кривые для Брестской,

Гомельской, Гродненской, и Минской областей. Это явно показывает, что и в самих данных по участкам в указанных областях имеются серьезные аномалии и эти данные нельзя было бы использовать, даже чтобы утверждать, что А. Лукашенко набрал около 60% голосов избирателей.

Однако, это скорее эстетический критерий. Как уже упоминалось, количественный анализ был проведен в работах [1-3].

Что касается юридической стороны дела, то, разумеется, нельзя говорить о фальсификации выборов, поскольку согласно белорусским законам фальсификация выборов является уголовным преступлением. На данный момент не было открыто каких-либо уголовных дел на эту тему. Тем не менее, каждый вправе сформировать свое мнение о достоверности рассматриваемых выборов. Приведенные оценки вероятностей, а также форма функций распределения вероятностей, основанные только на официальных данных могут быть полезными с этой точки зрения, так же как анализ неофициальных данных [2] и результаты социальных опросов [1,3].

СПИСОК ЛИТЕРАТУРЫ

1. *A. Zahorski*, Multilevel regression with poststratification for the national level Viber/Street poll on the 2020 presidential election in Belarus, arXiv: 2009.06615
2. *M. Gongalsky*. We have counted votes in Belarus according to the protocols of 1 million voters, 2020a. URL <https://www.facebook.com/maxim.gongalsky/posts/10221361216434274>. [Online (in Russian); accessed 20-August-2020].
3. *Charter97*. Independent Exit Poll Data: Tsikhanouskaya - 71,1%, Lukashenka - 15,7%, 2020. URL <https://charter97.org/en/news/2020/8/9/388815/>. [Online (in English); accessed 20-August-2020].
4. Фотографии 1527 протоколов участковых избирательных комиссий / <https://docs.google.com/spreadsheets/d/17aK3JxBTGtzULB0-YZGOF0hJwhuViHO3/>
5. *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>
6. *W. Feller*, An introduction to probability theory and its applications. Vol. I, New York-London-Sydney, John Wiley and Sons, Inc., 1966, 626 pp.

7. Дж. Мэтьюз, Р. Уокер. Математические методы физики, Мир, 1977.
8. J. N. Katz, G. King. A Statistical Model for Multiparty Electoral Data
American Political Science Review Vol. 93, No. 1 P.15-32
9. E. Rosenman, N. Viswanathan. Using Poisson Binomial GLMs to Reveal
Voter Preferences. arXiv:1802.01053 [stat.ML]
10. V. Zipunnikov, 2020. Статистические аномалии протоколов комиссий.
URL <https://42.tut.by/697128>. [Online (in
Russian); accessed 20-August-2020]