Probability Experts May Decide Vote in Pennsylvania

By PETER PASSELL

Fraud, virtually everyone agrees, should not be allowed to decide the election of a public official. But how is a court to determine whether the fraud was decisive? How certain must it be to take the extreme step of seating the nominal loser?

That problem faces Judge Clarence Newcomer of Federal District Court in Philadelphia, who will soon decide whether to order a new election in Pennsylvania's Second State Senatorial District in Philadelphia or declare the losing candidate, Bruce Marks, a Republican, to be the winner last November. To help, Judge Newcomer has hired a consultant from Princeton University who is an expert in the mathematics of probability.

The analysis of the expert, Orley Ashenfelter, an econometrician, along with that of statisticians hired by the Republicans and Democrats, may or may not sway the judge.

At the very least, though, the work of the statisticians will give the judge a rigorous basis for arguing about the probabilities that the election was stolen, and the case offers a fascinating glimpse at the no-man's-land where the law's wish for certainty meets a reality that many events can be explained only in terms of chance.

Absentee Ballot Problem

In the special election to fill a Senate vacancy, Mr. Marks received 19,691 votes on the voting machines to the 19,127 for William Stinson, his Democratic opponent. But in absentee ballots, Mr. Stinson outpolled Mr. Marks by 1,391 to 366, giving him a victory by 461 votes. And in spite of charges of fraud by the Republicans, Mr. Stinson, a former assistant Deputy Mayor of Philadelphia, was sworn into office, temporarily giving the Democrats exactly half the seats in the State Senate.

In February, Judge Newcomer ruled that many of the absentee ballots had been improperly obtained and processed by the Democratic-controlled Philadelphia County Board of Elections as part of a "massive scheme" to control the outcome of the election. In some cases, for example, registered Democrats were given absentee ballots and told that they need not go to the polls to vote. In other cases they were specifically guided to vote Democratic. Judge Newcomer declared Mr. Marks, a former aide to Senator Arlen Specter, to be the winner.

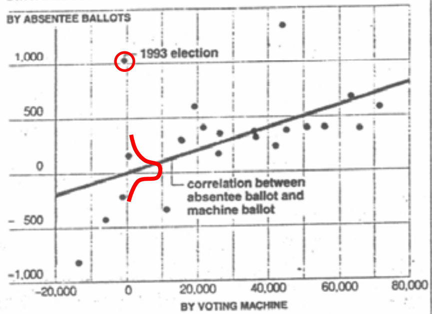
Then, on March 12, a Federal appellate court ruled on the case, letting the decision to void the results of the election stand, but ordering Judge Newcomer to reconsider his decision to seat Mr. Marks rather than to call a new election. Shortly thereafter, Mr. Stinson and two Democratic campaign workers were charged with election fraud by the Pennsylvania Attorney General, a Republican.

STATISTICS

Looking for Fraud in Philadelphia

The dots show the margin of victory or defeat for the Democrats, by both absentee ballot and voting-machine ballot, in 22 elections in Philadelphia's senatorial districts over the last decade. The plot of the absentee and machine votes shows a rough relationship between the two: the bigger the Democratic majority on machine votes, the larger their margin on absentee ballots. The 1993 election and a few others produced results that are far from typical. The probability that the unusual results of the 1993 election were simply caused by random variations in voting patterns is just 6%.

Difference between Democratic and Republican tallies



Sources: Orley Ashenfelter, Princeton University

Looking for answers in years of data on absentee ballots and those on machines.

Analysis of Votes

How might statistical theory be used to determine the probability of the Democrats' stealing the election?

The Question

Professor Ashenfelter's approach is to analyze the relationship between machine-counted votes and absentee ballots in state senatorial contests in Philadelphia over the last decade.

There have been 22 such contests since 1982, including the disputed 1993 election. And using the standard statistical technique of regression analysis, Professor Ashenfelter found that the difference between the Democratic and Republican tallies in the machine-based vote has been a good indicator of the difference between the two parties' absentee vote.

Assuming this relationship in the 21 previous elections had held in the most recent, Professor Ashenfelter estimates that the Republican's 564-vote edge on the machines should have led to a 133-vote advantage in absentee ballots. In fact, the Democrat ended up with a majority of 1,025 absentee votes.

Chance and Statistics

But as Professor Ashenfelter readily acknowledges, the predicted relationship between machine and absentee balloting is statistical in nature and dependent on the stability of the historical relationship. There is some chance that random variations alone could explain a 1,158-vote swing in the 1993 contest — the difference between the predicted 133-vote Republican advantage and the 1,025-Democratic edge that was reported.

More to the point, there is some larger probability that chance alone would lead to a sufficiently large Democratic edge on the absentee ballots to overcome the Republican margin on the machine balloting. And the probability of such a swing of 697 votes from the expected result, Professor Ashenfelter calculates, was about 6 percent.

Putting it another way, if past elections are a reliable guide to current voting behavior, there is a 94 percent chance that irregularities in the absentee ballots, not chance alone, swung the election to the Democrat, Professor Ashenfelter concludes.

The Answer

P = 94 %

Paul Shaman, a professor of statistics at the Wharton School at University of Pennsylvania who is the statistical consultant retained by the Democrats, exploits the limits in Professor Ashenfelter's reasoning. Relationships between machine and absentee voting that held in the past, he argues, need not hold in the present. Could not the difference, he asks, be explained by Mr. Stinson's "engaging in aggressive efforts to obtain absentee votes?"

Looking at the Findings

Elizabeth Holtzman, New York City's former Controller who has had a practical education in the vagaries of election statistics, agrees. With the exceptionally low turnout in the State Senate contest — no surprise in a special election — she believes it is unfair to project voter sentiment from historical relationships.

Brian Sullivan, the statistical consultant working for the Republicans who is an economist with the Center for Forensic Economic Studies in Philadelphia, responds to some of Professor Shaman's and Ms. Holtzman's objections. He analyzed the findings of a post-election poll of registered voters who had applied for absentee ballots. From the survey responses that he judged most reliable, he estimated that 84 percent of the absentee ballots had been illegally solicited or cast. By this reckoning, even if every legitimate absentee vote had gone to the Democrat, the Republican would still have won a majority in the overall vote.

Professor Shaman questions the quality of the survey, which was conducted by Republican campaign workers without supervision by neutral professionals. He notes, among other things, that the survey implies a total of 1,395 absentee ballots were illegitimate — more than the total absentee ballot vote for Mr. Stinson.

Polluted by Irregularities

Where does this leave an impartial observer? There is certainly a widespread impression that the election was badly polluted by irregularities. Were Mr. Stinson convicted of election fraud he would not be eligible to serve in the Senate, even if an analysis later found that the fraud was insufficient to change the election result. But those observations are of little help in deciding whether to seat Mr. Marks or call another election — an election that would be won by the Democratic candidate if registered voters cast their ballots along party lines.

Lowell Finley, a partner in the San Francisco firm of Altshuler, Berzon who specializes in election law, thinks the burden of choice should rest on those who would declare a winner in the absence of certainty. "In a democracy voters ought to have the last word," he argues, "not judges."

Practical Uses of χ^2

- To infer σ from S_x
- To establish a rejection criterion
- To compare a sample to an assumed population

- $\chi^2 = f(N) >> f(\nu)$. Thus, there are an infinite number of χ^2 distributions (like for Student's *t*).
- χ^2 is the fractional difference between measured and expected frequencies of occurrence.

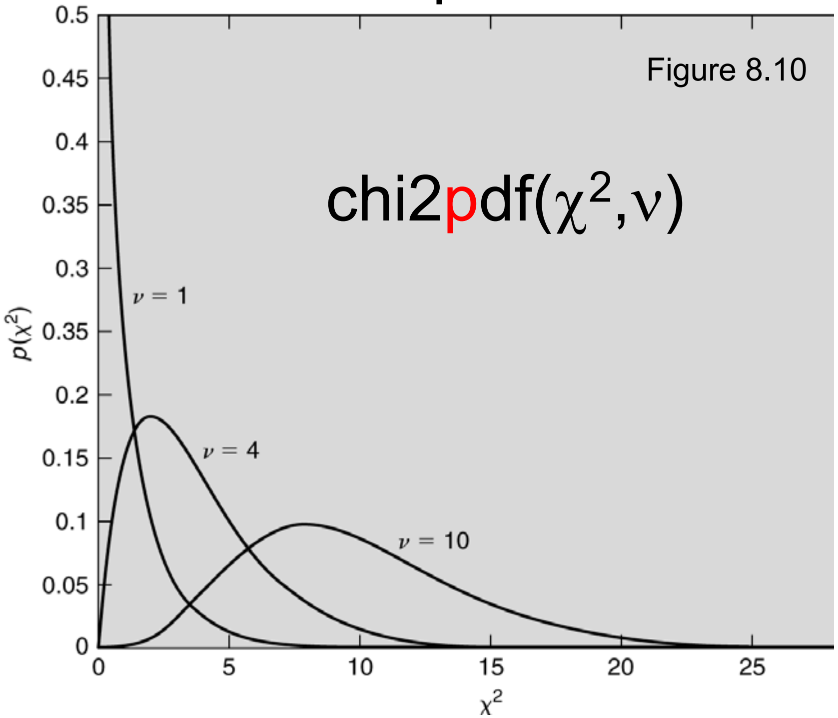
The χ^2 Variable

- The random variable, χ^2, is defined as

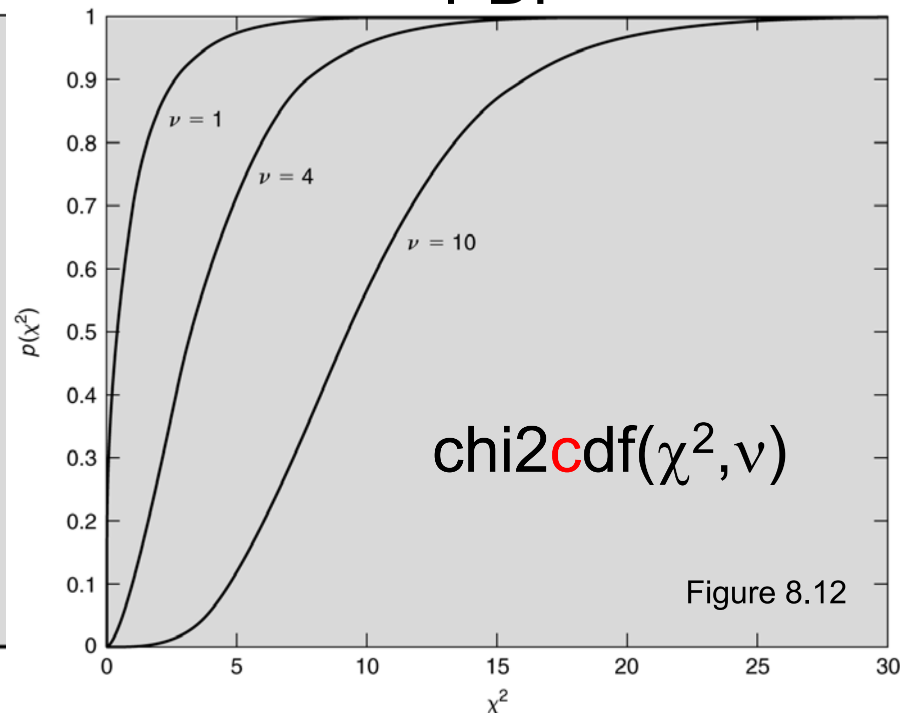
- Thus, as N$\rightarrow\infty$, $S_x^2 \rightarrow \sigma^2 \Rightarrow \chi^2 \rightarrow \nu$

The χ^2 Distribution

pdf



PDF



Determine Pr[$\chi^2 \leq 10$] for N = 11:

Determine Pr[$\chi^2 \leq 10$] for N = 5:

Determine χ^2 for P = 50 % and N = 5:

χ^2 Probability Density Function

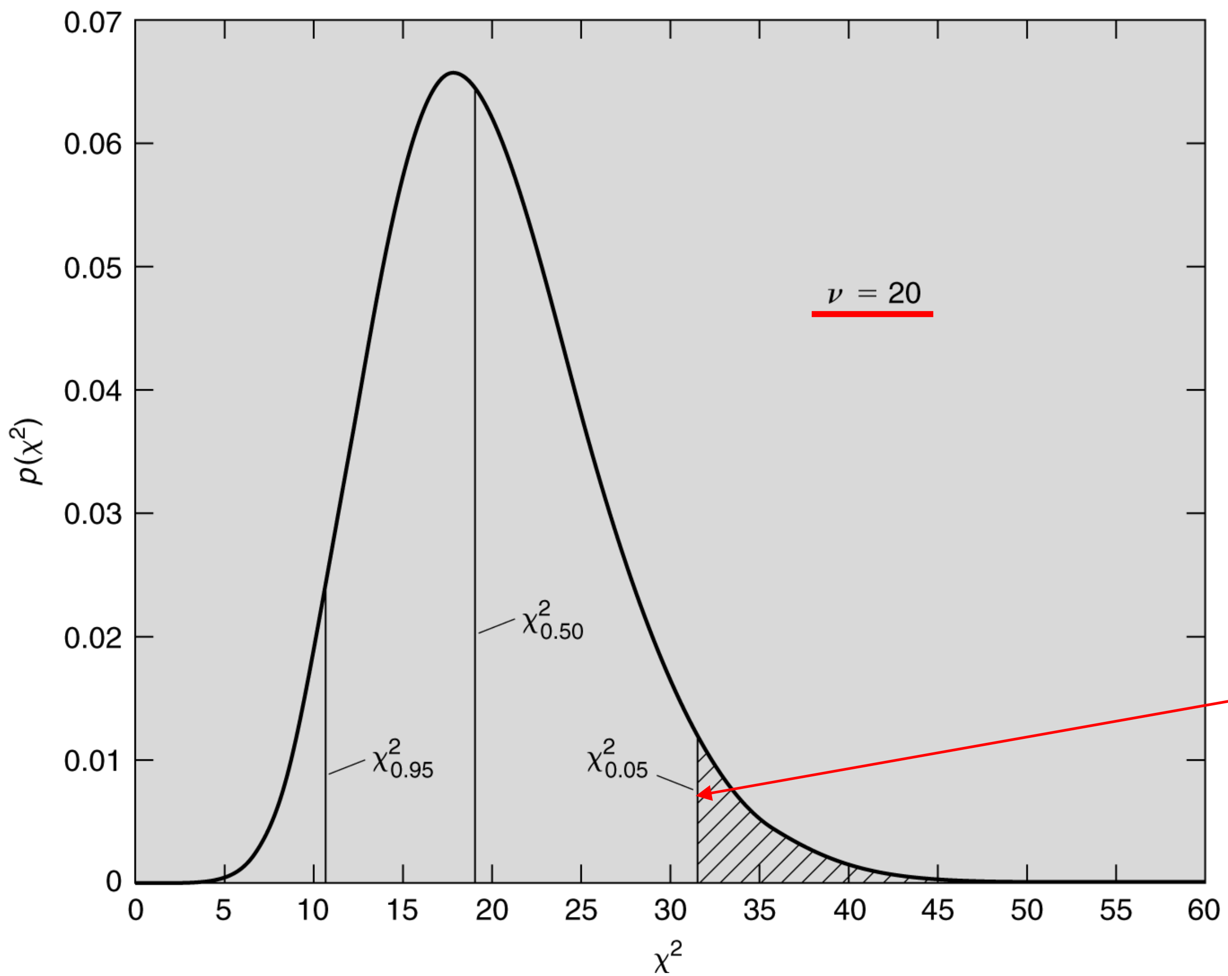


Figure 8.11

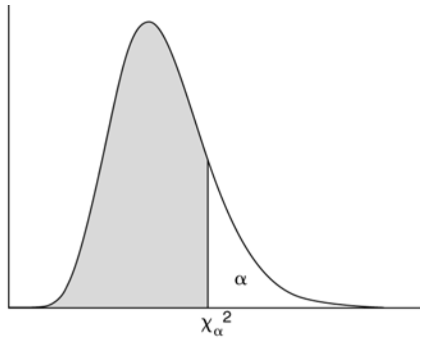
α

level of significance

probability that differences causing the value of $\chi^2{}_\alpha$ result solely from random effects

χ^2 Table

ν	$\chi^2_{0.99}$	$\chi^2_{0.975}$	$\chi^2_{0.95}$	$\chi^2_{0.90}$	$\chi^2_{0.50}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$	$\chi^2_{0.01}$
1	0.000	0.000	0.000	0.016	0.455	3.84	5.02	6.63
2	0.020	0.051	0.103	0.211	1.39	5.99	7.38	9.21
3	0.115	0.216	0.352	0.584	2.37	7.81	9.35	11.3
4	0.297	0.484	0.711	1.06	3.36	9.49	11.1	13.3
5	0.554	0.831	1.15	1.61	4.35	11.1	12.8	15.1
6	0.872	1.24	1.64	2.20	5.35	12.6	14.4	16.8
7	1.24	1.69	2.17	2.83	6.35	14.1	16.0	18.5
8	1.65	2.18	2.73	3.49	7.34	15.5	17.5	20.1
9	2.09	2.70	3.33	4.17	8.34	16.9	19.0	21.7
10	2.56	3.25	3.94	4.78	9.34	18.3	20.5	23.2
11	3.05	3.82	4.57	5.58	10.3	19.7	21.9	24.7
12	3.57	4.40	5.23	6.30	11.3	21.0	23.3	26.2
13	4.11	5.01	5.89	7.04	12.3	22.4	24.7	27.7
14	4.66	5.63	6.57	7.79	13.3	23.7	26.1	29.1
15	5.23	6.26	7.26	8.55	14.3	25.0	27.5	30.6
16	5.81	6.91	7.96	9.31	15.3	26.3	28.8	32.0
17	6.41	7.56	8.67	10.1	16.3	27.6	30.2	33.4
18	7.01	8.23	9.39	10.9	17.3	28.9	31.5	34.8
19	7.63	8.91	10.1	11.7	18.3	30.1	32.9	36.2
20	8.26	9.59	10.9	12.4	19.3	31.4	34.2	37.6
30	15.0	16.8	18.5	20.6	29.3	43.8	47.0	50.9
40	22.2	24.4	26.5	29.1	39.3	55.8	59.3	63.7
50	29.7	32.4	34.8	37.7	49.3	67.5	71.4	76.2
60	37.5	40.5	43.2	46.5	59.3	79.1	83.3	88.4
70	45.4	48.8	51.7	55.3	69.3	90.5	95.0	100.4
80	53.5	57.2	60.4	64.3	79.3	101.9	106.6	112.3
90	61.8	65.6	69.1	73.3	89.3	113.1	118.1	124.1
100	70.1	74.2	77.9	82.4	99.3	124.3	129.6	135.8



For N = 13, find α when χ^2 = 21.0

For P = 5 %, find χ^2 if N = 20

Table 8.8

In-Class Example

A property sales group claims that they are 95 % confident that the standard deviation in the horizontal displacement of a floor in their new ‘safe’ apartment during an earthquake will be less than 6 inches. The data that they use to support their claim consists of a sample size of 21 and a measured standard deviation of 5 inches. Does their data support their claim?

In-Class Example

ν	$\chi^2_{0.99}$	$\chi^2_{0.975}$	$\chi^2_{0.95}$	$\chi^2_{0.90}$	$\chi^2_{0.50}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$	$\chi^2_{0.01}$
1	0.000	0.000	0.000	0.016	0.455	3.84	5.02	6.63
2	0.020	0.051	0.103	0.211	1.39	5.99	7.38	9.21
3	0.115	0.216	0.352	0.584	2.37	7.81	9.35	11.3
4	0.297	0.484	0.711	1.06	3.36	9.49	11.1	13.3
5	0.554	0.831	1.15	1.61	4.35	11.1	12.8	15.1
6	0.872	1.24	1.64	2.20	5.35	12.6	14.4	16.8
7	1.24	1.69	2.17	2.83	6.35	14.1	16.0	18.5
8	1.65	2.18	2.73	3.49	7.34	15.5	17.5	20.1
9	2.09	2.70	3.33	4.17	8.34	16.9	19.0	21.7
10	2.56	3.25	3.94	4.78	9.34	18.3	20.5	23.2
11	3.05	3.82	4.57	5.58	10.3	19.7	21.9	24.7
12	3.57	4.40	5.23	6.30	11.3	21.0	23.3	26.2
13	4.11	5.01	5.89	7.04	12.3	22.4	24.7	27.7
14	4.66	5.63	6.57	7.79	13.3	23.7	26.1	29.1
15	5.23	6.26	7.26	8.55	14.3	25.0	27.5	30.6
16	5.81	6.91	7.96	9.31	15.3	26.3	28.8	32.0
17	6.41	7.56	8.67	10.1	16.3	27.6	30.2	33.4
18	7.01	8.23	9.39	10.9	17.3	28.9	31.5	34.8
19	7.63	8.91	10.1	11.7	18.3	30.1	32.9	36.2
20	8.26	9.59	10.9	12.4	19.3	31.4	34.2	37.6
30	15.0	16.8	18.5	20.6	29.3	43.8	47.0	50.9
40	22.2	24.4	26.5	29.1	39.3	55.8	59.3	63.7
50	29.7	32.4	34.8	37.7	49.3	67.5	71.4	76.2
60	37.5	40.5	43.2	46.5	59.3	79.1	83.3	88.4
70	45.4	48.8	51.7	55.3	69.3	90.5	95.0	100.4
80	53.5	57.2	60.4	64.3	79.3	101.9	106.6	112.3
90	61.8	65.6	69.1	73.3	89.3	113.1	118.1	124.1
100	70.1	74.2	77.9	82.4	99.3	124.3	129.6	135.8