Apportion 2020: Montana vs. New York

Ian McDonald

12/28/2020

https://www.census.gov/prod/cen2010/briefs/c2010br-08.pdf

https://www.brookings.edu/research/dividing-the-house-why-congress-should-reinstate-an-old-reapportionment-formula/

the js calculator is here: https://isr.umich.edu/apportionment-calculator-for-us-census/

https://www.census.gov/programs-surveys/popest/data/tables.html

Abstract: New York will certainly lose at least one seat coming out of the 2020 census, but it could lose two. Will the apportionment method rob New York of its 26th seat in 2020? Will Montana unfairly receive a second seat? Thanks to the measurable bias for smaller states in the House apportionment method adopted in 1941, perhaps they will.

Today, reports appeared about delays in final 2020 Census apportionment for the US House, which also changes state distribution of electors for the Electoral College. New population estimates were published on December 22, and they project New York will lose two of its current 27 seats. Based on 2019 estimates, New York was projected to lose one seat, but not two.

Once the census determines the country's population, apportionment should be simple, Proportionality is the only requirement, aside from a guarantee of one representative. But rounding introduces controversy and complexity. The current process was enacted in 1941, and seems linked to a modest but consistent small state bias.

Consider each state's share of the vote as

$$Quota_{state} = \frac{StatePopulation}{USPopulation} \times 435$$

a result we call its "quota". This number includes a fractional remainder, and any reliable apportionment scheme will assign the whole number immediately above or below the quota. Using last week's population estimate, New York's seat quota would be:

$$\text{New York Quota} = \frac{19,336,776}{328,771,307} \times 435 \approx 25.59$$

$$\text{Montana Quota} = \frac{1,080,577}{328,771,307} \times 435 \approx 1.430$$

New York's quota rounds to 26, but the current apportionment formula somehow assigns only 25 seats. Montana's quota rounds to 1, but show the formula assigns two seats?

Webster, Huntington and Hill, and Method of Equal Proportions

The choice of an apportionment rounding method can potentially favor big states or small ones. The question dates back to the 1790's, and the important solutions were associated with famous statesmen, including Hamilton, Jefferson, John Quincy Adams, and Daniel Webster. In the decades before the 1940's, Congress used a formula designed by Daniel Webster based on earlier work by Hamilton.[^1] [Note about the lawsuit and the Dean Method]

The Webster formula applies the intuition that the allocation should round up when the remainder is greater than .5, and round down if the remainder is less. Suppose three states divide fifty seats:

[three state example]

The states simply round up and from remainders of .5, and the assignments add up to 435. If the remainders distributed differently, however, the numbrer of seats may add to a number greater or less than 435. If so, we adjust the denominator so that some assignments are added or subtracted until the sum hits 435.

[three state example with adjustments]

The Webster method produced hypothetical paradoxes and quirks that were apparent when the number of seats changed from census to census, and was replaced in 1941 by a design by statistician Joseph Hill and adapated by Harvard mathemitician Edward Huntington.

Hill's design applies the geometric mean of the two whole numbers closest state's quota. This method changes the rounding threshold from .5 to the geometric mean between the greater and lower whole numbers. If the quota exceeds the geometric mean, we round up, and if not, we round down. Since remainder of the geometric mean ranges between .41 and .5, and the approach addresses a couple of quirks that are more apparent when the threshold is .5, it has problems and isn't as easy to understand or explain to the general public.

For two consecutive whole numbers, the geometric mean is [sqrt(n*(n+1))]

Every method has to allow for the possibility that the number of representatives produces sum that differs from the number of seats allocated. For example, in the case of 2020, the Huntington and Hill formula described below generates a sum of 437 representatives on the first try.

Somehow, in that scenario, the formula has to find states who give up two representatives. An adjustment must be made and the choice of this adjustment can dramatically change apportionments. Hamilton's method adds or subtracts seats based on the size of the remainder, while other methods adjust the divisor. [Young shows the math; other sources are....] [Buklinski and Young 1982 show that the paradoxes that Huntington and Hill hope to avoid are largely unavoidable.]

[Show graph with geometric mean]

Their revised method has a couple of mathematical advantages that resolved quirks in the process that were more apparent when the number of seats was routinely growing. The geometric means gives the process some interesting but obscure mathematical properties, and it gives the process some scientific gloss, but it does not eliminate the possibility that too few or too many seats will be allocated.

Young explaind the reasoning for this alternative and has argued that this method generates a bias for small states, despite claims from early in the 20c that the method is unbiased.

[graph from his paper]

[samee table as above with the different threshold] Take for example, the states of Montana and New York in 2020:

1. Montana in 2020 has population of 1,080,577. New York is 19,336,776.
2. The US Population is 328,771,307.
3. Montana's quota is MT population / US population * 435, or: (1,080,577 / 328,771,307) * 435, or 1.430.
4. New York's quota is NY population / US population * 435, or (19,336,776 / 328,771,307) * 435, or 25.584.

Intuitively we'd expect Montana to round down from 1.430 to 1 seat, and New York rounds up from 25.584 to 26 seats.

But the Huntington Hill method used since 1940 looks at the geometric means of the upper and lower whole numbers. In the case of Montana, we compute the geometric mean of 1 and 2, which gives:

sqrt(1*2) = sqrt(2) ~= 1.414.

Montana's quota is 1.430, which exceeds 1.414, the geometric mean of 1 and 2, which gives them two representives.

New York's threshold is the geometric mean of 25 and 26

sqrt(25*26) = sqrt(650) ~= 25.4951. [as n approches inf, remainder of sqrt(n*(n+1)) approaches .5]

New York's quota is 25.584 which like Montana exceeds the threshold (25.495 in this case).

But using this method the total number of assigned seats is 437. We eliminate assignments by gradually increasing the denominator, which reduces every state's quota. When this happens, two states eventually lose a representatives, which gets us to 435.

In this case, the two states are Minnesota and New York.

But I thought we assigned seats in ascending order?

A clever way to regenerate the Huntington Hill method is described on the census web page at:

This method produces the same results but eliminates any initial miscalculation and doesn't convey the possibility that a seat is somehow being taken or given after the fact:

1. Assign one seat to every one of the states.
2. Assign the remaining 385 seats as follows: a Compute a "priority score" that takes the population divided by the geometric mean of 1 and 2, or the seat the state already has times the number it would have with an additional seat. b Every state's initial priority score is its population / sqrt (1*2), or its population / ~ 1.41. c Identify the state with the highest priority score and assign the 51st seat to that state. The 51st seat goes to the state with the greatest population. d That state now has two seats. Recompute that state's priority score as its population divided by the geometric mean of n and n+1, where n is the new, higher number of the state's seats.

   e For this state with the 51st seat, it's the geometric mean 2 and 3, which is sqrt(2*3) or sqrt(6), 2.449. Its priority score goes down accordingly.

   f Reorder the priority scores. The 52nd seat goes to the state with the newly determined high priority score. In all likelihood, the 52nd seat goes to a different state that now has 2 seats. Its priority score is also recomputed to a lower value.

   g Repeat steps e and f until all 435 seats have been allocated. In each iteration, the priority score is recalculated and reduced, and the state awarded a given seats (probably) moves somewhere further back in the line.

And mirable dictu, you end up with the same result as the Huntington Hill method described abov e, but without the inelegance of changing the denominator after the initial assignment. In this way, the seat assignment are fun and nothing looks like a process that gives a seat and then takes it away. Some states will nearly miss the coveted "435th" seat assignment but the mechanism is prettier.

Look at assignment if the continued; we get NY at 436 and MN at 437.

Why is MT and AL the relevant comparison? Because MT gets the biggest break in the adjustment to a geometric mean threshold. Rounding gets it nowhere near a second seat.

The Webster method gets it right on the first try with 435. The difference is MT gets 1 and NY gets 26.

[Explain Webster method]

Usually the result rounds accurately. Notice the residuals for each method. Rounding up from less than .5 or rounding down with a residual greater than .5 is infrequent. [Distributions of Webster and HH method residuals.]

[Distribution of residuals]

How many states have been screwed as badly as NY?

Since the Huntington-Hill method came into effect for the 1940 census, nine censuses have produced 446 individual state apportionments. The method rounded down the seat allocation from quotas in 215 of the 446. Only 16 had residuals greater of than .5. The largest residual (i.e., the state losing a seat with the biggest rounding error) was California in the 1950 census. If the current estimate holds, New York will be second greatest.

Twelve of these 16 unlucky states had seat counts of 10 or more.

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0.9000
## v readr   1.4.0     v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

## # A tibble: 16 x 6
##    State      Pop Year  Seats Quota Residual
##    <chr>    <dbl> <chr> <dbl> <dbl>    <dbl>
##  1 CA    10586223 1950     30 30.7    -0.722
##  2 NY    19336776 2020     25 25.6    -0.585
##  3 CA    23668562 1980     45 45.6    -0.584
##  4 PA    11319366 1960     27 27.6    -0.576
##  5 IN     5490179 1980     10 10.6    -0.574
##  6 IL    10081158 1960     24 24.6    -0.559
##  7 TN     3291718 1950      9  9.55   -0.553
##  8 KY     2944806 1950      8  8.55   -0.546
##  9 MA     5148578 1960     12 12.5    -0.543
## 10 NJ     7748634 1990     13 13.5    -0.536
## 11 MA     6029051 1990     10 10.5    -0.532
## 12 MO     4319813 1960     10 10.5    -0.524
## 13 GA     5464265 1980     10 10.5    -0.524
## 14 NY    18044505 1990     31 31.5    -0.521
## 15 CT     3050693 1970      6  6.51   -0.505
## 16 OR     2110810 1970      4  4.50   -0.501
```

To compare with states receiving an additional seat with a residual less than .5:

```
## # A tibble: 11 x 6
##    State      Pop Year  Seats  Quota Residual
##    <chr>    <dbl> <chr> <dbl>  <dbl>    <dbl>
##  1 NV      110247 1940      1  0.366    0.366
##  2 MT     1080577 2020      2  1.43     0.430
##  3 SD      673247 1970      2  1.44     0.436
##  4 CA    33930798 2000     53 52.4      0.447
##  5 NV      160083 1950      1  0.465    0.465
##  6 NC     8067673 2000     13 12.5      0.470
##  7 AR     1949387 1940      7  6.47     0.473
##  8 MN     5314879 2010      8  7.48     0.478
##  9 NH      606921 1960      2  1.48     0.479
## 10 RI     1055247 2010      2  1.48     0.485
## 11 MT      701573 1970      2  1.50     0.496
```

Theoretically, the quota would round down to zero for a very small state, but we have only seen this occur twice since 1940 (Nevada in 1940 and 1950). The Huntington-Hill method guarantees at least one seat for every state, without any adjustment, regardless of its population.

The second most generous seat allocation will be 2020 if the population figures hold: Montana, winning a second seat despite a small residual of .43. Both New York's smaller apportionment and Montana's larger one are historic.

How many states have been gimmeed as goodly as MT? Montana will make history for the lowest quota that yields from 1 to 2 seats. Montana is always kind of on the bubble. But a quota below 1.5 that yields two seats just doesn't happen very often.

only states with quotas below 1.5?
How many times has HH differred from Webster?

```
## # A tibble: 18 x 6
##    st_name        year  quota hunhill webster web_v_hun
##    <chr>          <chr> <dbl>   <dbl>   <dbl>     <dbl>
##  1 Arkansas       1940   6.47       7       6        -1
##  2 Michigan       1940  17.5       17      18         1
##  3 Kansas         1950   5.53       6       5        -1
##  4 California     1950  30.7       30      31         1
##  5 New Hampshire  1960   1.48       2       1        -1
##  6 Massachusetts  1960  12.5       12      13         1
##  7 Montana        1970   1.50       2       1        -1
##  8 South Dakota   1970   1.44       2       1        -1
##  9 Connecticut    1970   6.51       6       7         1
## 10 Oregon         1970   4.50       4       5         1
## 11 New Mexico     1980   2.50       3       2        -1
## 12 Indiana        1980  10.6       10      11         1
## 13 Oklahoma       1990   5.52       6       5        -1
## 14 Massachusetts  1990  10.5       10      11         1
## 15 Rhode Island   2010   1.48       2       1        -1
## 16 North Carolina 2010  13.5       13      14         1
## 17 Montana        2020   1.43       2       1        -1
## 18 New York       2020  25.6       25      26         1
```

We dedicated an entire chamber to protecting the numerical advantage of small states. The House doesn't need to help.