

Minimum Roughness: A New Approach to Interpolation and Smoothing

Griffith Feeney

Demography–Statistics–Information Technology, Letter No. 6, 11 June 2014

INTERPOLATION is a way of filling in a series of given values with "in between" values.

Point interpolation fills in points between a series of given points- Given numbers of life table survivors at ages 1, 5 and 10, for example, values for ages 1, 2, 3 and 4 and 6, 7, 8 and 9 may be required.

Group interpolation breaks down grouped data on a continuous variable into smaller constituent groups. Given a population age distribution in five year age groups, it may be required to interpolate numbers at single years of age.

Demographers have traditionally used polynomial interpolation methods, but these have several disadvantages.

- There are many methods; there may be no good reason for choosing one over the others
- The formulas are complex and tedious; understanding is difficult
- The methods may require that the values be given at regular intervals
- Most importantly, the methods break down for some data patterns

1. Example of Polynomial Interpolation Breakdown

Shryock and Siegel present an example in their old but classic and still useful compendium (see the RESOURCES section below). They observe that "In census or survey counts of population by 5-year age groups there may be a large drop from one age group to the next, followed by a much smaller drop". The text then displays Figure 1 below.

The interpolation breaks down because nothing in the data suggests the fall-rise-fall pattern in the interpolated values. The pattern is an artifact of the interpolation procedure.

I have developed a new approach to interpolation based on the idea of "minimum roughness". It has several advantages over traditional methods. In particular, it gives satisfactory results for observed data for which polynomial methods break down, as shown in Figure 2. The superiority of the interpolated points in Figure 2 is immediately evident.

2. Point Interpolation: Annual Population Numbers

The same problem may occur with point interpolation. The solid dots in Figure 3 below show mid-year population of persons aged 0-4 at 5 year intervals for 1965 through 1990 for Sri Lanka, as estimated by the United Nations Population Division. The hollow circles show the result of fifth degree polynomial interpolation.

Figure 4 below shows minimum roughness interpolation applied to the same given points. The difference between Figures 3 and 4 is not as dramatic as between Figures 1 and 2, but

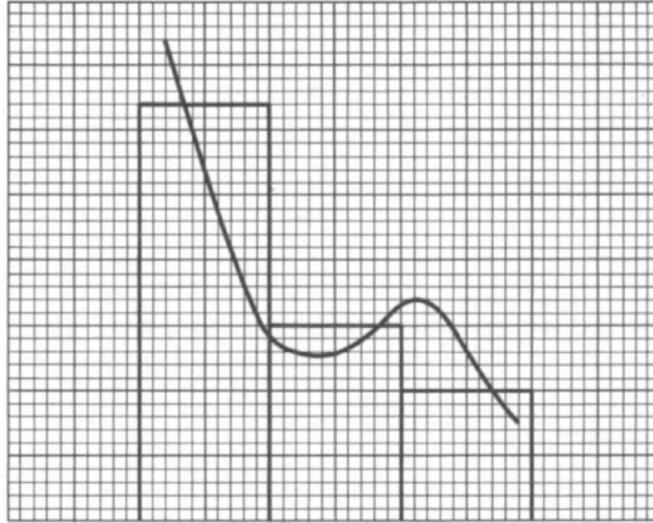


Figure 1 Shryock and Siegel example of polynomial interpolation breakdown

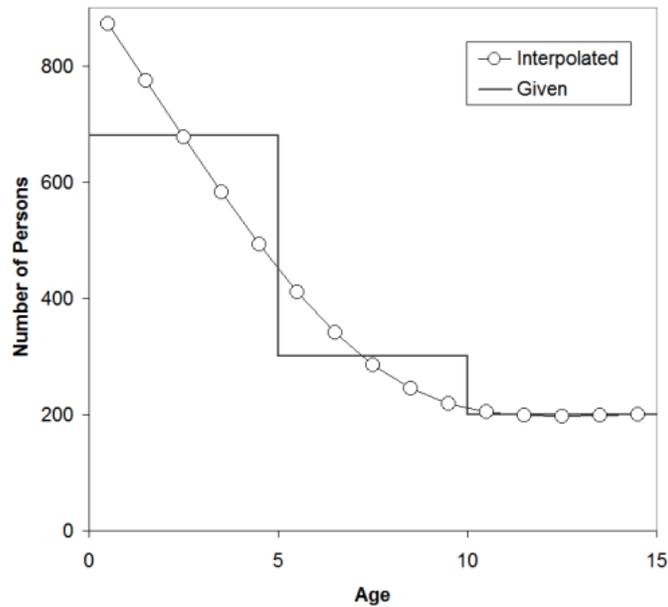


Figure 2 Minimum roughness interpolation applied to Shryock and Siegel example

we will probably prefer the minimum roughness interpolated values for the more sensible pattern of interpolated points between 1965 and 1975.

3. Implementing Minimum Roughness Interpolation

How does minimum roughness interpolation work? We don't need any special mathematical prowess, only a bit of experience using formulas in spreadsheets and the Excel SOLVER Add-In or an equivalent facility.

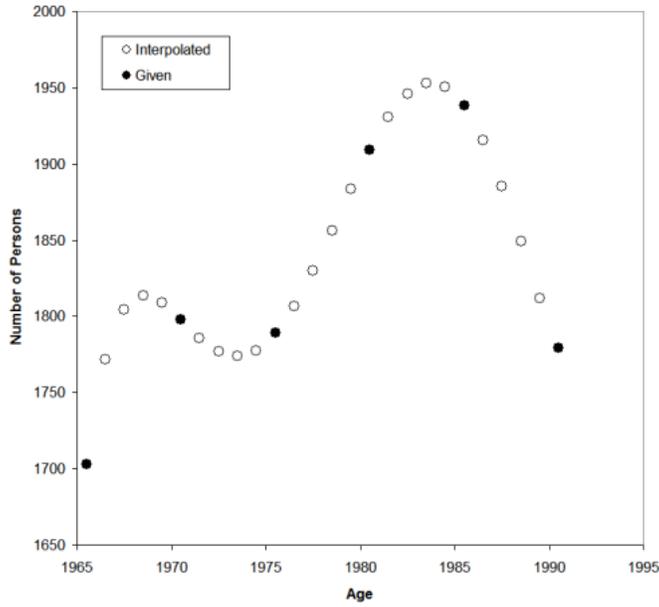


Figure 3 Example of polynomial point interpolation

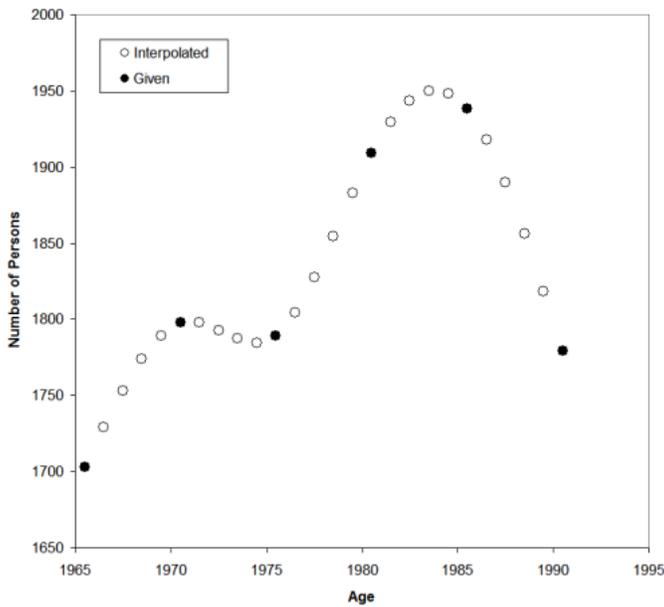


Figure 4 Minimum roughness point interpolation applied to Figure 3 example

We begin with the idea that the interpolated points together with the given points should form as smooth a curve as the given points allow. To get smoothness, we define a measure of “roughness” and then chose interpolated values to minimize this measure of roughness.

Given a series of points at one year intervals of age or time, calculate the following for each point except the first and the last.

1. The average of the y-values for the preceding and following points.

2. The difference between this average and the y-value of the point.
3. The square of this difference.

Now sum these squared differences. The result is a measure of roughness. To find the interpolated values, follow these steps.

- A. Create a spreadsheet with cells for the observed and interpolated points.
- B. Enter the observed values in the observed value cells and sensible initial values in the interpolated value cells.
- C. Put a formula that calculates the roughness measure described above in a cell labeled “Measure of Roughness”.
- D. Use SOLVER to find interpolated values (“By Changing Cells” in SOLVER) that minimize this measure of roughness “Set Target Cell” in SOLVER).

For point interpolation, the initial values between two given points may be set equal to the value of the first point. For group interpolation they may be set to the number in the given group divided by the number of constituent groups to be interpolated.

For group interpolation, constraints should be used to ensure that the sum of the interpolated values for each group equals the given number in the group. Constraints may also be used for point interpolation to ensure, for example, monotonicity of interpolated values where this is appropriate.

4. Conclusion

Minimum roughness interpolation has several advantages.

- It gives good results for data for which other methods break down, as illustrated by the above examples.
- It provides a single, unified approach to interpolation. The only formula required is the one defining the measure of roughness.
- It is not restricted to observed data given at equal intervals.
- It is easily implemented with a computer spreadsheet program (no tables of constants needed).

Why wasn’t minimum roughness interpolation invented long ago? One obvious answer is that calculating the interpolated values in Figure 4 requires numerical minimization of a function of 20 variables. This took less than one second on my modest laptop computer, but 40 years ago, when Shryock and Siegel’s book was first published, very few people had access to this kind of computing power.

Minimum roughness methods do have the disadvantage that the numerical minimization algorithm may fail to converge. Often this is easily fixed by tinkering with the goodness of fit measure, usually nothing more than multiplying differences by a multiple of ten before squaring. But it is possible one will hit a dead end. Another limitation is that problems involving larger numbers of parameters may choke SOLVER—for now.

A final *caveat* applies to all interpolation methods. We use them because the data don’t provide as much information as we would like. We can use interpolation to supply what

is missing, but the accuracy of the interpolated values is uncertain. There is no Taylor Remainder Theorem for demographic data!

5. Resources

Shryock and Siegel Volume 2 (1973) is available on [Google Books](#) (follow the link and search the authors' names). The plot shown in Figure 1 above appears in Chapter 22 on page 701.

Readers familiar with Excel and the SOLVER tool (not installed by default) may download and study [minimum-roughness-group-interpolation.xls](#), which illustrates the implementation of group interpolation, and [minimum-roughness-point-interpolation.xls](#), which illustrates point interpolation.

DEMOGRAPHY, STATISTICS, and INFORMATION TECHNOLOGY is an every-other-month email letter from Griffith Feeney, Ph.D., an international consultant based in Scarsdale, New York, USA. Email me at feeney@feeney.com, telephone me +1 914 725 2218, or text message me at +1 914 721 3950. This letter was scheduled to go out on 15 May 2014 but was delayed under pressure of other work.

This and past letters are available at demographer.com. To UNSUBSCRIBE, send email to feeney@feeney.com with “unsubscribe” in the subject field. To SUBSCRIBE, send email to feeney@feeney.com with “subscribe” in the subject field.