Practical Use of Interpolatory Cubic and Rational Spline Functions for Fertility Data

Zenji Nanjo

NUPRI Research Paper Series No. 34

December 1986

Zenji Nanjo
Professor
Fukushima Medical College
and
Affiliated Researcher
Nihon University
Population Research Institute

CONTENTS

TABLES

A B S T R A C T

Some studies have been undertaken for the purpose of obtaining data for single years of age applying interpolatory spline function to births by five-year age groups of mothers, generally, which graduate approximately bell-shaped distribution data. Neither cubic nor quintic spline interpolation always provides good results easily. In short, this type of birth data belongs to that type which cannot be interpolated easily. Therefore, giving up the idea of applying the interpolation method suitable for the given data itself, and making the data easier to interpolate using a suitable transformation, the transformed values at each age are obtainable by using rational (or cubic) spline interpolation and then obtaining the interpolated values by single years of age of the given data after the inverse transformation. There may be many formulae available for the transformation, but among them, the angular transformation was found to be the most practical and suitable for our purpose. In the present paper, a few examples of the method and the results of a rather severe test for it are given.

It should be noted that the rational spline function, which includes the cubic spline function as its special case, is generally useful for demographic study.

I. Introduction

The Pearson III-type distribution has been long known as a graduation formula for interpolating figures, by single years of age of mothers from approximately bell-shaped distribution data such as births by five-year age groups of mothers (Wicksell, 1931; Keyfitz et al., 1971). Concerning this problem, there are also some other studies, including one by Hoem et al. (1981). The present paper will discuss the method of obtaining the number of births by single years of age from births by five-year age groups of mothers by means of spline interpolation.

Such birth data are those to which ordinary interpolation formulae are hard to apply. Boneva, Kendall, and Stefanov (BKS) (1971) showed an example, especially for that application, where the birth data of Bulgaria (1963) similar to the data mentioned above was interpolated, using their computer program /HISTO/SPLINE. Schoenberg (1973) reported an example where the same data were interpolated using the specially prepared computer program SPLINT (the quintic B-spline). But it cannot necessarily be said that their results are very satisfactory (cf. (IV) below). McNeil et al. (1977) used an approximate spline, relaxing the "Area Matching" restriction, so that the values by single years of age (interpolated by applying a quintic spline function to the fertility rates by five-year age groups), might not be negative.

In the present study, interpolatory spline functions will be used as BKS and Schoenberg used them in order to satisfy the Area Matching Property (Boneva et al., 1971; Schoenberg, 1973). However, the idea of using spline interpolation directly for the given data will not be applied. Use of the spline interpolation will be made after transforming the given data by a certain method so that it will be easier to interpolate data. Here, the meaning of the interpolation may change a little mathematically, but practically-speaking, our method is suitable. The transformations used here are the α-th root,

---

Note: This paper is a revision of "Practical Use of Interpolatory Cubic and Rational Spline Functions in Demography," Paper presented at the IUSSP Conference, Florence, 7-12 June, 1985.

the logit, and the angular transformation. In using these, it is necessary first to adjust data (cf. (III) below). The latter two are known as the transformation for linearizing the S-shaped distribution in Bioassay. The data used in the present study are distributed leaning considerably to the left, and do not become linear after the transformation. However, if a cubic or rational spline interpolation is used for the data (especially after the angular transformation), fairly good interpolated values may be obtained. The method of interpolation is described in (III) below and some examples are shown in (IV). It is found that cubic spline interpolation tends generally to suffice for the present purpose, but in some cases, rather inappropriate results are obtained. For example, when the spline function obtained fluctuates or some of the interpolated values become negative, rational spline interpolation (which is more flexible), may be desirable.

II. Rational Spline Interpolation

Use of the rational spline interpolation was attempted by Späth (1971). When y_0, $y_1,\ldots,y_N$ are the values of function at N+1 points $x_0, x_1 \ldots x_N$ (where $a = x_0 < x_1 < \ldots < x_N = b$) and the differential coefficients of 1st degree y'_0 and y'_N at x_0 and x_N are given, the function s(x) is called a rational spline function, if the function:

$$s(x) = a_k + b_k(x-x_k) + c_k(x-x_k)^2 + \frac{d_k}{x - x_k + p_k} \qquad \text{(i)}$$

defined in the interval (x_k, x_{k+1}) $(k = 0,1,2,\ldots,N-1)$, satisfies the *conditions:*

(1) $s(x_k) = y_k (k = 0, 1, 2, \ldots, N)$; (2) $s'(x_0) = y'_0$, $s'(x_N) = y'_N$; and (3) the second derivative of s(x) is continuous in the interval (a,b), where a_k, b_k, c_k, d_k and p_k are parameters. We call it a rational spline interpolation to obtain such an s(x). Here p_k are parameters which satisfy $-\infty < p_k < -(x_{k+1}-x_k)$ or $0 < p_k < \infty$. The values of p_k should be decided before the interpolation.

In the present paper a slightly modified formula is used instead of the formula (i). Generally speaking, it may not be said that the rational spline interpolation is a very useful one, for it is not always easy for us to decide the suitable values of p_k from the given data. However, in the interval (x_k, x_{k+1}), when $p_k \to \pm\infty$, $s(x)$ becomes a cubic spline function, when $p_k \to 0$, it becomes linear, and when $p_k \to -(x_{k+1}-x_k)$, it becomes quadratic. So, we take advantage of this property, and use it as a rule, setting $p_k \geq 10^5$, instead of a cubic spline function. When the results are not satisfactory, it is only necessary to slightly change some of the parameters $p_k (k=1,2,\ldots,N)$, so as to correct the results. Thus, this is a very convenient method of interpolation which may be said to be useful in reducing the fluctuation of $s(x)$ which often occurs in polynomial spline interpolations. See Examples 1 and 3 for the choice of values of p_k. It would not be very difficult for us to choose a few parameters p_k, especially after the said transformation.

It may be of interest to note that the algorithm of the rational spline interpolation is as simple to use as the cubic spline interpolation.

III. Method of Interpolation

First the number of births by five-year age groups of mothers will be dealt with here. While the age of mothers is x, let $c(x)$ be the proportion of the number of births by age of mothers to the total births. Let the value of $c(x)$ be cumulated as age x advances, and denoted by $p(x)$. Then $p(15) = 0$ and $p(50) = 1$. (The data are given in seven five-year age groups, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, and 45-49.) Usually $p(x)$ has an S-shaped distribution. If the following transformations (there can be many others), are applied to $p(x)$, the transformed values (not always linear) tend to be more easily interpolated by the cubic spline function.

(1) The α-th root transformation $\phi(u) = \sqrt[\alpha]{u}, (\alpha=3.5, 3.6, \ldots), u \geqq 0$,

(2) The logit transformation $\phi(u) = \frac{1}{2} \ln \frac{u}{1-u}$, $0 < u < 1$,

(3) The angular transformation $\phi(u) = \sin^{-1}\sqrt{u}$, $0 \leq u \leq 1$

The actual calculation of values of $p(x)$ at every fifth age was done in the following two ways:

(1) Let the number of births of age groups 15-19,...,45-49 be denoted respectively by B_1, B_2,...,B_7, and let

$$T = B_1 + B_2 + \ldots + B_7, \quad p(15) = 0, \quad p(20) = B_1/T,$$

$$p(25) = (B_1 + B_2)/T, \ldots, P(50) = (B_1 + \ldots + B_7)/T.$$

(2) The birth data are generally given for ages 15-49, but actually, there can be births for under age 14 or above age 50 (cf. data for 1970 and 1980 in Japan). Let these births be respectively α % and β % of births of ages 15-19 and ages 45-49. Now, let the number of births of age 14 and under, 15-19,...,45-49,50 and over be denoted respectively by B_0,B_1, B_2,...,B_7, and B_8, where $B_0 = \alpha B_1/100$, and $B_8 = \beta B_7/100$. Then, let

$$T = B_0 + B_1 + \ldots + B_8, \quad p(15) = B_0/T,$$

$$p(20) = (B_0 + B_1)/T, \ldots, \quad p(50) = (B_0 + B_1 + \ldots + B_7)/T.$$

Therefore, when $\alpha \neq 0$ and $\beta \neq 0$, p(15) is nearly zero but is positive, and p(50) closely approximates to one but is less than one. α and β can be decided according to the given data. Case (1) is a special case of (2) with $\alpha = \beta = 0$. The angular transformations with $u = p(x)$ in the cases 1 and 2 are called the angular transformations (1) and (2) respectively. In (2), we put $\alpha = 0.1, \beta = 1$ tentatively. In the logit transformation is used p(x) with α and β that $\alpha > \delta$ and $\beta > \delta$ for some $\delta > 0$, e.g. $\delta = 1$ or 1.5.

The present method of interpolation using p(x) obtained in this way is as follows:

First we obtain the values of ϕ at every fifth age using p(x) instead of u in these transformations, and then obtain the values of at each age by applying a rational or cubic interpolation to the values at every fifth age, and then obtain the interpolated values of p(x) at each age, applying the inverse transformation to the values of ϕ at each age. If we multiply "the value of T" by the values of p(x) at each age, we can obtain the cumulated number of births at each age, and from this we obtain the number of births at each age. Especially the angular transformations (1) and (2) give fairly stable and good results. As a cubic spline function is a special case of a rational spline function with parameter p_k which are sufficiently large (usually $p_k \geq 10^5$) for K=1,2,...,7, the latter will be used hereafter.

Notice that as end conditions, $y'_0 = 0$ and $y'_7 = 0$ are typically used (cf. McNeil et al., 1977; Schoenberg, 1973; Boneva et al, 1971). But it should be noted that the choice of the values of y'_0 and y'_7 as well as α and β is arbitrary and it tends to affect the interpolated values by single years of age especially in age groups 15-19 and 45-49 (cf. Example 4).

Furthermore, the following should be noted: When the angular transformation (1) is used, at least one of y'_0 and y'_7 should be a suitable value which is not zero in order to get better interpolated values in the vicinity of the end points (in Example 1, $y'_0 = 0.003$, $y'_7 = 0$). When better results are needed, what is necessary is to change some of the parameters p_k a little (cf. Examples 1 and 3).

IV. Examples

Three transformation methods will be examined here by means of some examples. Hereafter, a rational spline interpolation is generally used with the parameters $p_k = 10^5$ ($k = 1,\ldots,7$).

Example 1. Births by Five-Year Age Group of Japanese Mothers, 1980

Ages	15-19	20-24	25-29	30-34	35-39	40-44	45-49	Total
Births	14,576	296,854	810,204	388,935	59,127	6,911	257	1,576,864
Percent	0.9	18.8	51.3	24.7	3.7	0.4	0.2	100

From these values, births by age are obtained by the present method.

Table 1. True and Interpolated Births by Age, Japan, 1980

Age	True values (ξ_i)	Interpolated values using logit (x_i)	angular(2) (x_i)	angular(1) (x_i)(*)
15	41	9	38	41
16	416	60	302	336
17	1,620	374	1,277	1,334
18	3,895	2,392	3,813	3,827
19	8,604	11,741	9,145	9,038
20	16,758	33,037	18,774	18,507
21	30,457	55,786	33,637	33,254
22	49,000	65,699	54,150	53,823
23	79,823	66,350	80,075	80,136
24	120,816	75,982	110,218	111,134
25	151,933	109,165	141,300	142,934
26	168,685	150,747	165,425	165,679
27	173,995	183,898	176,663	175,014
28	165,610	193,167	172,559	170,906
29	149,981	173,227	154,256	155,671
30	130,615	135,897	126,912	132,405
31	104,776	99,594	98,996	103,720
32	78,740	70,561	73,935	74,639
33	50,382	49,086	52,892	49,155
34	24,422	33,798	36,199	29,017
35	19,138	23,028	23,745	22,187
36	16,266	15,371	15,250	14,162
37	10,840	10,069	9,733	10,559
38	7,702	6,503	6,260	7,349
39	5,181	4,157	4,140	4,871
40	3,004	2,675	2,828	3,067
41	1,725	1,814	1,876	1,854
42	1,200	1,232	1,173	1,076
43	650	773	679	598
44	332	417	355	317
45	163	179	163	157
46	53	57	64	68
47	29	16	22	24
48	5	4	6	6
49	7	1	1	1

(*) $\alpha = \beta = 0$; $y'_0 = 0.003$, $y'_7 = 0$; $P_1 = P_2 = P_6 = P_7 = 10^5$, $P_3 = 10$, $P_4 = 15$, $P_5 = -0.5$.

The values of $Q = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \xi_i)^2}$ are computed for comparison within each example.

The values of Q/S are computed for comparison among examples where S is the total number of births in the age interval 15-49.

Table 2. Values of Q by the Present Method, Japanese Births, 1980

Transformation	$\sqrt[3.5]{p}$ (1)	$\sqrt[3.5]{p}$ (2)	logit	angular (1)	angular (2)	angular (1) (*)
$Q \times 10^4$	0.64	0.64	1.45	0.422	0.420	0.324
$Q/S \times 10^3$	4.06	4.06	9.20	2.68	2.66	2.06

In this table, the values of Q using the α-th root transformation were computed only in case 1 (i.e. $\alpha = \beta = 0$). In $\sqrt[3.5]{p}$ (1) the value for age 49 was negative. Accordingly, in $\sqrt[3.5]{p}$ (2), values 10 and 0.5 were used respectively instead of 10^5 and 10^5 for ages 40-44 and 45-49 as the parameters p_k for the rational spline interpolation in order to remove negative interpolated values. Further, the transformation $\sqrt[\alpha]{p}$ is often found to be useful but not very recommendable, for it is hard to decide the suitable values of α. Therefore, only the logit and the angular transformation will be dealt with below.

Example 2. Interpolation of the number of births by age group of Bulgarian mothers of 1963. This is an examnple used by BKS (1971) and Schoenberg (1973).

Ages	15-19	20-24	25-29	30-34	35-39	40-44	45-49	Total
Births	7,422	19,261	14,385	6,547	2,123	451	17	50,226
Percent	14.8	38.3	28.7	13.1	4.2	0.9	0.0	100

BKS compared the interpolated values by single years of age obtained by the computer program /HISTO/SPLINE with the true (given) values by single years of age. According to it, the interpolated values of age 15 and 16 are much larger than the true values, while those of ages 18-20 are smaller, and those of ages 48 and 49 are negative. Schoenberg applied the quintic B-spline interpolation to the same data.

(Program, SPLINT). His graphic curve shows that the results are pretty good, but the values of each year of age in ages 45-49 are rather inappropriate, those of ages 48 and 49 being negative. From the interpolated values and the true values, we have the following table.

Table 3. Values of Q by Some Methods, Bulgarian Births, 1963

Method	The Present Method logit	angular(1)	angular(2)	BKS Method	Schoenberg Method
$Q \times 10^2$	13.5	0.882	0.853	3.16	*
$Q/S \times 10^3$	26.9	1.76	1.70	6.29	*

* There are no values by single years of age, so it is impossible to get the value of Q.

The data of Example 1 are distributed leaning considerably to the left in comparison with that of Example 2. Therefore, the values of $Q/S \times 10^3$ of the former are larger than those of the latter.

Example 3. Using the Weibull distribution, the present method is put to a rather severe test. As in the case of the number of births, giving seven groups of data, we obtain the values of 35 small groups by means of the present interpolation. The reason of adopting the Weibull distribution is that its shape changes with parameter m, so that it has a distribution shape very similar to that of the foregoing examples, and the computation of it is very easy.

The Weibull distribution is the one with the density function:

$$f(x) = \begin{cases} (mx^{m-1}/a)e^{-(x^m/a)} & (x \geq 0), \\ 0 & (x < 0), \end{cases}$$

where a and m are parameters. Therefore, the distribution function is

$$F(x) = 1 - e^{-(x^m/a)}$$

When we divide the interval $(0, m_1)$(where m_1 = 2 or 2.5) into seven equal parts on the x-axis and let their dividing points be denoted by $0 = x_0, x_1, \ldots, x_7 = m_1$, and dividing the same interval into 35 equal parts, we let their dividing points be denoted by $0 = x'_0, x'_1, \ldots, x'_{35} = m_1$, we set

$$F(x_i) - F(x_{i-1}) = H_i \quad (i=1,2,\ldots,7)$$
$$F(x'_i) - F(x'_{i-1}) = H'_i \quad (i=1,2,\ldots,35). \qquad \text{(ii)}$$

Our problem is to compare 35 interpolated values obtained by the present method from the given data H_i $(i=1,2,\ldots,7)$ with the true values H'_i $(i=1,2,\ldots 35)$. For brevity's sake, we put a = 1.

The cases are dealt with where $m_1 = 2$, $m_1 = 2.5$; m = 3, m = 4, that is, Case 1 (m = 3, m_1 = 2), Case 2 (m = 4, m_1 = 2), Case 3 (m = 3, m_1 = 2.5), and Case 4 (m = 4, m_1 = 2.5).

Let the following be given data (using f(x) x 1000) (cf. (ii) above).

	H_1	H_2	H_3	H_4	H_5	H_6	H_7	Total (approx.)
Case 1	23.05	147.16	297.05	307.97	170.58	47.69	6.15	1000
Case 2	6.64	94.49	315.99	401.28	166.07	15.45	0.18	1000
Case 3	44.53	260.88	404.49	238.13	50.81	3.31	0.05	1000
Case 4	16.14	213.05	503.09	252.19	15.49	0.04	0.00	1000

From this table it will be apparent that each case has a very different shape of distribution. The data in each case above were interpolated by the present method. We compared these interpolated values with H'_i and then computed the values of Q and Q/S.

Table 4. Values of Q by the Present Method Using Weibell Distribution (using f(x) x 1000)

Transformation	logit	angular(1)	angular(2)
Case 1	10.3	.351	.260
Case 2	5.2	.147	.150
Case 3	16.0	.524	.390
Case 4	120.5	.417	.439

Note: For example, the value of $Q/S \times 10^3$ is .390 in case (3), for angular (2).

From Table 4 above, it is found that the angular transformation always gives better results than the logit transformation. There is not much difference between the angular transformations (1) and (2). Case 4 is so far from the distribution of the number of births by age of mothers that we need not take it into consideration here, but it is shown for the reader's information. In Cases 1, 2 and 3, there are no negative values, but the data in Case 4 may not be desirable for interpolation. In fact, some negative values appear and there are slight fluctuations when $x \geq 1.36$ in the logit transformation, and when $x \geq 1.93$ in the angular (1) and (2). But even these can be corrected considerably using slightly changed parameters p_k (e.g., $p_1 = p_2 = p_3 = p_4 = 10^5$, $p_5 = 10$, $p_6=0.5$ and $p_7=0.5$).

Example 4. The present method may also be used to provide single-year fertility rates for populations in which age-specific fertility is tabulated only by five-year age groups. Data used here are two kinds of fertility rates multiplied by five for seven five-year age groups taken from the model fertility tables by Coale and Trussell (1974): the first data (data (1)) are the case with mean age m = 28.0, standard deviation σ = 4.5, and the ratio of fertility at ages 15-19 to fertility at ages 20-24, $R_1 = 0.0481$; in the case of the second (data (2)), m = 27.0, σ = 6.0, and $R_1 = 0.3161$. The data are given as follows:

(per million women)

Ages	15-19	20-24	25-29	30-34	35-39	40-44	45-49	Total (approx.)
Data(1)	12,435	258,315	437,770	214,040	64,685	12,050	710	10^6
Data(2)	102,595	324,535	289,165	169,000	83,925	27,570	2,800	10^6

The cumulative fertility rates at 15, 20,..., and 50 were computed as McNeil et al. did (see McNeil et al. 1977). For the rest, the forgoing computation (i.e. transformation, interpolation, etc.) is followed (cf.(III) above). Estimated single-year fertility rates were compared with the single-year rates taken from Coale and Trussell's tables graphically and numerically. As a result, the estimated values seemed to be reasonable. For data (1) which has a small standard deviation (σ = 4.5), the angular transformation(1) was used, for which we have the values of Q = .842 and Q/S x 10^3 = .842. For data (2), which has a large standard deviation (σ = 6.0), the angular transformation (2) was used, and we have the value of Q = .909 and Q/S x 10^3 = .909.

V. Concluding Remarks

The present paper is based on the idea that the data which are hard to interpolate can be interpolated after being suitably transformed. For the purpose of illustration, the birth data were used. The present method seems to give fairly good and stable results when using the angular transformation.

Here are some points to bear in mind regarding this research.

1. The present paper deals with data where ages are equidistant, but data where ages are at unequal intervals are also applicable, of course.
2. In the present method, a cubic spline interpolation may suffice for a way of interpolation, but the rational spline interpolation is more desirable when better interpolated values are needed. The rational spline interpolation will be useful in the study of demography.
3. The present method satisfies to a great extent what is called "Area Matching Property."

4. From Examples 1-4, it will be seen that (1) the values of Q/S for actual data are larger than those for non-actual data; (2) the Q/S values of the birth distribution by mother's age are larger when the distribution leans considerably to the left, than when it leans only slightly to the left.

5. The method is effectual for the interpolation of not only the births in this study but also other histogram-type data having an approximately bell-shaped distribution.

References

Ahlberg, J.H., E.N. Nilson, and J.L. Walsh. 1967. The Theory of Splines and Their Applications. New York: Academic Press.

Boneva, L.I., D.G. Kendall and I. Stefanov. 1971. "Spline transformations: Three new diagnostic aids for the statistical data analyst." Journal of Royal Statistical Society, Series B. 33: 1-70.

Coale, A.J. and T.J. Trussell. 1974. "Model fertility schedules: Variations in the age structure of childbearing in human populations," Population Index Vol. 40, No.2, pp.185-258.

Hoem, J.M. et al. 1981. "Experiments in modelling recent Danish fertility curves," Demography Vol.18, No.2, pp. 231-244.

Ichida, K. and F. Yoshimoto. 1979. Spline Functions and Their Applications, Tokyo: Kyoiku-Shuppan (in Japanese).

Keyfitz, N. and W. Flieger. 1971. Population: Facts and Methods of Demography. San Francisco: Freeman.

McNeil, D. R., T. J. Trussell and J. C. Turner. 1977. "Spline interpolation of demographic data." Demography 14:245-252.

Schoenberg, I.J. 1973. "Splines and histograms." pp. 277-327 in A. Meir and A. Sharma, Spline Functions and Approximation Theory. Basel, Stuttgart: Birkhauser Verlag.

Spath, H. 1971. "Rationale Spline-Interpolation." Angewandte Informatik 8:357-359.

Wicksell, S.D. 1931. "Nuptiality, fertility, and reproductivity." Skandinavisk Aktuarietidskrift Vol.14, No. 3, pp. 125-157.