Global Journal of Science Frontier Research: F
Mathematics and Decision Sciences
Volume 21 Issue 2 Version 1.0 Year 2021
Type : Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

Estimating Distributions using the Theory of Relative Increment Functions

By Sereko Kooepile-Reikeletseng
University of Botswana

Abstract- Bounded growth processes can be modelled, approximately by different mathematical models. The challenge for statisticians and mathematicians is finding suitable models for these processes. In this paper we illustrate a non-parametric method, using the the theory of relative increment functions, of estimating density functions of these processes. For a long time, mathematicians attempted to describe the cumulative prevalence of caries with the assumption that there is a mathematical model that would describe the caries prevalence and may be used for predicting caries incidences. In 1960 Porter and Dudman [12] introduced The relative increment function and called it the relative increment of decay as they designed it to compare dental caries increments among children. Further studies of this led to the motivation that the best suitable model for describing the cumulative prevalence of caries should be chosen from a set of distributions that have relative increment functions with the same monotonic behaviour as the relative increment of decay [1]. We illustrate how relative increment functions may be used to estimate the unknown indefinitely smooth probability density function of unimodal populations.

Keywords: relative increments, distribution, continuous. unimodal.

GJSFR-F Classification: MSC 2010: 12E12

ESTIMATINGDISTRIBUTIONSUSINGTHETHEORYOFRELATIVEINCREMENTFUNCTIONS

Strictly as per the compliance and regulations of:

Estimating Distributions using the Theory of Relative Increment Functions

Sereko Kooepile-Reikeletseng

Abstract- Bounded growth processes can be modelled, approximately by different mathematical models. The challenge for statisticians and mathematicians is finding suitable models for these processes. In this paper we illustrate a non-parametric method, using the the theory of relative increment functions, of estimating density functions of these processes. For a long time, mathematicians attempted to describe the cumulative prevalence of caries with the assumption that there is a mathematical model that would describe the caries prevalence and may be used for predicting caries incidences. In 1960 Porter and Dudman [12] introduced The relative increment function and called it the relative increment of decay as they designed it to compare dental caries increments among children. Further studies of this led to the motivation that the best suitable model for describing the cumulative prevalence of caries should be chosen from a set of distributions that have relative increment functions with the same monotonic behaviour as the relative increment of decay [1]. We illustrate how relative increment functions may be used to estimate the unknown indefinitely smooth probability density function of unimodal populations.

Keywords: *relative increments, distribution, continuous. unimodal.*

I. Introduction

Density estimation using non-parametric methods was first proposed by Fix and Hodges in 1951 as a way of moving away from distributional assumptions which at times restrict estimation. The methods proposed by Fix and Hodges were the Histograms, Naive estimator, Kernel estimator , Nearest neighbour estimator,Variable Kernel estimator and many others. [14].

In 1989 [16] proposed the use of relative increment functions for density estimation.

The relative increment function, h, of a distribution function, F, is defined as

$$h_{F(x)}(x) = \frac{F(x+a) - F(x)}{1 - F(x)} \text{ where } a = x_{k+1} - x_k.$$

He defined the function

$$\Psi(x) = \frac{(F(x) - 1).f'(x)}{f^2(x)}$$

and used the fact that if $\Psi < 1 (\Psi > 1)$, then the function h strictly increases (strictly decreases) to classify some well known distribution functions according to their monotonic behaviour. Szabo [20] developed an algorithm for finding the distributions of unknown unimodal population by eliminating a large class of continuous distributions whose behaviour of relative increments do not match the behaviour

Author: *Department of Mathematics, University of Botswana. e-mail: reikeletsengsn@mopipi.ub.bw*

Ref

14. Silverman, B.W.: Density estimation for statistics and data analysis. Pub. Monographs on Statistics and Applied Probability, London: Chapman and Hall, 1986.

of the empirical relative increment functions of the population being investigated. We illustrate this method by two numerical examples that showed that when the distribution relative increments behaves the same way as the empirical relative increments, the fit is superior to the ones with different monotonic behaviour of relative increments.

In this section we explain how we use relative increment functions to estimate density. Assume we have a large sample of a continuous random variable. We form the empirical cumulative distribution function F_{emp} at equidistant points x_k. Our aim is to find the smooth unimodal distribution which our sample belongs to. We assume that the distribution is a indefinitely smooth unimodal distribution whose probability density function has at most two points of inflection.

Suppose all the intervals $I_k = [x_{k-1}, x_k]$ have the same length a, for $k = 1, 2, ..., n$. Let v_k be the frequency distribution defined as the number of sample values in I_k, then we have relative frequency $r_k = \frac{v_k}{N}$ and the cumulative relative frequency as $y_k = \sum_{j \le k} r_j$, for all k.

The empirical cumulative distribution function F_{emp} whose points of discontinuity are at equidistant points, x_k, is therefore given by y_k, so

$$F_{emp}(x_k) = y_k \text{ for } k = 0, 1, ..., n$$

Define the relative increment function, h, for a distribution with cumulative distribution function $F(x)$ as

$$h_F(x) = \frac{F(x+a) - F(x)}{1 - F(x)} \text{ where } a = x_{k+1} - x_k$$

and define the empirical relative increment function, h_{emp}, for our population as

$$h_{emp}(x_k) = \frac{y_{k+1} - y_k}{1 - y_k} \text{ for } k = 0, 1, ..., n-1$$

Assume we have a large sample of a continuous random variable. We form the empirical cumulative distribution function F_{emp} at equidistant points x_k. Our aim is to find the smooth unimodal distribution which our sample belongs to. We assume that the distribution is a indefinitely smooth unimodal distribution whose probability density function has at most two points of inflection.

From this sample we calculate the empirical relative increment function $h_{emp}(x_k)$. If the monotonic behaviour of the empirical relative increment function $h_{emp}(x_k)$ is different from the monotonic behaviour of the theoretical relative increment function $h(x_k)$ of the cumulative distribution function $F(x)$, then we drop the corresponding smooth distribution $F(x)$. If $h(x_k)$ and $h_{emp}(x_k)$ have the same monotonic behaviour, we keep the corresponding cumulative distribution function $F(x)$ and put them in a class of possibilities, S.

From the set S a best fitting function is found by using the method of least squares. A distribution function $F(x) \in S$ providing the best fit to the cumulative relative frequency y_k such that $\sum_{k=1}^{n} [F(x_k) - y_k]^2$ is minimal

is selected or a distribution whose probability density function $f(x)$ provides the best fit to the relative frequencies r_k such that $\sum_{k=1}^{n} [f(x_k) - r_k]^2$ is minimal is selected.

Notes

Notes

This method can also be used to model bounded growth processes. Let g_k be a sequence of values measured at some equidistant points x_k. An upper bound B for g_k has to be determined such that B is greater than any value of g_k). To model the growth process of (x_k, g_k) we consider the transformed data

$$y_k = \frac{g_k}{B}(< 1)$$

as the values of of the empirical cumulative distribution function F_{emp} at the points x_k. The upper bound B is determined by building a parameter,B, into the distribution functions we want to fit, so instead of fitting $F(x)$, we fit $B * F(x)$. The estimated value of B gives the upper bound.

To use this method, we need to know the monotonic behaviour of relative increment functions h of distributions. A great number of classical smooth unimodal distributions has been investigated and classified according to the behaviour of their relative increments. These are listed in the following section.

II. Summary of Investigated Distributions

Here is a summary of distributions grouped according to the monotonic behaviour of their relative increment functions investigated by Szabo Z.[20] and myself.

2.1. The following probability distributions have increasing RIFs:

1. $F(x) = 1 - (-x)^k$ where $I = (-1, 0), k \in \mathbb{N}$
2. $F(x) = \sin x$ where $I = (0, \frac{\pi}{2})$
3. $F(x) = 1 + \tan x$ where $I = (-\frac{\pi}{4}, 0)$
4. $F(x) = 1 + \sinh x$ where $I = (\ln(\sqrt{2} - 1), 0)$
5. $F(x) = 2 - \cosh x$ where $I = (\ln(2 - \sqrt{3}), 0)$
6. $F(x) = 1 - x^2$ where $I = (-1, 0)$
7. $F(x) = \ln x$ where $I = (1, e)$
8. Uniform Distribution
9. $F(x) = (1 - \exp(-\lambda x))^k$ where $I = (0, \infty), \lambda > 0,\ k > 1$
10. $F(x) = 1 - \exp(-\lambda . e^x)$ where $I = (-\infty, \infty),\ \lambda > 0$
11. $F(x) = (1 + e^{-x})^{-k}$ where $I = (-\infty, \infty),\ k > 0$
12. $F(x) = 2^{-k}.(1 - \tanh(x))^k$ where $I = (-\infty, \infty),\ k > 0$
13. Logistic Distribution

$$F(x) = (1 + e^{-\lambda x})^{-1} \text{ where } I = (-\infty, \infty),\ \lambda > 0$$

14. Fisher Distribution (or z-distribution)

$$F(x) = C.\int_{-\infty}^{x} e^{nt}.(1 + k.e^{2t})^{-\alpha} dt,\ \text{where } I = (-\infty, \infty),\ n, n' \in \mathbb{N}\ k = \tfrac{n}{n'},$$

$$\alpha = \tfrac{n+n'}{2},\ C = 2.k^{\frac{n}{2}}.\Gamma(\alpha).\left[\Gamma(\tfrac{n}{2}).\Gamma(\tfrac{n'}{2})\right]^{-1}$$

15. Weibull Distribution when $\alpha > 1$

$$F(x) = 1 - \exp(-\lambda x^{\alpha}) \text{ where } I = (0, \infty),\ \lambda > 0$$

16. Extreme value Distribution

Global Journal of Science Frontier Research (F) Volume XXI Issue II Version I Year 2021

$$F(x)=\int_{-\infty}^{x}\exp(-t-e^{-t})dt \text{ where } I=(-\infty,\infty)$$

17. $F(x)=1-2[c.(1+e^{x})^{k}-c+2]^{-1}$ where $I=(-\infty,\infty),\ c>0, k=1,2$
18. Normal Distribution

$$F(x)=K.\int_{-\infty}^{x}\exp(-\tfrac{1}{2}.\sigma^{-2}.(t-m)^{2})dt, \text{ where } I=(-\infty,\infty),\ \sigma>0,\ K=\tfrac{1}{\sigma\sqrt{2\pi}}$$

19. Special Gamma Distribution

$$F(x)=K.\int_{0}^{x}t^{\alpha-1}\exp(-\lambda t)dt \text{ where } I=(0,\infty),\ \alpha>1,\ \lambda>0,\ K=\tfrac{\lambda^{\alpha}}{\Gamma(\alpha)}$$

20. Beta Distribution of the first kind

$$F(x)=C.\int_{0}^{x}t^{\alpha}(1-t)^{\beta}dt \text{ where } I=(0,1),\ \alpha,\beta>-1,\ C=\tfrac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}$$

21. $F(x)=C.\int_{-s}^{x}(1-\frac{t^{2}}{s^{2}})^{n}dt$ where $I=(-s,s),\ s>0,\ C=\left[s.\beta(\tfrac{1}{2},n+1)\right]^{-1}$

    where n is a positive integer.
22. Maxwell Boltzmann distribution

$$f(x)=\sqrt{\tfrac{2}{\pi}}\tfrac{1}{\sigma^{3}}x^{2}\exp\left(\tfrac{-x^{2}}{2\sigma^{2}}\right) \text{ where } I=(0,\infty),\ \sigma>0$$

23. $f(x)=2\lambda e^{-\lambda x(1-e^{-\lambda x})}$ where $I=(0,\infty)$
24. $f(x)=c(1-\frac{x^{2}}{a^{2}})^{n}$ where $I=(-a,a),\ a>0,\ n\in\mathbb{N}$ and $c=\frac{1}{a.B\left(\frac{1}{2},n+1\right)}$
25. Rayleigh Distribution $f(x)=\frac{x}{\sigma^{2}}\exp\left(\frac{-x^{2}}{2\sigma^{2}}\right)$ where $I=(0,\infty),\ \sigma>0$
26. $f(x)=\frac{x}{\sqrt{1-x^{2}}}$ where $I=(0,1)$
27. $f(x)=kx^{c-1}e^{-\frac{x^{2}}{2}}$ where $I=(0,\infty),\ c>1,$ and $k>0$
28. Reciprocal distribution

$$f(x)=\tfrac{\ln x-\ln a}{\ln b-\ln a} \text{ where } I=[a,b],\ a,b\in\mathbb{R},\ 0<a<b, \text{ and } \tfrac{a}{b}<e$$

29. $f(x)=c.\exp\left(\arctan\left(x\right)\right)$ where $c=\frac{1}{\int_{0}^{a}\exp\left(\arctan x\right)dx}$, and $I=(0,a),\ a>0$
30. $F(x)=1-x^{2}$ where $I=(-1,0)$

312. $F(x)=\ln x$ where $I=(1,e)$

32. Nakagami distribution

$$f(x)=\tfrac{2n^{n}}{\Gamma(n)\Omega^{n}}x^{2n-1}\exp\left(\tfrac{-n}{\Omega}x^{2}\right) \text{ where } I=(0,\infty)\ n>\tfrac{1}{2}, \text{ and } \Omega>0$$

Notes

2.2. The following probability distributions have decreasing RIFs:

1. $F(x) = 1 - x^{-\lambda}$, where $I = (1, \infty)$, and $\lambda > 0$
2. $F(x) = 1 - (\ln x)^{-\lambda}$, where $I = (e, \infty)$, and $\lambda > 0$
3. $F(x) = 1 - (\ln(\ln x))^{-\lambda}$, where $I = (e^e, \infty)$, and $\lambda > 0$
4. Weibull Distribution when $0 < \alpha < 1$

$$F(x) = 1 - \exp(-\lambda.x^{\alpha}),\ \text{where } I = (0, \infty),\ \text{and } \lambda > 0$$

5. $F(x) = 1 - a\exp(-b.x) - c.\exp(-d.x)$, where $I = (0, \infty)$, and $a, b, c, d > 0, a + c = 1$
6. $F(x) = 1 - \sum_{j=1}^{N} a_j \exp(-b_j.x)$, where $I = (0, \infty)$, and $a_j, b_j > 0, \sum_{j=1}^{N} a_j = 1$
7. Pareto Distribution of the third kind

$$F(x) = 1 - k.\exp(-b.x).x^{-a},\ \text{where } I = (k, \infty),\ \text{and } a, b, k > 0$$

8. Special Chi-Square Distribution

$$F(x) = K.\int_0^x t^{\frac{-1}{2}}.\exp(\tfrac{-t}{2})dt,\ \text{where } I = (0, \infty),\ \text{and } K = \frac{1}{\sqrt{2}\Gamma(\frac{1}{2})}$$

9. Pareto Distribution of the second kind

$$F(x) = 1 - x^{-k},\ \text{where } I = (1, \infty),\ \text{and } k > 0$$

10. Special Gamma Distribution

$$F(x) = K.\int_0^x t^{\alpha-1}.\exp(-\lambda t)dt,\ \text{where } I = (0, \infty),\ K = \frac{\lambda^{\alpha}}{\Gamma(\alpha)},\ \lambda > 0, \alpha < 1$$

11. $f(x) = \lambda x^{-1}(\ln x)^{-\lambda-1}$ where $I = (\exp, \infty)$, and $\lambda > 0$

2.3. The Exponential Distribution Function

$$F(x) = 1 - \exp[-\lambda(x - a)],\ \text{where } I = (a, \infty),\ \text{and } \lambda > 0$$

has a constant relative increment function.

2.4. The following probability distributions have RIFs that increase first and, having culminated, they decrease:

1. Cauchy Distribution

$$F(x) = \tfrac{1}{2} + \tfrac{1}{\pi}\tan^{-1} x,\ \text{where } I = (-\infty, \infty)$$

2. Inverse Gaussian Distribution

$$F(x) = \int_0^x (\tfrac{2\pi t^3}{\lambda})^{\frac{-1}{2}}.\exp(-\lambda.\tfrac{(t-m)^2}{(2m^2.t)})dt,\ \text{where } I = (0, \infty),\ \lambda > 0, m > 0$$

3. Lognormal Distribution

$$F(x) = \int_0^x \frac{1}{t\sigma\sqrt{2\pi}}\exp\{-\frac{(\ln t - \mu)^2}{2\sigma^2}\}dt,\ \text{where } I = (0, \infty),\ ,\ \sigma > 0, -\infty <$$

$\mu < \infty$

Notes

4. Beta type II distribution

$$f(x) = C\frac{x^p}{(1+x)^{p+q}} \text{ where } I = (0,\infty),\ p,q > 0,\ \text{and } C > 0$$

5. Burr type XII distribution

$$f(x) = ck\frac{x^{c-1}}{(1+x^c)^{k+1}} \text{ where } I = (0,\infty)\ c > 0,\ \text{and } k > 0$$

6. $f(x) = \frac{k}{1+x^4}$ where $I = \mathbb{R}$ and $k > 0$

7. $f(x) = cx^{-n}\exp\left(\frac{-k}{x}\right)$ where $I = (0,\infty)$ and $c,k > 0$

8. Frechet Distribution $F(x) = \exp\left(-(\frac{x-n}{t})^{-a}\right)$ where $I = (n,\infty),\ a,t \in (0,\infty),$ and $n = t\left(\frac{a}{a+1}\right)^{\frac{1}{a}}$.

9. Gumbel Distribution $F(x) = e^{-bx^{-a}}$ where $I = (0,\infty)$ and $a,b \in \mathbb{R}^+$ and $m = \sqrt[a]{\frac{ab}{a+1}}$

2.5. The following probability distributions have RIFs that decrease first and, having reached their minima, they increase:

1. $F(x) = 1 + \frac{2}{\pi}\arcsin x,$ where $I = (-1,0)$
2. $F(x) = \sqrt{x},$ where $I = (0,1)$
3. $F(x) = (1-x^2)^{\frac{1}{2}}$ where $I = (-1,0)$
4. Reciprocal distribution

$$f(x) = \frac{\ln x - \ln a}{\ln b - \ln a} \text{ where } I = [a,b],\ a,b > 0,\ \text{and } \frac{a}{b} > e$$

III. Numerical Examples

In this section, we illustrate the method described in section 2 by two examples.

1. The distribution of the the population of Botswana.
2. The distribution of people living with HIV globally.

Numerical Example 1: Botswana Population Growth. Below is a table of the population of Botswana from 1960 to 2012 in 5 year periods. This data was obtained from the World data bank.

Table 6.1: Botswana Population

x_k(*Time*)	g_k(Population)
5	579729
10	671416
15	793164
20	960807
25	1146205
30	1343440
35	1544865
40	1724924
45	1854739
50	1951715

We wish to find the probability distribution function of this sample.

Notes

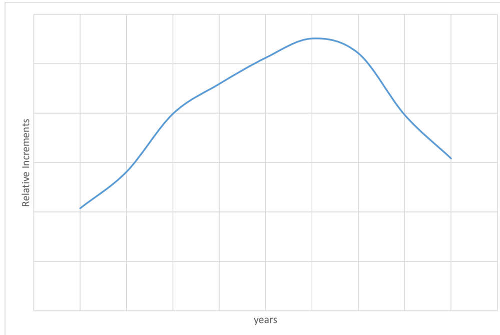


Figure 6.1: Graph of empirical relative increments

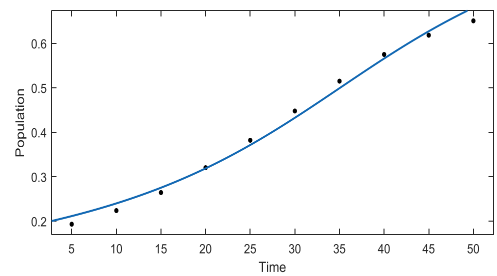
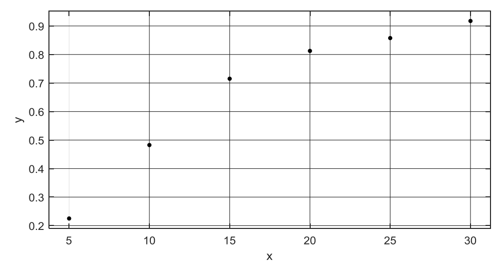
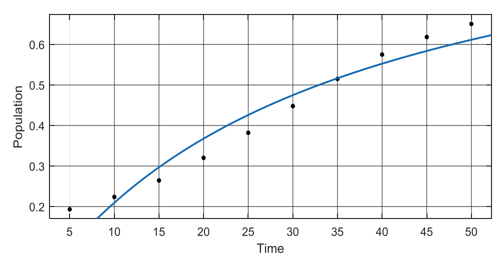
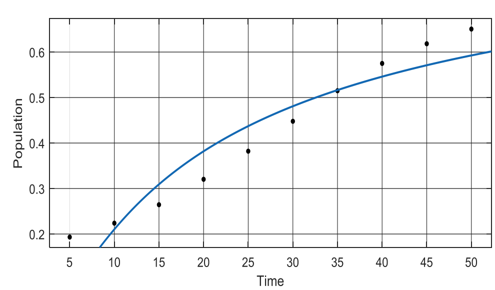
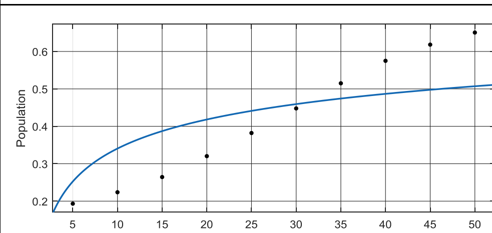
Figure 6.1 shows the empirical relative increment function of the distribution of the population. We see that the relative increments increase and then decrease. The distribution functions of section 2.4 display the same monotonic behaviour of first increasing and then decreasing. For these models, the values of B ranged between $24,900,000$ and $28,600,000$. We therefore picked $3,000,000$ as a reasonable estimate of the upper bound for all the models. Below is the table of time in years and the adjusted values of the population.

Table 6.2: Adjusted Population

$x_k(Time)$	y_k(Adjusted Population)
5	0.193243
10	0.223805
15	0.264388
20	0.320269
25	0.382068
30	0.447813
35	0.514955
40	0.574975
45	0.618246
50	0.650572

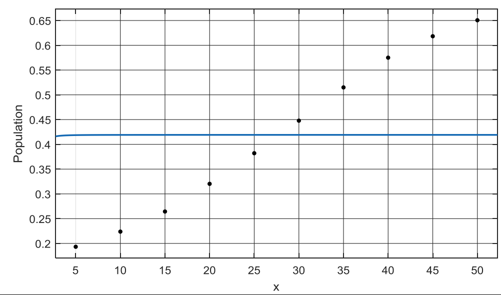
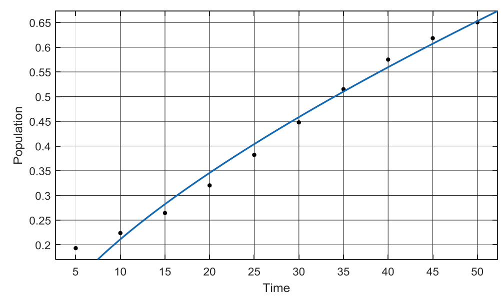
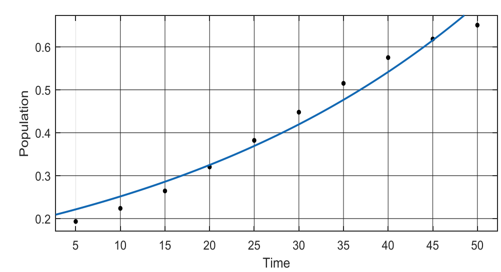
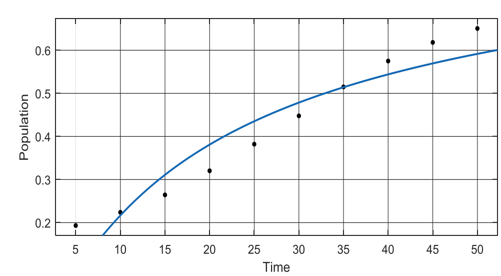
The adjusted population values were fitted to the distributions in section 6.3.4 as they exhibit the same monotonic behaviour of relative increment functions as that of our data. Table 6.2 shows the fitted values of the distributions.

Table 6.3: Fitted Models (Botswana Population)

Distribution function	Fit Results	Fitted Curve
Cauchy Distribution $\frac{1}{2}+\frac{1}{\pi}*\tan^{-1}\left(\frac{x-\alpha}{\beta}\right)$	Coefficients (with 95% confidence bounds): $p1\ =\ 35.06$ $p2\ =\ 23.53$ *Goodness of fit:* SSE: 0.002344	
Inverse Gaussian Distribution General model: $1/2\ \ (1+erf(\sqrt{(\lambda/2x)}\ (x/\mu-1))) + 1/2\ \ e\hat{}(2\lambda/\mu)\ (1 - erf(\sqrt{(\lambda/2x)}\ (x/\mu + 1)))$	Coefficients (with 95% confidence bounds): a = 0.3371 b = 0.1622 Goodness of fit: SSE: 1.012	
Lognormal Distribution $\frac{1}{2}*erfc\left(-\frac{\ln(x)-\mu}{\sigma\sqrt{2}}\right)$ Where $erfc$ is the complementary error function.	Coefficients (with 95% confidence bounds): $\sigma\ =\ 2.95$ $\mu\ =\ 3.494$ Goodness of fit: SSE: 0.01796	
Beta type II Distrribution $I_{\frac{x}{1+x}}(\alpha,\beta)$ Where $I_x(\alpha,\beta)$ is the incomplete beta function.	Coefficients (with 95% confidence bounds): $\alpha\ =\ 9.512$ $\beta\ =\ 0.5614$ Goodness of fit: SSE: 0.02997	
Burr type XII Distribution $1-(1+x^c)^{-k}$ Where $c,k>0$	Coefficients (with 95% confidence bounds): $c\ =\ 7.823$ $k\ =\ 0.02311$ Goodness of fit: SSE: 0.09004	

Notes

Notes

Number 6 $\frac{1}{8}k\sqrt{2}\left(\ln-\frac{x^2+x\sqrt{2}+1}{-x^2+x\sqrt{2}-1}\right.$ $+2\tan^{-1}(x\sqrt{2}+1)$ $\left.+2\tan^{-1}(x\sqrt{2}-1)\right)$	Coefficients (with 95% confidence bounds): $k\ =\ 0.6022$ Goodness of fit: SSE: 0.3504	
Number 7 $cx^{-n}x\left(-\frac{k\,\mathrm{Ei}\left(1,\frac{k}{x}\right)}{x}+\mathrm{e}^{-\frac{k}{x}}\right)$ Where Ei is the exponential Integral	Coefficients (with 95% confidence bounds): $c\ =\ 0.03116$ $k\ =\ -3.3$ $n\ =\ 0.1954$ Goodness of fit: SSE: 0.006749	
Frechet $e^{-\left(\frac{x-m}{s}\right)^{-a}}$ Where a is the shape parameter s is the scale parameter m is the location parameter	Coefficients (with 95% confidence bounds): $a\ =\ 2.002$ $s\ =\ 78.36$ Goodness of fit: SSE: 0.007987	
Gumbel $e^{-bx^{-a}}$ where a is real and b is the shape parameter	Coefficients (with 95% confidence bounds): $a\ =\ 0.6644$ $b\ =\ 7.06$ Goodness of fit: SSE: 0.02744	

Amongst these, the Cauchy distribution function,

$$F(x)=\tfrac{1}{2}+\tfrac{1}{\pi}\tan^{-1}\left(\tfrac{x-35.06959}{23.5291}\right)$$

provided the best least squares fit

$$\sum_{k=1}^{10}[F(x_k-y_k]^2\approx 0.002341.$$

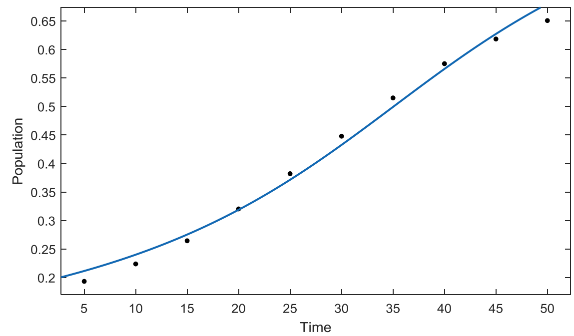


Figure 6.2: Graph of F and (x_k, y_k)

Figure 6.2 shows the fitted curve of the Cauchy model.

Numerical Example 2: Number of people living with HIV/AIDS (1990 to 2015): In the 1990s almost 3.5 million people were diagnosed with HIV every year. In 1997 the number declined and was reduced to 2.1 million in 2015. This table, obtained from UNAIDS data, shows the number of people living with HIV/AIDS from 1990 to 2015 in millions. The estimated value of B, which we adjust g_k by, gives the upper bound. For our models in section 2.4 the estimated values of B ranged between 36.82 and 39.77, we therefore picked 40 as a reasonable estimate of the upper bound for all the models.

Table 6.4

x_k(Time)	g_k(in Millions)(People living with HIV)
5	9
10	19.3
15	28.6
20	32.5
25	34.4
30	36.7

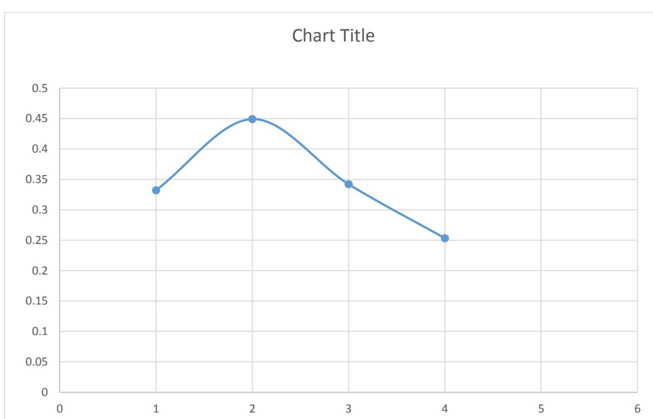


Figure 6.3: Graph of empirical relative increments

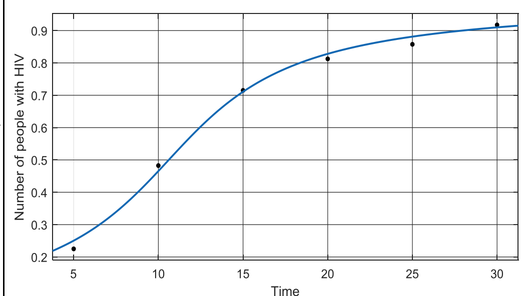
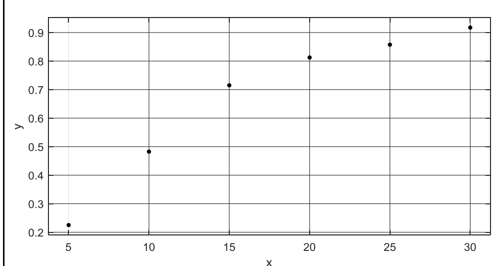
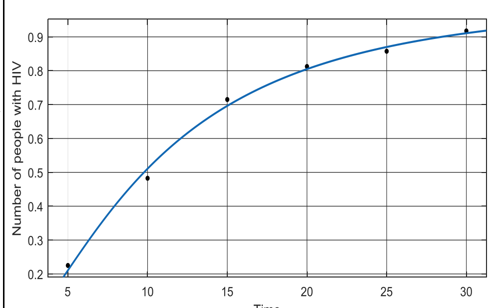
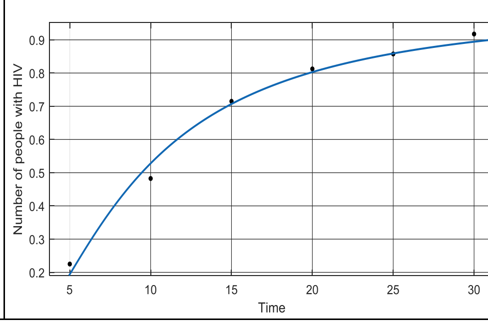
As we can see, the empirical relative increments increase up to a certain point and then decrease. Below is the table of distributions with the same behavior of relative Increments, the distributions in *section*6.3.4. The table shows the fitted values of the distributions.

Notes

Table 6.5

x_k(Time)	g_k(in Millions)(People living with HIV)	y_k (Adjusted g_k)
5	9	0.225
10	19.3	0.4825
15	28.6	0.715
20	32.5	0.8125
25	34.4	0.86
30	36.7	0.9175

Notes

Distribution function	Fit Results	Fitted Curve
Cauchy Distribution $\frac{1}{2}+\frac{1}{\pi}*\tan^{-1}\left(\frac{x-\alpha}{\beta}\right)$	Coefficients (with 95% confidence bounds): $\alpha\ =\ 10.61$ $\beta\ =\ 5.623$ Goodness of fit: SSE: 0.001816	
Inverse Gaussian Distribution General model: $1/2\ \ (1+erf(\sqrt{(\lambda/2x)}\ (x/\mu-1)))+1/2\ \ e^{\wedge}(2\lambda/\mu)\ (1-erf(\sqrt{(\lambda/2x)}\ (x/\mu+1)))$	Coefficients (with 95% confidence bounds): a = 0.3371 b = 0.1622 Goodness of fit: SSE: 1.012	
Lognormal Distribution $\frac{1}{2}*erfc\left(-\frac{\ln(x)-\mu}{\sigma\sqrt{2}}\right)$ Where $erfc$ is the complementary error function.	Coefficients (with 95% confidence bounds): $\sigma\ =\ 1.667$ $\mu\ =\ 2.28$ Goodness of fit: SSE: 0.001625	
Beta type II Distrribution $I_{\frac{x}{1+x}}(\alpha,\beta)$ Where $I_x(\alpha,\beta)$ is the incomplete beta function.	Coefficients (with 95% confidence bounds): $\alpha\ =\ 16.06$ $\beta\ =\ 1.986$ Goodness of fit: SSE: 0.003733	

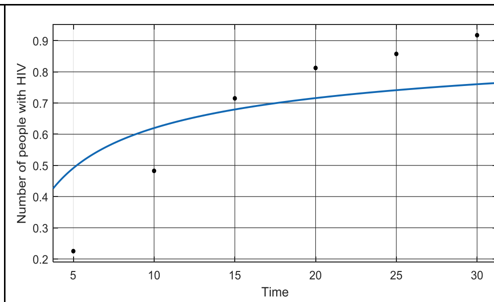
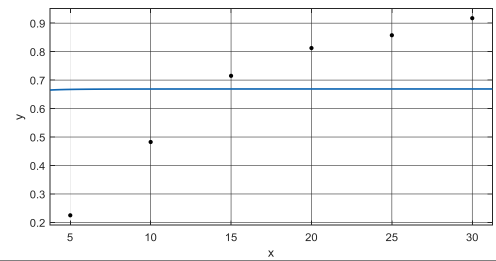
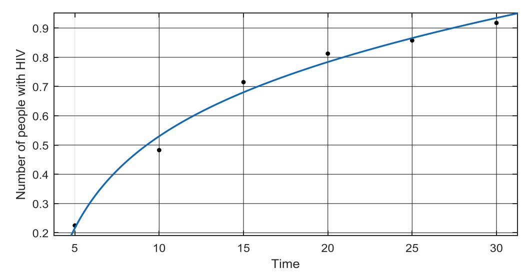
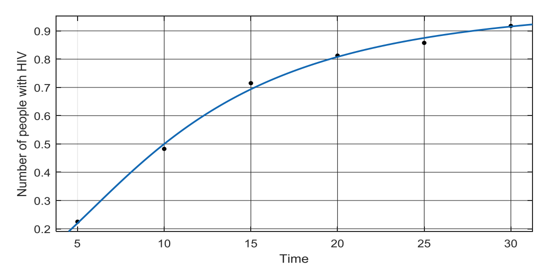
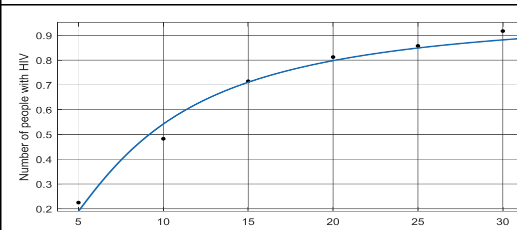
Burr type XII Distribution $1-(1+x^c)^{-k}$ Where $c,k>0$	Coefficients (with 95% confidence bounds): $c\ =\ 7.078$ $k\ =\ 0.05934$ Goodness of fit: SSE: 0.1386	
Number 6 $\frac{1}{8}k\sqrt{2}\left(\ln-\frac{x^2+x\sqrt{2}+1}{-x^2+x\sqrt{2}-1}\right)$ $+\,2\tan^{-1}(x\sqrt{2}+1)$ $+\,2\tan^{-1}(x\sqrt{2}-1))$	Coefficients (with 95% confidence bounds): $k\ =\ 0.6022$ Goodness of fit: SSE: 0.3504	
Number 7 $cx^{-n}x\left(-\frac{k\,\mathrm{Ei}\left(1,\frac{k}{x}\right)}{x}+e^{-\frac{k}{x}}\right)$ Where Ei is the exponential Integral	Coefficients (with 95% confidence bounds): $a\ =\ 0.218$ $k\ =\ 2.24$ $n\ =\ 0.5929$ Goodness of fit: SSE: 0.00472	
Frechet $e^{-\left(\frac{x-m}{s}\right)^{-a}}$ Where a is the shape parameter s is the scale parameter m is the location parameter	Coefficients (with 95% confidence bounds): $a\ =\ 3.506$ $m\ =\ -15.01$ $s\ =\ 22.53$ Goodness of fit: SSE: 0.001123	
Gumbel $e^{-bx^{-a}}$ where a is real and b is the shape parameter	Coefficients (with 95% confidence bounds): $a\ =\ 1.44$ $b\ =\ 16.85$ Goodness of fit: SSE: 0.006365	

Figure 6.4: Graph of F and (x_k, y_k)

Notes

Notes

Amongst these distributions , the Lognormal distribution function,

$$F(x) = K.\int_0^x \exp\frac{(-\frac{1}{2}.\sigma^2.(\ln t)^2)}{t}dt$$

provided the best least squares fit

$$\sum_{k=1}^{6}[F(x_k - y_k]^2 \approx 0.001625$$

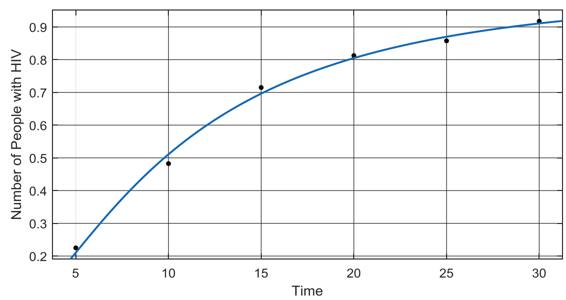


Figure 6.4: Graph of F and (x_k, y_k)

According to our model, the number of people with HIV/AIDS will not reach $40,000,000$.

IV. Discussions

This method is different from the classical methods of density estimation and kernel estimation because some of these methods do not provide indefinitely smooth models and some provide some twice differentiable models with much more than two points of inflection. In addition our model will give us tools to estimate the behaviour of the considered (natural, social and others) phenomena in the future if the environment and other conditions or circumstances do not change.This method cam only be used for relative increment functions with at most two phases. Further analysis of the algorithm to accommodate sample whose relative increment functions are more than two phased may prove useful to our work of finding continuous distributions. Results from this study sheds light on the use of relative increment functions in determining distributions which would be very helpful in forecasting and estimating percentiles of growth processes. By using indefinitely smooth model of growth processes, one can predict how the process in question will behave in the future.

References Références Referencias

1. Adler, P. and Szabo, Z.: The Suitability of Different Mathematical Models in Describing Cumulative Caries Prevalence Curves in Individual Teeth. European Meeting of Statisticians, Budapest, Proceedings in Colloquia Mathematica Soc. J. Bolyai (1972), pp. 43-49.
2. Adler, P. and Szabo, Z.: A Mathematical Description of Cumulative Caries Prevalence Curves of Individual Teeth with Regard to the Relative Growth of Caries. Dental Review 68 (1975), pp. 8-9.

3. Adler, P.-Szabo, Z.: Interpretation of a Cauchy distribution function describing cumulative caries prevalence curves. Community Dent. Oral Epid. 7 (1979) 82-86.
4. Adler, P.-Szabo, Z.: Comparison of two mathematical descriptions of the cumulative caries prevalence. Acta Medica Acad. Sci. Hung. 36 (1979) 211-222.
5. Adler, P.-Szabo, Z.: Mathematical description and use in forecasting of the caries curve. Community Dent. Oral Epid. 12 (1984) 96-102.
6. Carlos, J.P., Gittelsohn, A.M.: Longitudinal Studies of Natural History of Caries. II. A life-table study of caries incidence in permanent teeth. Arch. Oral Biol. (1965) pp. 739-751.
7. Johnson N.L. and Kotz, S.: Distributions in Statistics. Continuous Univariate Distributions, vol. 1, 2, Houghton Mi˙in, Boston - New York, 1970.
8. Knutson, J.W.: An Index of the Prevalence of Dental Caries in School Children. Pub. Hlth. Rep. 59, 253-263(1944).
9. Kooepile-Reikeletseng, S. Szabo, Z. I.: Monotonic Properties of Relative Increments of Cumulative Distribution Functions with some Applications in Modelling Growth Processes. Presented at IASTED 2016 African Conferences Modelling and Simulation (AfricaMS 2016) in September 5-6, 2016 in Gaborone, Botswana.
10. Kooepile-Reikeletseng, S. N. Szabo, Z. I.: Applications of Relativ e Increments of Cumulative Distribution Functions. Presented at TAAG 2017 Conference held at University of Botswana, Sep 18-20, 2017 in Gaborone, Botswana.
11. Kooepile-Reikeletseng, S.N.: Monotonic Behaviour of Relative Increments of Pearson Distributions. Global Journal of Science Frontier Research: Mathematics and Decision Sciences. 18. Issue 5. Version 1.0(2018). Online ISSN: 2249-4626, pp.1-6.
12. Porter, D.R. and Dudman, J.A.: Assessment of Caries Increments I. Constitution of the R.I.D. Index. J. Dent.Res. 39(1960),1056-1061.
13. Sobkowiak, E.M., Szabo,Z, Radtke, G. and Adler, P: Analysis of Cummulative Caries Prevalance in Primary Molar . Commumunity Dent. Oral Epidemiol.1973,1,1, pp.127-131.
14. Silverman, B.W.: Density estimation for statistics and data analysis. Pub. Monographs on Statistics and Applied Probability, London: Chapman and Hall, 1986.
15. Szabo, Z.: On Relative Growth of Distribution Functions. Biometrics Journal 18 (1976), pp. 154-155
16. Szabo Z.: Relative Increments of Distribution Functions and Their Use in Biometrics, Joint Meeting of American Statistical Association at Washington, D.C. Proceedings, Biopharmaceutical Section, 1989, pp. 212-216.
17. Szabo, Z.: Investigation of Relative Increments of Distribution Functions. Publ. Math. Debrecen 49 (1996), pp. 99-112.

Notes

18. Szabo, Z.: Relative Increments of Pearson Distributions. Acta Math. Aca. Paed. Nyiregyh. 15 (1999), pp. 45-54. [Electronic Journal, websites www.bgytf.hu/amapn, www.emis.de/journals]
19. Szabo Z.: On Monotonic Behaviour of Relative Increments of Unimodal Distributions. Acta Math. Acad. Paed. Nyiregyh. 22 (2006), pp. 113-119. [Electronic Journal, website www.emis.de/journals]
20. Szabo Z. Algorithm for Finding the Proper Continuous Distribution Function to a Unimodal Empirical Distribution Function. Global Jounal of Science Frontier Research. 10. (2010) pp. 20-26.

Notes