A SHAPE-PRESERVING AREA-TRUE APPROXIMATION OF HISTOGRAM

Rep. Fac. Sci.,Kagoshima Univ., (Math., Phys., & Chem.),
No.21, p 1 -12, 1988.

A SHAPE-PRESERVING AREA-TRUE APPROXIMATION OF HISTOGRAM

M. ATAKI* and M. SAKAI*

(Received September 1, 1988)

Abstract.

We shall consider an application of a simple rational spline to a shape-preserving area-true approximation of a histogram. Our method is situated between the one on a polynomial quadratic spline and on a piecewise linear function. Some numerical examples are given at the end of this paper.

1. Introduction and Description of method

We are concerned with a smoothing of a histogram by a simple rational spline. For the histogram with a height h_i on an interval $[x_i, \ x_{i+1}]$ $(0 \leqq i \leqq n-1)$, we demand that a smooth (at least continously differentiable) function f satisfies the area true conditions:

$$\int_{x_i}^{x_{i+1}} f(x)\,dx = h_i\,\Delta x_i \quad (0 \leq i \leq n-1) \tag{1}$$

where $\Delta x_i = x_{i+1} - x_i$.
Usually we choose f to be quadratic or quartic splines from a computational point of view since then the coefficient matrices of the determination of them are diagonally dominant. However, they don't always preserve the shape (for example, monotonicity) of the histogram.

Now by making use of the following rational spline, we consider a shape-preserving area-true approximation of the histogram. For $p_i > -1$ $(0 \leqq i \leqq n-1)$, the spline s is defined by

* Department of Mathematics, Faculty of Science, Kagoshima, University, Kagoshima, 890, Japan.

(i) $s(x) \in C^1 [x_0, x_n]$

(ii) $s(x)$ is a linear combination of 1, x, $t^2/(1+p_i t)$ ($t = (x-x_i)/\Delta x_i$) on $[x_i, x_{i+1}]$.

Then, by a simple calculation we have

$$\phi(p_i)\ \Delta x_i\ s'_{i+1} + [\{1/2 - \phi(p_i)\}\ \Delta x_i + \{\frac{1+p_{i-1}}{2+p_{i-1}} - \phi(p_{i-1})\ \Delta x_{i-1}]\ s'_i + \{\phi(p_{i-1}) - \frac{p_{i-1}}{2(2+p_{i-1})}\}\ \Delta x_{i-1}\ s'_{i-1}$$
$$= \frac{1}{\Delta x_i} \int_{x_i}^{x_{i+1}} s(x)\,dx - \frac{1}{\Delta x_{i-1}} \int_{x_{i-1}}^{x_i} s(x)\,dx \qquad (2)$$

where $s_i' = s'(x_i)$ and $\phi(p) = \frac{(1+p)^2}{(2+p)} \{\frac{1}{2p} - \frac{1}{p^2} + \frac{\log(1+p)}{p^3}\}$.

Since s depends upon $n+2$ parameters, there are two additional conditions to the area true ones (1) required for a unique determination of it. Here we take these to be end ones :

$$s'_0 = \alpha, \quad s'_n = \beta. \qquad (3)$$

For $p_i = p$ $(0 \leq i \leq n-1)$, we have

Theorem 1. *Suppose that* $\Delta h_0 \leq \Delta h_1 \leq \cdots \leq \Delta h_{n-2} < 0$, $\alpha \leq 0$ *and* $2\Delta h_{n-2}/\Delta x_{n-1} \leq \beta \leq 0$. *Then the spline under* (1) *and* (3) *is decreasing on* $[x_0, x_n]$ *for sufficiently large* p.

By replacing h_i with h_{n-1-i} (*or* $-h_{n-1-i}$), x with $-x$, p with q $(=p/(1+p))$ and swapping α, β, from above we have

Corollary. *Suppose that* $0 < \Delta h_0 \leq \Delta h_1 \leq \cdots \leq \Delta h_{n-2}$, $0 \leq \alpha \leq 2\Delta h_0/\Delta x_0$ *and* $0 < \beta$ (*or* $\Delta h_{n-2} \leq \Delta h_{n-3} \leq \cdots \leq \Delta h_0 < 0$, $2\Delta h_0/\Delta x_0 \leq \alpha \leq 0$ *and* $\beta \leq 0$). *Then the spline under* (1) *and* (3) *is increasing* (*or decreasing*) *on* $[x_0, x_n]$ *for sufficiently large* q.

Next we consider the case when $\Delta h_0 \leq \Delta h_1 \leq \cdots \leq \Delta h_{k-1} < 0 < \Delta h_k \leq \cdots \leq \Delta h_{n-2}$. (4)

Then, for $p_i = p$ $(0 \leq i \leq k-1)$, $p_k = 0$ and $p_i = q$ $(= -p/(1+p))$ we have

Theorem 2. *Suppose that* (4), $\alpha \leq 0$ *and* $0 \leq \beta$. *If*

$$-\frac{(\Delta x_k+2\Delta x_{k+1})}{\Delta x_{k-1}}<\frac{\Delta h_k}{\Delta h_{k-1}}<-\frac{\Delta x_{k+1}}{(\Delta x_k+2\Delta x_{k-1})},$$

then the spline under (1) *and* (3) *is decreasing on* $[x_0, c]$ *and increasing on* $[c, x_n]$ *for sufficiently large* p, *where* $c\in(x_k, x_{k+1})$.

2. Proof of Theorems

Before we proceed with analysis, we shall require the following simple lemma.

Lemma. *Let s be a linear combination of* 1, x, $x^2/(1+px)$. *Then* $s'(x)\leq 0$ *on* $[0, 1]$ *if and only if* $s'(0)$, $s'(1)\leq 0$.

Proof. We only have to notice the equation which is easily obtained by a direct calculation :

$$\begin{aligned} s'(x) &= s'(0)+\frac{(2x+px^2)\ (1+p)^2}{(1+px)^2\ (2+p)}\{s'(1)-s'(0)\} \\ &= \{1-\psi(\mathrm{x})\}s'(0)+\psi(\mathrm{x})s'(1) \end{aligned} \tag{5}$$

where

$$0\leq\psi(x)\ \left(=\frac{(2x+px^2)\ (1+p)^2}{(1+px)^2\ (2+p)}\right)\leq 1 \tag{6}$$

Now, from (1)– (3) we have a system of linear equations in $s'_i\ (p)$ ($=s'(x_i)$ with the parameter p) whose coefficient matrix is diagonally dominant for any $p>-1$. Therefore, $s'_i\ (p)$ is a continous function of p on $(-1,\infty)$. Letting p be $+\infty$, we have

$$\begin{aligned} &\text{(i)}\ \ s'_0(+\infty)=\alpha \\ &\text{(ii)}\ \ \Delta x_i s'_{i+1}(+\infty)+\Delta x_{i-1}s'_i(+\infty)=2\Delta h_{i-1}\ \ (1\leq i\leq n-1) \\ &\text{(iii)}\ \ s'_n(+\infty)=\beta \end{aligned} \tag{7}$$

If $2\Delta h_{n-2}/\Delta x_{n-1}<\beta<0$, from above we have

$$s'_i(+\infty)<0\ \ (1\leq i\leq n-1) \tag{8}$$

from which follows

$$s'_i(p)<0\ \text{ for } p>p_0(1\leq i\leq n-1) \tag{9}$$

where, p_0 means a sufficiently large generic constant.
If $2\Delta h_{n-2}/\Delta x_{n-1}=\beta$, we only have

$$2\Delta h_{i-1}/\Delta x_{i-1}\leq s'_i\ (+\infty)\ \leq 0\ \ (1\leq i\leq n-1). \tag{10}$$

In this case, a detailed analysis is required to show the desired inequalities (9). Now, letting $\Delta x_{i-1}s'_i\ (+\infty)$ by d_i, from (2) we have the following asymptotic relation :

$$d_i = 2\Delta h_{i-1} - \Delta h_{i+1} + (2/p)\{1 + (\Delta x_i/\Delta x_{i-1})d_{i+1} - 2(\Delta x_i/\Delta x_{i-1})\Delta h_{i-1}\} + O(1/p^2)\quad (1 \leq i \leq n-1). \tag{11}$$

Hence we have

$$\begin{aligned} &\text{(i)}\ d_{n-1} = (4/p)c_{n-1} + O(1/p^2) \\ &\text{(ii)}\ d_{n-2} = 2\Delta h_{n-3} - (4/p)c_{n-2} + O(1/p^2) \end{aligned} \tag{12}$$

where $c_{n-1} = \Delta h_{n-2}/\Delta x_{n-2}\ (<0)$ and $c_{n-2} = c_{n-1} + (\Delta x_{n-2}/\Delta x_{n-3})\Delta h_{n-3}\ (<0)$. Hence, we have

$$s'_{n-1}(p),\quad s'_{n-2}(p) < 0 \ \text{ for } p > p_0. \tag{13}$$

If $\Delta h_{n-4}/\Delta h_{n-3} <$, by 12 (ii)

$$2\Delta h_{n-4}/\Delta x_{n-3} < s'_{n-2}\ (+\infty) < 0 \tag{14}$$

from which follows by 7 (ii) and

$$s'_i\ (+\infty) < 0\quad (1 \leq i \leq n-3). \tag{15}$$

Thus we have

$$s'_i(p) < 0 \ \text{ for } p > p_0\quad (1 \leq i \leq n-3). \tag{16}$$

If $\Delta h_{n-4} = \Delta h_{n-3}$, from (ii) we have

$$\begin{aligned} &\text{(i)}\ d_{n-3} = (4/p)c_{n-3} + O(1/p^2) \\ &\text{(ii)}\ d_{n-4} = 2\Delta h_{n-5} - (4/p)c_{n-4} + O(1/p^2) \end{aligned} \tag{17}$$

where $c_{n-3} = c_{n-2} + \Delta h_{n-3}\ (<0)$ and $c_{n-4} = c_{n-3} + (\Delta x_{n-4}/\Delta x_{n-5})\Delta h_{n-5}\ (<0)$. Hence, we have

$$s'_{n-3}(p),\quad s'_{n-4}(p) < 0 \ \text{ for } p > p_0. \tag{18}$$

By continuing these processes of mathematical induction, we have the desired inequalities (9).

For $\beta = 0$, as in the proof for the above case when $\beta = 2\Delta h_{n-2}/\Delta x_{n-1}$, we have the desired inequalities (9), by making full use of the asymptotic relation (12). Combining Lemma and (9), we have the complete proof of Theorem 1.

Next, we shall prove Theorem 2. Letting p be $+\infty$ in (2) $(i=k,\ k+1)$, we have

$$\text{(i)}\ \ (2\Delta x_k + 3\Delta x_{k-1})\, s'_k(+\infty) + \Delta x_k s_{k+1}(+\infty) = 6\Delta h_{k-1}$$
$$\text{(ii)}\ \ \Delta x_k s'_k(+\infty) + (2\Delta x_{k+1})\, s'_{k+1}(+\infty) = 6\Delta h_k\,. \tag{19}$$

From above, we have

$$\text{(i)}\ \ s'_k(+\infty) = (6/D_k)\,\{(2\Delta x_k + 3\Delta x_{k+1})\ \Delta h_{k-1} - \Delta x_k \Delta h_k\}$$
$$\text{(ii)}\ \ s'_{k+1}(+\infty) = (6/D_k)\,\{(2\Delta x_k + 3\Delta x_{k-1})\ \Delta h_k - \Delta x_k \Delta h_{k-1}\} \tag{20}$$

where

$$D_k = (2\Delta x_k + 3\Delta x_{k-1})(2\Delta x_k + 3\Delta x_{k+1}) - (\Delta x_k)^2.$$

Similarly as in the proof of Theorem 1, if

$$\text{(i)}\ \ 2\Delta h_{k-2}/\Delta x_{k-1} < s'_k(+\infty) < 0$$
$$\text{(ii)}\ \ 0 < s'_{k+1}(+\infty) < 2\Delta h_{k+1}\ (+\infty) < 2\Delta h_{k+1}/\Delta x_{k+1}, \tag{21}$$

we have for $p > p_0$

$$\text{(i)}\ \ s'_i(p) < 0 \qquad (1 < i < k-1)$$
$$\text{(ii)}\ \ s'_i(p) > 0 \qquad (k+2 < i < n-1). \tag{22}$$

Since $\Delta h_{k-2} < \Delta h_{k-1}$ and $\Delta h_k < \Delta h_{k+1}$, conditions 21 (i) – (ii) may be replaced by

$$\text{(i)}\ \ 2\Delta h_{k-1}/\Delta x_{k-1} < s'_k(+\infty) < 0$$
$$\text{(ii)}\ \ 0 < s'_{k+1}(+\infty) < 2\Delta h_k/\Delta x_{k+1}. \tag{23}$$

By a simple calculation, 23 (i) – (ii) are equivalent to the following inequalities :

$$-\frac{(\Delta x_k + 2\Delta x_{k+1})}{\Delta x_{k+1}} < \frac{\Delta h_k}{\Delta h_{k-1}} < \frac{-\Delta x_{k+1}}{(\Delta x_k + 2\Delta x_{k-1})}. \tag{24}$$

Hence, under (24) we have

$$s'(x) < 0 \quad (x_0 < x < x_k),\quad s'(x) > 0 \quad (x_{k+1} < x < x_n) \ \ \textit{for}\ p > p_0. \tag{25}$$

Since s is a quadratic polynomical on $[x_0,\ x_n]$, there exists a constant c where s' changes its sign from $-$ to $+$.

This completes the proof of Theorem 2.

Finally we notice that our spline is represented in terms of $R_{3,p}$ in the case when $x_i=i$ and $p_i=p$.
Let χ be an indicator of $(0, 1]$ and $\phi_{m,p}$ (x) be a simple rational function:

$$\phi_{m,p}(x)=\begin{cases} c_m(p)/(1+px)^{m+1} & (0<x\leq 1) \\ 0 & (\text{otherwise}) \end{cases} \tag{26}$$

where $c_m(p)$ is determined by

$$\int_0^1 \phi_{m,p}(x)\,dx=1 \tag{27}$$

Then, $R_{m+1,p}(x)$ is defined by

$$R_{m+1,p}(x)=(\underbrace{\chi*\chi*\cdots*\chi}_{m}*\phi_{m,p})(x) \tag{28}$$

Remark. At the end of this Section, let $x_i=i$. Then we may also consider another area-true shape preserving approximation of the histogram by use of $R_{4,p}$ ([2]).
That is, we take a spline s of the form:

$$s(x)=\sum_{i=-3}^{n-1}\alpha_i R_{4,p}(x-i) \tag{29}$$

with undermined coefficients $(\alpha_{-3}, \alpha_{-2}, \cdots, \alpha_{n-1})$ so that

$$s(i)=\sum_{j=1}^{i-1}h_j \quad (0<i<n) \tag{30}$$

for any constant h_{-1}.
Then, s' is an area-true shape preserving approximation of the histogram with the height h_i over $[i, i+1]$ for p sufficientry large or close to -1 according to the shape of it. For $e=f-s'$, we have

$$\begin{aligned}&\frac{1}{2(3+3p+p^2)}\{(1+p)^2e_i+e_{i+1}+(2+p)^2e_{i-1}\}\\ &=\frac{p^2}{12(3+3p+p^2)}f_i''+\cdots\end{aligned} \tag{31}$$

where $h_i=\int_i^{i+1}f(x)\,dx \quad (0\leq i\leq n-1)$.

On the other hand, by a simple calculation we have the following consistency relation for $R_{3,p}$:

$$\phi(p)s_{i+1}+\{\frac{4+3p}{2(2+p)}-2\phi(p)\}s_i+\{\phi(p)-\frac{p}{2(2+p)}\}s_{i-1}$$
$$=(1/2)\left(\int_i^{i+1}s(x)\,dx+\int_{i-1}^{i}s(x)\,dx\right). \tag{32}$$

Hence, we have

$$\phi(p)\bar{e}_{i+1}+\{\frac{4+3p}{2(2+p)}-2\phi(p)\}\bar{e}_i+\{\phi(p)-\frac{p}{2(2+p)}\}\bar{e}_{i-1}$$
$$=\{\phi(p)-\frac{4+5p}{12(2+p)}\}f_i''+\cdots \tag{33}$$

where $\bar{e}=f-s$.

Here, by a simple calculation and a numerical computation, we see that the ratio of the coefficient of f_i'' in (33) and the one in (31) would go from $1/5$ to 1 as p goes from 0 to $+\infty$ (or $-1+$), where the coefficients in (33) and (31) are given by $p^2/60+O(p^3)$ and $p^2/12+O(p^3)$, respectively. For example, the above ratio is about $0.6\sim0.8$ at $p=1\sim10$. Therefore, our method would be superior to the one with $R_{4,p}$ if we would use the same value of p.

3. Numerical Illustration

For a determination of an appropriate initiial value of p, we shall show the following inequality that implies a positiveness of the coefficient of f_i'' in (33) :

$$\frac{(1+p)^2}{(2+p)}\{\frac{1}{2p}-\frac{1}{p^2}+\frac{\log(1+p)}{p^3}\}>\frac{4+5p}{12(1+p)}$$
$$(p>-1,\ p\neq0). \tag{34}$$

This is equivalent to

$$(1+p)^2\log(1+p)>p+3p^2/2+p^3/3-p^4/12 \qquad (p>0) \tag{35}$$

where for $-1<p<0$, we take the reverse inequality in (35).
Here, let us denote a difference of the left hand side and the right one of (35) by $2g(p)$. Then

$$g'(0)=g''(0)=0,\ g^{(3)}(p)=p^2/(1+p) \tag{36}$$

from which follows

$$g(p) = (1/3!)\, g^{(3)}(\xi) \quad (0 \lesseqgtr \xi \lesseqgtr p). \tag{37}$$

This completes the proof of (34).

Since the coefficient of f_i'' in (33) gets its minimum at $p=0$ and then the method is of order 3, in practical computation it would be sufficient to increase (or decrease) the parameter p, starting at zero, until the curve is satisfactory. The last six examples are for the histograms given by $f(x)=e^{rx}$ $(r>-1)$. Especially the last three ones (FIGS 7 −9) show that the method is reliable and efficient even for not so well behaved histograms obtained from integrals in which integrands have endpoint singularities.

References

[1] C. DEBOOR. A Practical Guide to Splines, Springer-Verlag, New York, 1978.
[2] M. SAKAI & M. C. LO'PEZ de SILANES. A simple rational spline and its application to monotonic interpolation to monotonic data. Numer. Math., **50**, 171-182 (1987).

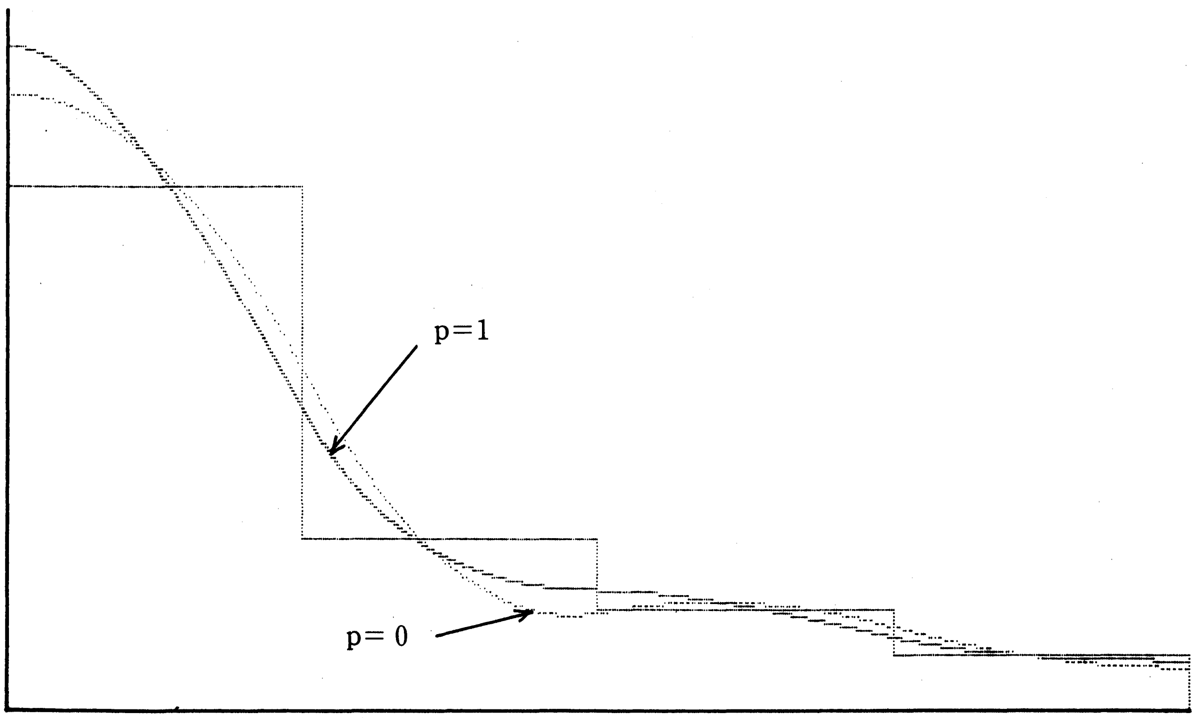

FIG. 1

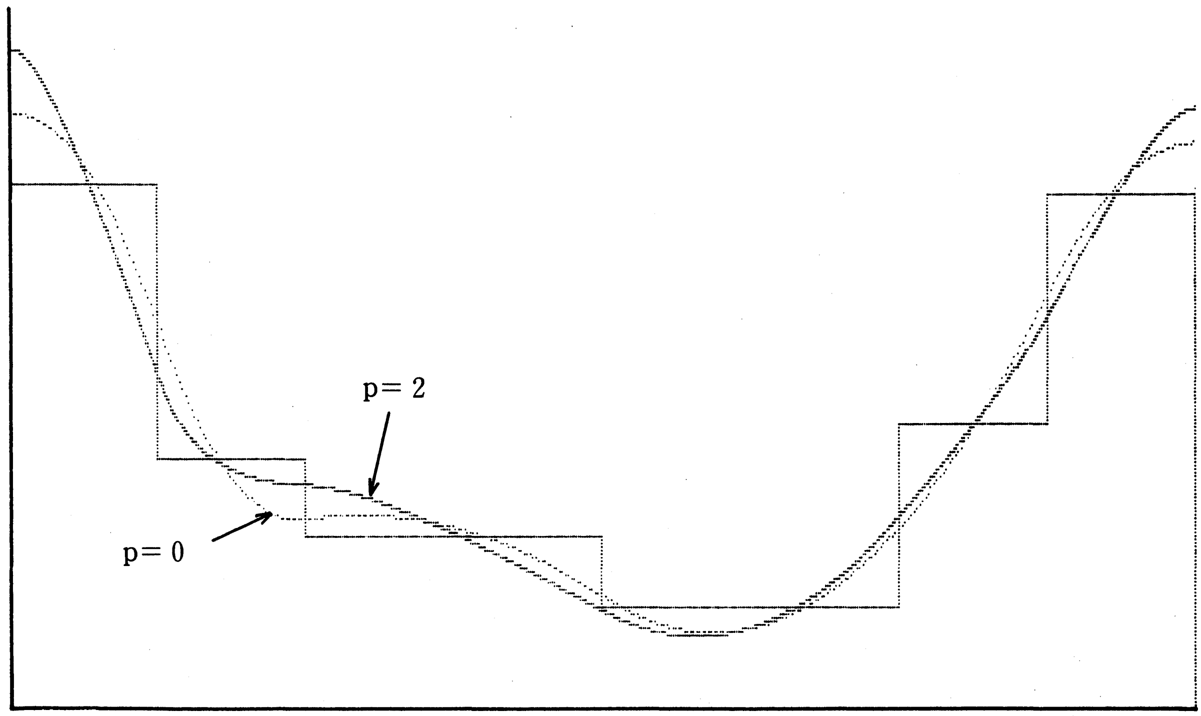

FIG. 2

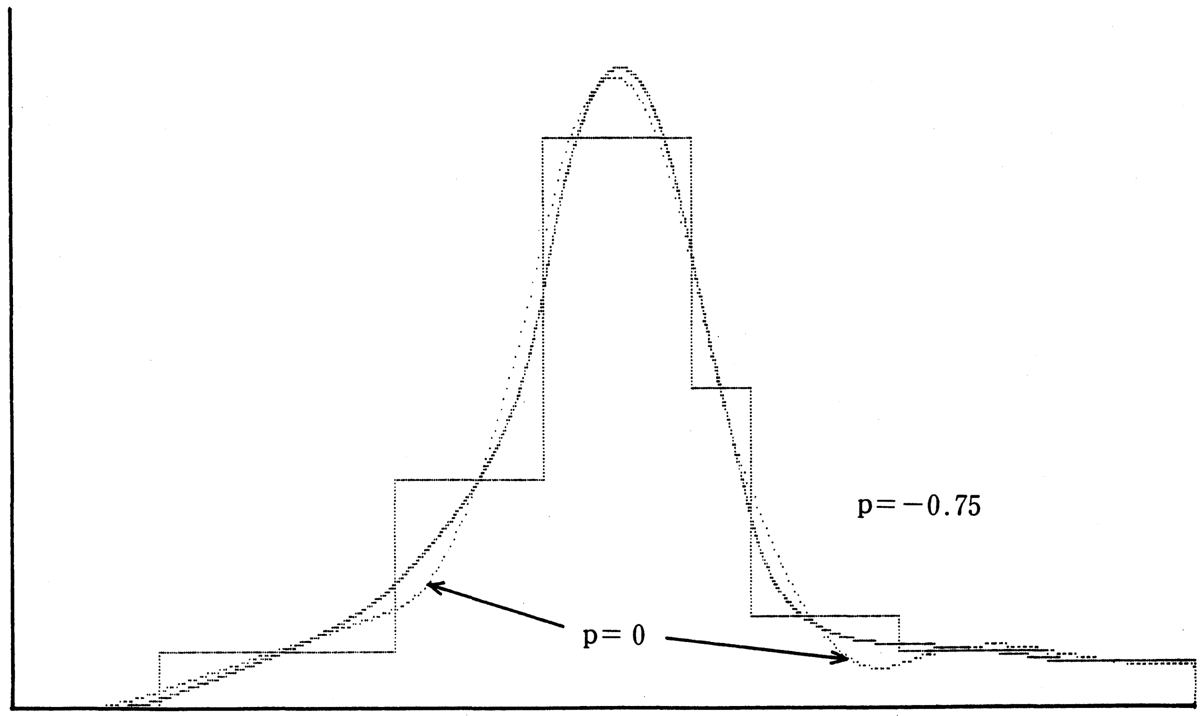

FIG. 3

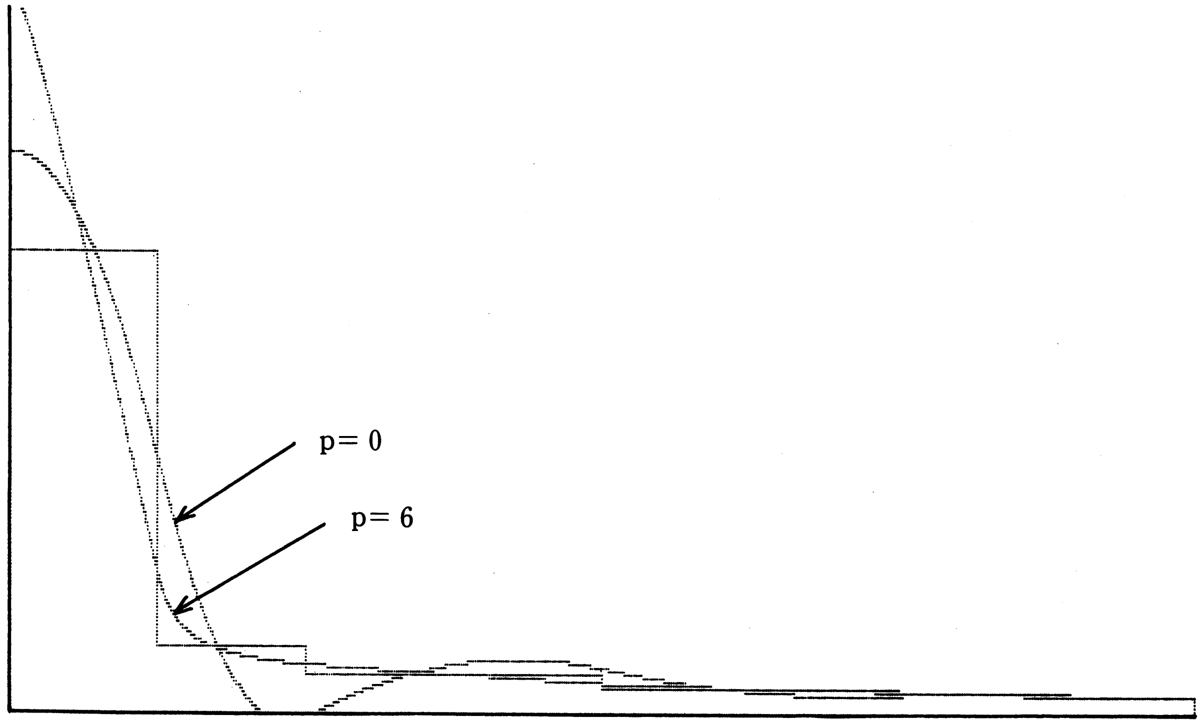

FIG. 4 $(f(x)=x^{-0.8})$

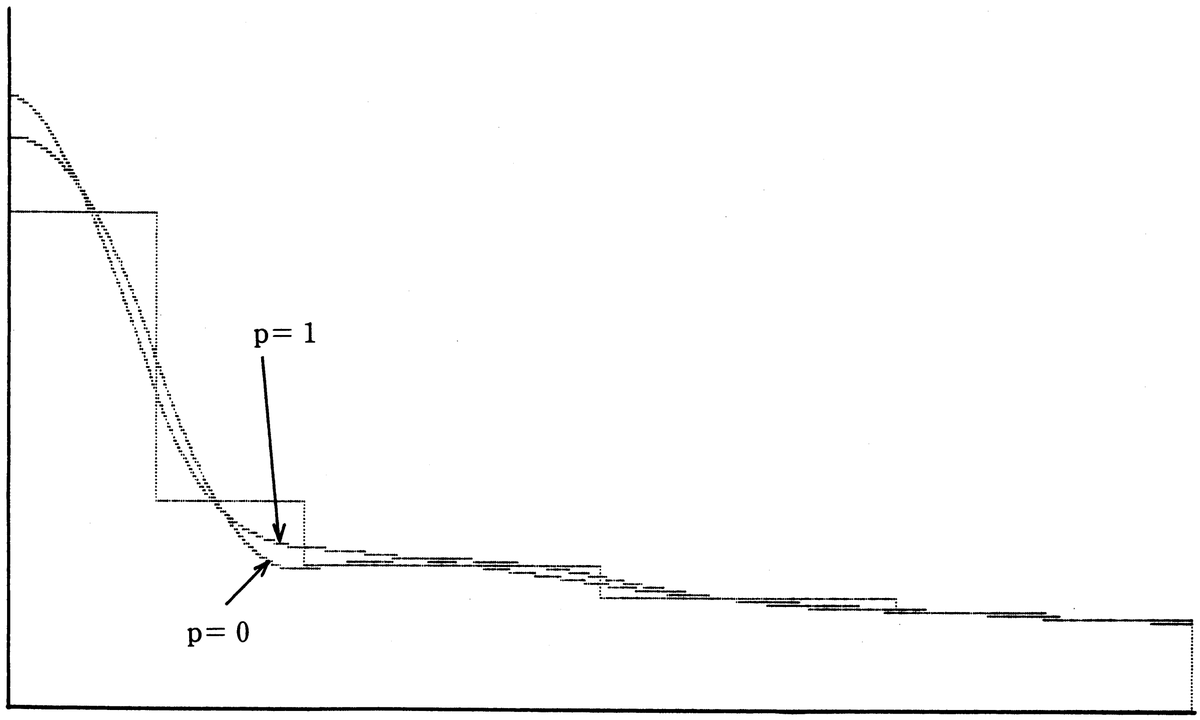

FIG. 5 $(f(x)=x^{-0.5})$

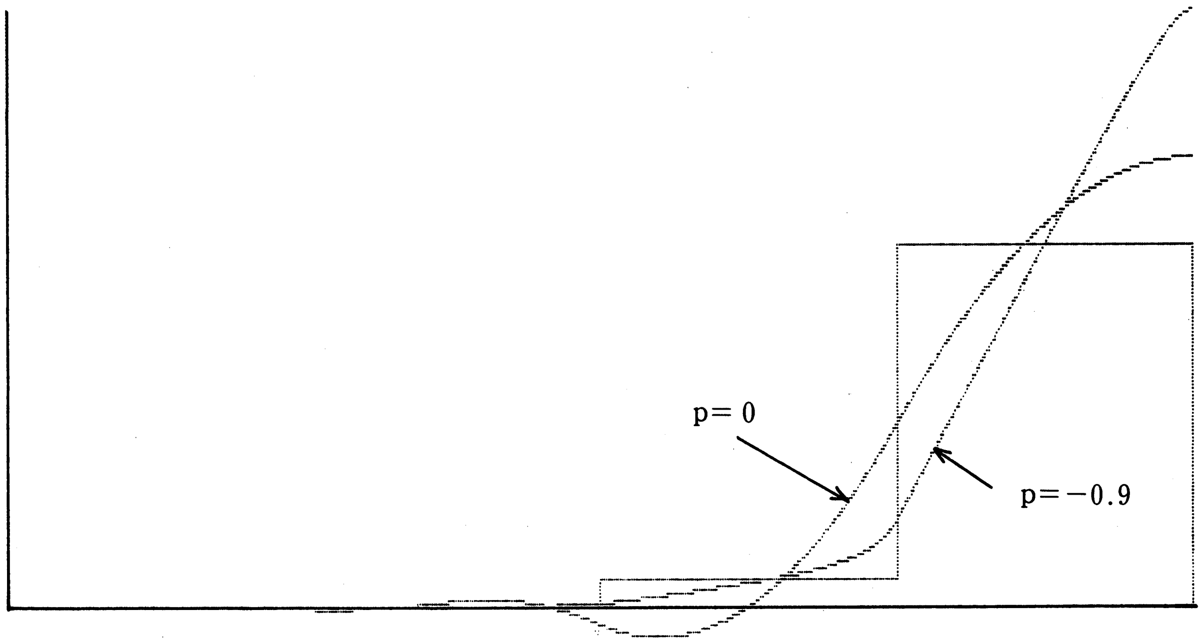

FIG. 6 $(f(x)=x^8)$

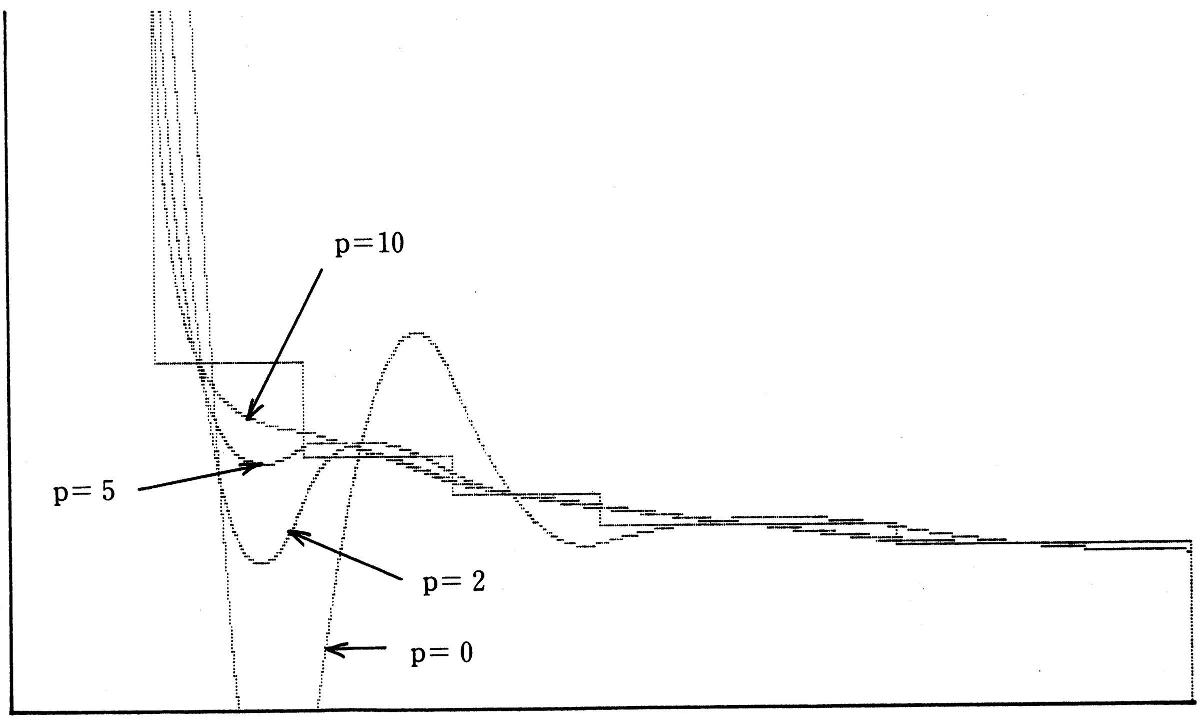

FIG. 7 $(f(x)=x^{-0.9})$

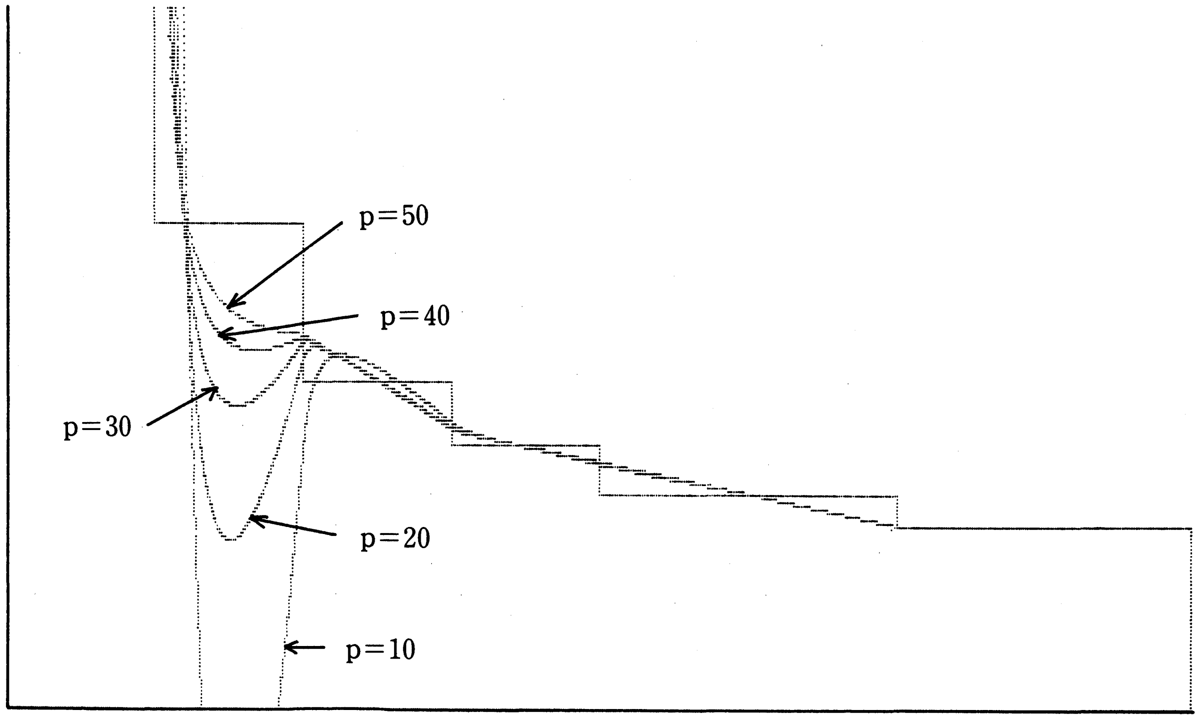

FIG. 8 $(f(x)=x^{-0.99})$

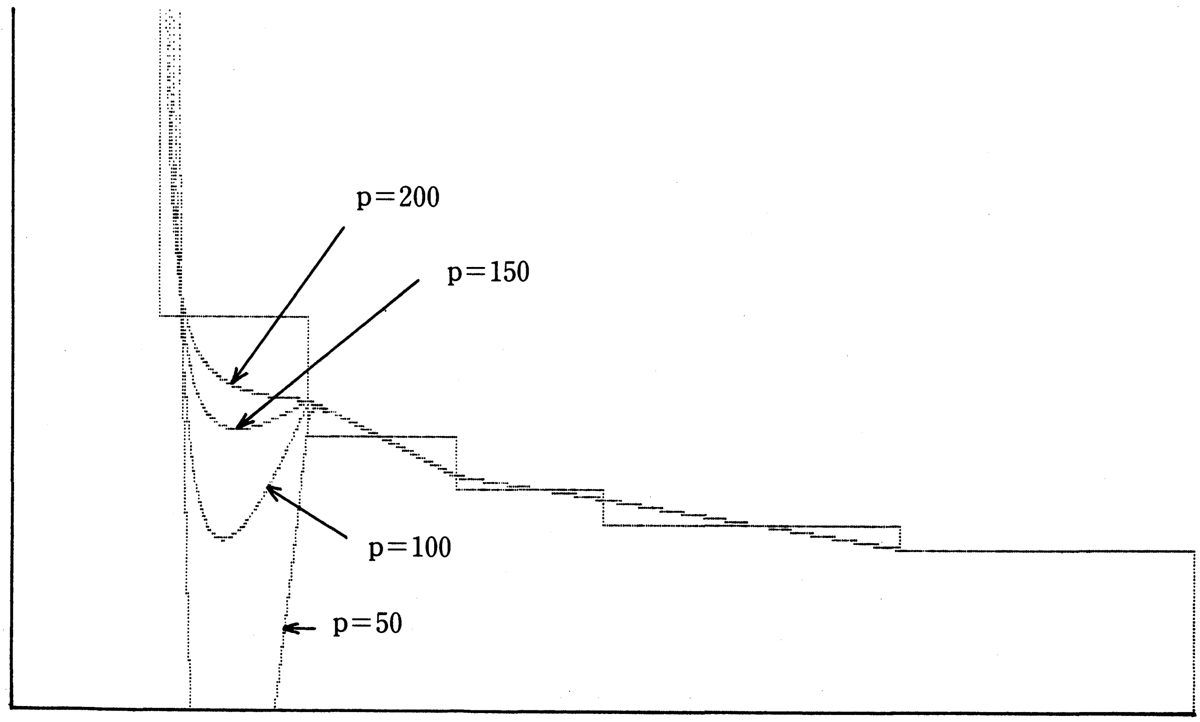

FIG. 9 $(f(x)=x^{-0.999})$