

高橋・森の DE 変換公式と ある積分表示された
関数に適用し その関数近似式を求めること

千葉大・工 戸田英雄 (Hideo Toda)

都立農芸高 小野令美 (Hatumi Ono)

ある統計分布関数の確率積分の数値計算に DE 変換公式
が非常に有効で、これを土台にして分布関数の最小 2 乗近
似式を導いた数値例について報告し、高橋秀俊先生の業績
を思い、私どもはここに 今は亡き先生に感謝の気持ちを
捧げる。

1. 正規分布からの独立な大きさ $n=4$ の標本歪度
 $\sqrt{b_1}$ の確率分布について

$(x_1, x_2, \dots, x_n)$ を標準正規分布 $N(0, 1)$ からの
独立な大きさ n の標本とする。このとき 次の統計量を

$$\bar{x} = \sum_{i=1}^n x_i / n, \quad m_2 = \sum_{i=1}^n (x_i - \bar{x})^2 / n, \quad m_3 = \sum_{i=1}^n (x_i - \bar{x})^3 / n$$

と置く。標本歪度 $\sqrt{b_1}$ は、

$$\begin{aligned} \sqrt{b_1} &= m_3 / (m_2)^{3/2} = t_n \\ &= \sqrt{n} \left\{ \sum_{i=1}^n (x_i - \bar{x})^3 \right\} / \left\{ \sum_{i=1}^n (x_i - \bar{x})^2 \right\}^{3/2} \end{aligned}$$

と定義される。

この統計量 t_n の確率分布関数を $f_n(t_n)$ とする。

標本の大きさ n が十分に大きいときは、 $f_n(t_n)$ は漸近的に正規分布に従うので、 $n \geq 30$ ではたとえば Edgeworth 展開等を求めることができる。 n が小さいときは、

R.C. Geary (1947)¹⁾ が、標本の大きさが $n-1$ のときの分布 $f_{n-1}(t_{n-1})$ と、標本の大きさが n のときの分布 $f_n(t_n)$ との間の関係式を導き、これによって実際に $n=3$ のときの分布 $f_3(t_3)$ から出発して $n=4(1)8$ までの近似分布の展開式を与えている。

そこで、これらの展開式の精度と数値計算的な手法で高めることを目標として、今回は $n=4$ の場合について述べる。DE変換公式を適用すれば数値的には $f_4(t_4)$ の値を高精度で求めることはできるが手回がかかっているので、いくつかの代表値を選んで計算しておく。これを土台として Geary の展開式¹⁾ (ここでは $g_4(t_4)$ と名づける) を補正した更に精度のよい近似式 $\overline{f_4(t_4)}$ が最小2乗法で得られたので報告する。

2. Geary による $f_4(t_4)$ の積分表示とその展開式

標本の大きさ $n=3$ のときの標本歪度 t_3 の分布 $f_3(t_3)$ は

$$(1) \quad f_3(t_3) = \frac{\sqrt{2}}{\pi \sqrt{1-2t_3^2}} \quad \text{で } |t_3| \leq \frac{1}{\sqrt{2}}$$

で与えられる。

$n=4$ のときの 標本歪度 t_4 の 分布 $f_4(t_4)$ は

$$(2) \quad f_4(t_4) = \frac{\sqrt{3}}{2\pi} \int_D 1/\sqrt{y(x;v)} \, dx, \quad |t_4| \leq \frac{2}{\sqrt{3}}$$

ここで,

$$(3) \quad y(x;v) = -(3x^2-1)^2(3x^2-2) - 2v(5x^3-3x) - v^2,$$

$$v = \sqrt{3} \, t_4$$

であり, (2) の積分の領域 D は, $|x| \leq 1$ で

$y(x;v) \geq 0$ を満たす x の範囲とする.

$v = \sqrt{3} \, t_4$ が小さいとき, Geary は (2) 式の積分表示の主要項を求めて, 展開式 $f_4(t_4)$ と次のようにして導いている: すなわち, $v \neq 0$ のときは $y(x;v) = 0$ の根は $-1/\sqrt{3}$ に近い二つの実根と, $1/\sqrt{3}$ に近い二つの複素根と, $-\sqrt{2/3}$ と $\sqrt{2/3}$ に近い単根とを持つので, $O(v^{3/2})$ の項を無視して 四つの実根は,

$$(4) \quad \begin{cases} a = -\sqrt{2/3} - v/q, & b = -1/\sqrt{3} - \alpha\sqrt{v}, \\ c = -1/\sqrt{3} + \alpha\sqrt{v}, & d = \sqrt{2/3} - v/q, \end{cases}$$

ここで, $\alpha^2 = 2/(9\sqrt{3})$ として (4) で

近似できる. Geary は (2) 式を次のように分けて,

$$f_4(t_4) = \frac{\sqrt{3}}{2\pi} \left\{ \int_a^b + \int_c^{1/\sqrt{3} - \alpha\sqrt{v}} + \int_{1/\sqrt{3} - \alpha\sqrt{v}}^{1/\sqrt{3} + \alpha\sqrt{v}} + \int_{1/\sqrt{3} + \alpha\sqrt{v}}^d \right\}$$

$O(v^{3/2})$ の項を無視して, $f_4(t_4)$ の展開式 $g_4(t_4)$ を (5) のように導いている.

$$(5) \quad g_4(t_4) = \frac{\sqrt{3}}{2\pi} \left[-\frac{1}{\sqrt{3}} \log\left(\frac{2}{3\sqrt{3}}v\right) + \frac{1}{\sqrt{3}} \log(1+\sqrt{2}) \right. \\ \left. + \frac{2}{\sqrt{3}} \left\{ \log 2 + \frac{1}{2} \log(\sqrt{2}-1) \right\} \right]$$

これは, $0 < v \leq 2^{-17} \doteq 0.75 \cdot 10^{-5}$ では 10^{-12} 以下の誤差となる
3 非常によい展開式である.

3. Geary の展開式 $g_4(t_4)$ の誤差 $E_4(t_4)$

$v = \sqrt{3} t_4$ なので, (2) から $|v| \leq 2$ であるが, $v > 0$ であり, v の範囲が $[2^{-17}, 2)$ ^{注1)} では, 当然 Geary の展開式 $g_4(t_4)$ の誤差は大きくなるが, (2) 式を直接 DE 変換公式で数値積分を行えば $f_4(t_4)$ の数値は相対誤差で 10^{-15} 程度まで求めることができる.

$[2^{-17}, 2)$ の区間の v に対して, (3) 式の

$$y(x; v) = -27x^6 + 36x^4 - 10x^3v - 15x^2 + 6xv + 2 - v^2 = 0$$

を満たす数値解を DKA 法で求め, 四つの実根を小さい方から a, b, c, d と名づける. これ等を用いれば,

$y(x; v) \geq 0$ を満たす領域 D が定まるので

$$(6) \quad f_4(t_4) = \frac{\sqrt{3}}{2\pi} \left\{ \int_a^b \frac{dx}{\sqrt{y(x; v)}} + \int_c^d \frac{dx}{\sqrt{y(x; v)}} \right\}, \quad v = \sqrt{3} t_4$$

を計算すればよい. 被積分関数は端点で特異性を示すが, DE 公式では問題なく, このようにして求めた値を, この報告では, 数値的に $f_4(t_4)$ の真の値とみなす.

Geary の展開式 $g_4(t_4)$ の誤差を $E_4(t_4)$ と書くと,

$$(7) \quad E_4(t_4) = f_4(t_4) - g_4(t_4), \quad t_4 = v/\sqrt{3}$$

は, 数値的には簡単に求められる.

そこで, $E_4(t_4)$ と v との関係をしらべるため,

$\log_2 v$ の値を $-17, -16, -15, \dots, -1, 0, 1$ として

$\log_{10}(E(v/\sqrt{3}))$ の値をプロットしたのが図 1

である. これから Geary の展開式は, v が大きくなると
対数目盛でほぼ直線的にゆるくなることが分かる.

そこで, $\log_{10}(E_4(t_4))$ を $\log_2 v$ の k 次の多項式で
最小二乗近似 $\overline{\log_{10}(E_4(t_4))}$ と求めることにしよう.

これを次節に述べる.

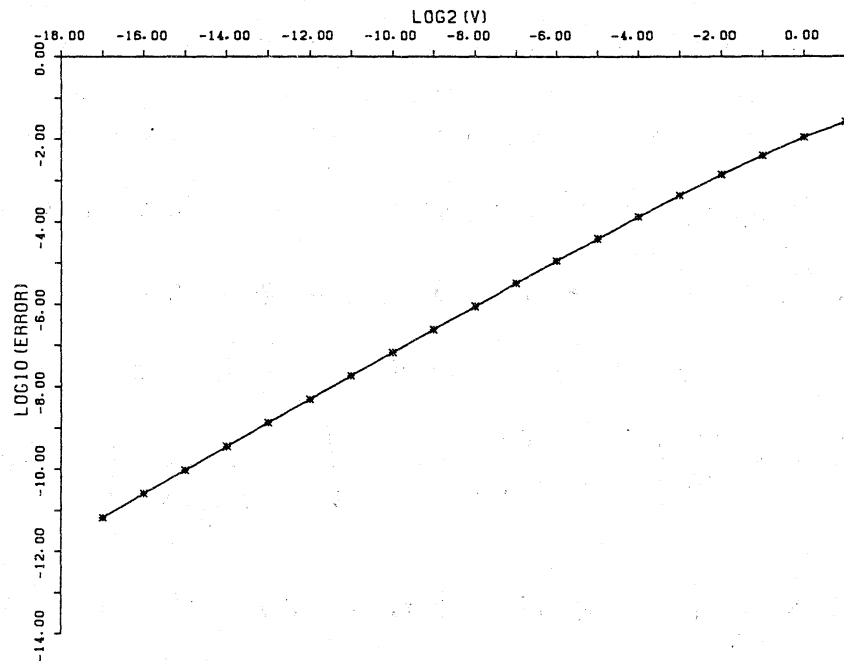


図 1 $g_4(v/\sqrt{3})$ の誤差

(注1) DE 変換公式でも v が 2^{-17} より小さく
なると精度がでにくくなる。また $v \rightarrow 2$ となる
 $t_4 \rightarrow \sqrt[3]{3}$ のときは $f_4(t_4) \rightarrow \frac{1}{2\sqrt{3}}$ と Geary により理論的に
求められている。

$v = 2$ のときは, $y(x; 0) = -(x+1)^2(3x-1)^2(3x^2-4x+2)$

となるので, $|x| \leq 1$ では常に $y(x; 0) \leq 0$ となる。

したがって, $f_4(t_4) = \frac{\sqrt{3}}{2\pi} \int_D \frac{dx}{\sqrt{y(x; 0)}} = 0$ となる。

(注2) $v = 0.01, 0.10, 1.00, 1.80, 2.00$ の場合
 $|x| \leq 1$ における $y(x; v)$ と $1/\sqrt{y(x; v)}$ の様子を(左, 右)に,
図 2.1, 図 2.2, 図 2.3, 図 2.4, 図 2.5 に与えておく。

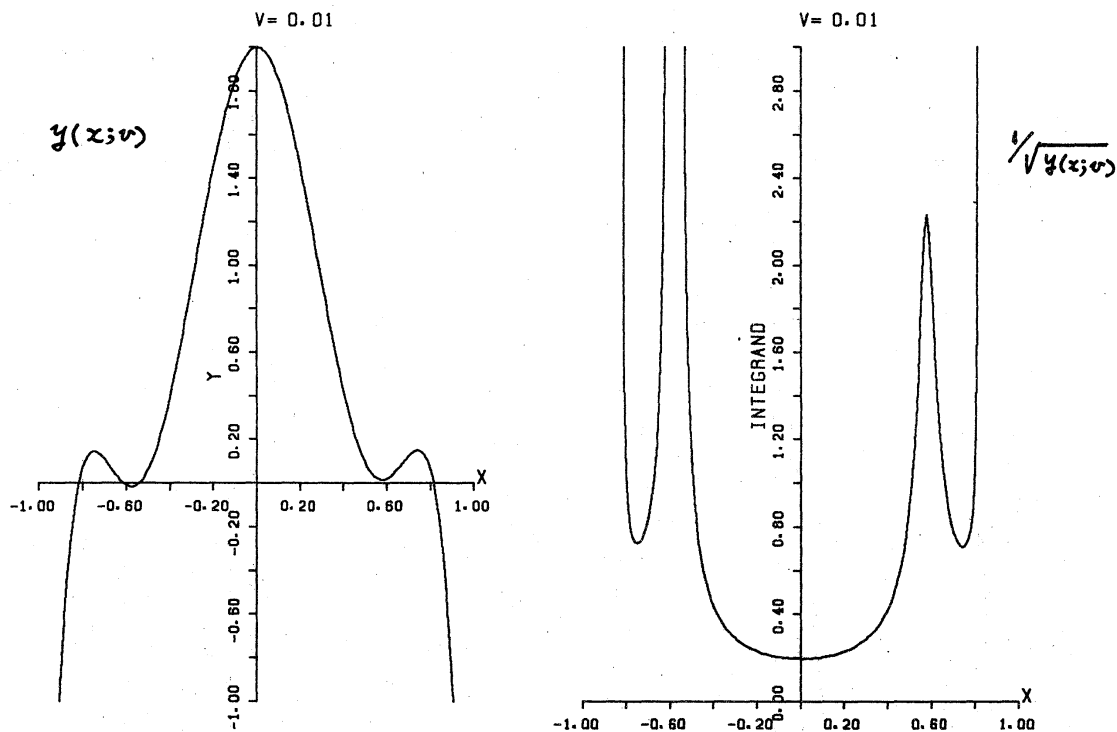
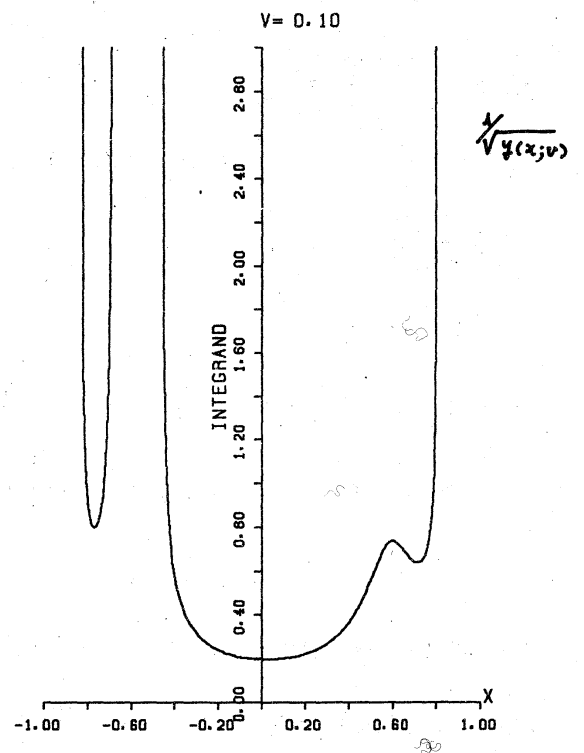
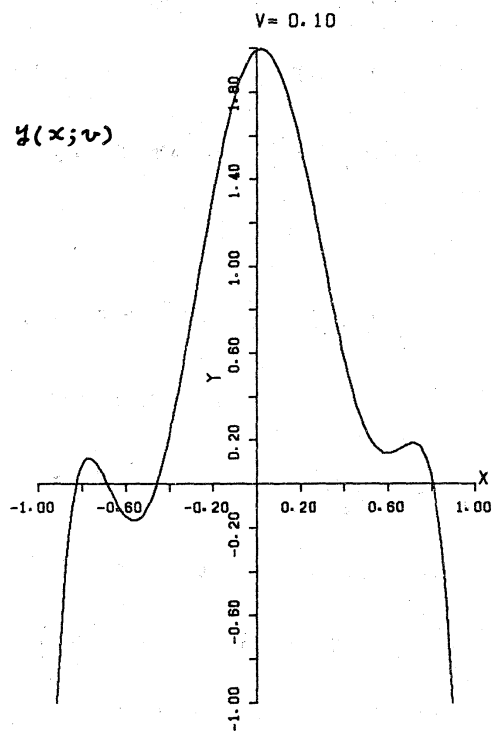
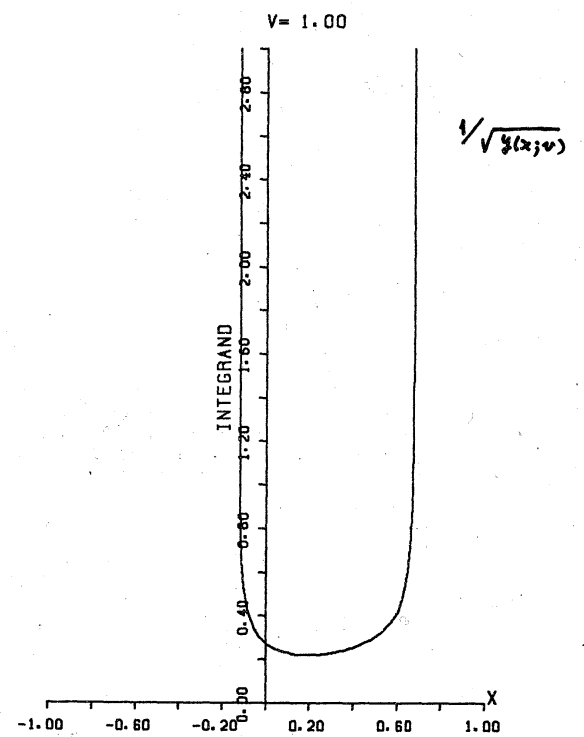
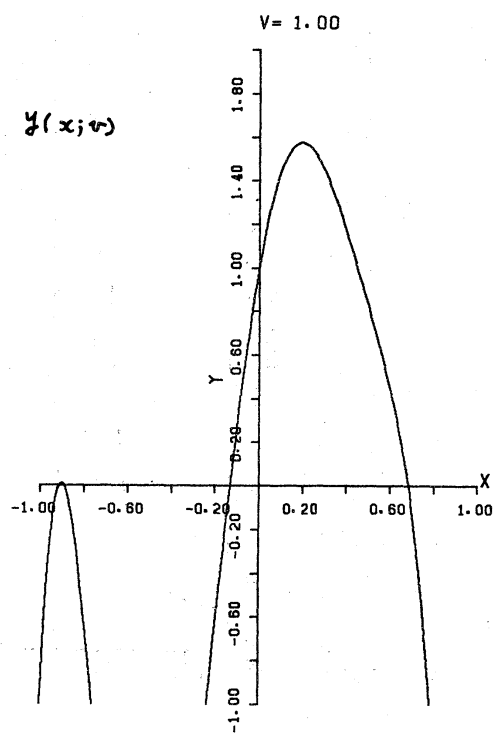


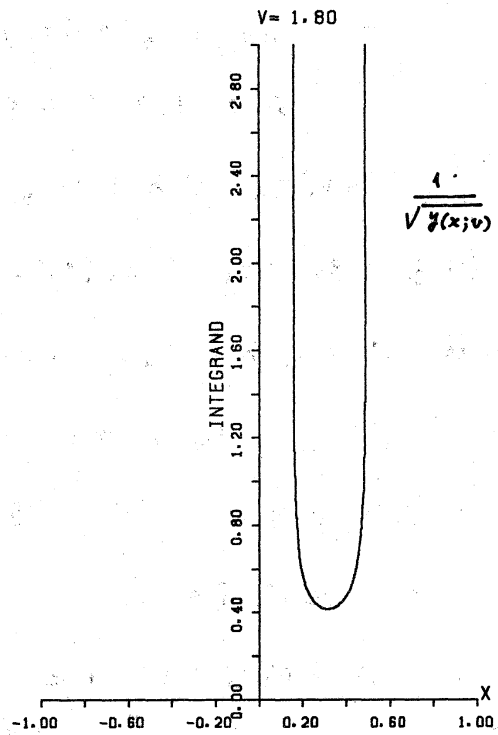
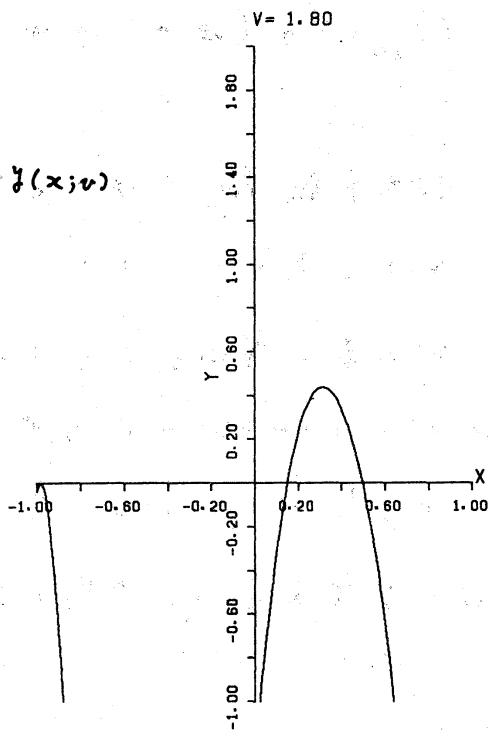
図 2.1



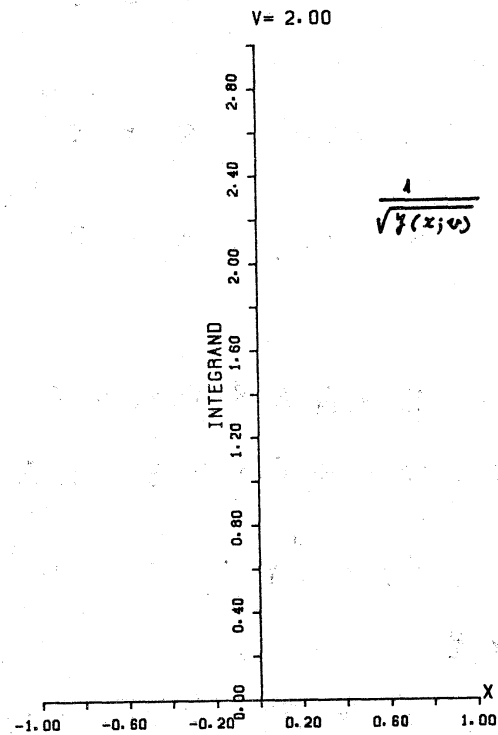
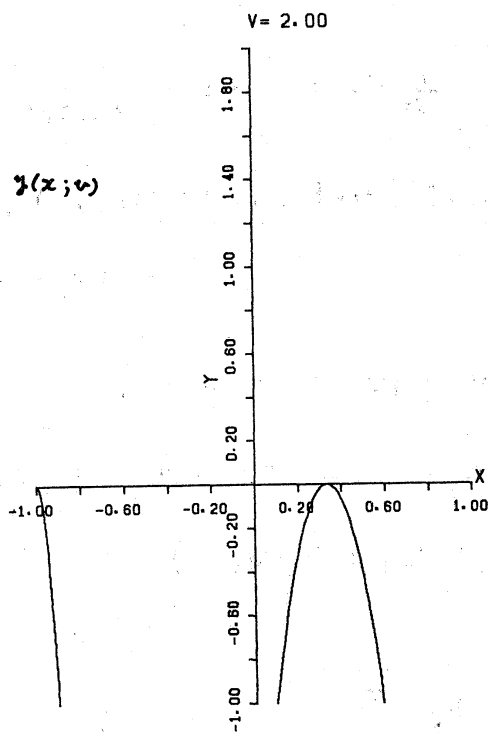
2-2



2-3



2-4



2-5

4. $\log_{10}(E_4(v/\sqrt{3}))$ を $(\log_2 v)$ の k 次の多項式であてはめること

$v = \sqrt{3}t_4$ であるが, 今 独立変数 $\log_2 v$ を X と書き, 従属変数 $\log_{10}(E_4(v/\sqrt{3}))$ を Y と書く. 4:2, 問題 を " $(X_1, X_2, \dots, X_N)$ と代表に選び, これに対して $(Y_1, Y_2, \dots, Y_N)$ が求められたとき, k 次までの選真直交多項式を用いて, Y を近似しよう" という ことにする.

v の 区間 $[2^{-17}, 2]$ における代表真 v_i の 選び方として 次の場合を試みた.

- (8) $N = 19$, $v_i = 2^{(i-18)}$, $i = 1, 2, \dots, 19$
- (9) $N = 16$, $v_i = 2^{(18 \cos(\frac{\pi}{32}i) - 17)}$, $i = 0, 1, \dots, 15$
- (10) $N = 16$, $v_i = 2^{(9 \cos(\frac{\pi}{32}(2i+1)) - 8)}$, $i = 0, 1, \dots, 15$
- (11) $N = 14$, $v_i = 2^{(1 - \tanh(\sinh(\sinh(8i))))}$,
 $i = 0, 1, \dots, 13$

選真直交多項式による不等間隔データのあてはめ

は, 次の手順で行う. 代表真 v_i , $i = 1, 2, \dots, N$ に対して

$$(12) \quad x_i = \log_2 v_i, \quad Y_i = \log_{10}(E_4(v_i/\sqrt{3}))$$

と書く. また,

$$(13) \quad \bar{X} = \sum x_i / N, \quad T_i = x_i - \bar{X}$$

とする. さらに,

$$(14) \quad \phi_0(T_i) = 1, \quad \phi_i(T_i) = T_i, \quad i=1, 2, \dots, N$$

とする。ここで ϕ_0, ϕ_1 はそれぞれ 0 次と 1 次の多項式である。そこで、

$$j = 2, 3, \dots, k; \quad k < N$$

に対して 次の (15), (16), (17) を計算する。

$$(15) \quad \phi_j(T_i) = (T_i - \alpha_{j-1})\phi_{j-1}(T_i) - \beta_{j-1}\phi_{j-2}(T_i)$$

ここで、

$$(16) \quad \alpha_{j-1} = \sum_i T_i \phi_{j-1}^2(T_i) / \sum_i \phi_{j-1}^2(T_i),$$

$$(17) \quad \beta_{j-1} = \sum_i \phi_{j-1}^2(T_i) / \sum_i \phi_{j-2}^2(T_i)$$

である。

このようにして得られた T について j 次の多項式 $\phi_j(T)$,

$j = 0, 1, 2, \dots, k$ は k 以下の整数 r, s に対して、

$$\sum_i \phi_r(T_i) \cdot \phi_s(T_i) = 0, \quad r \neq s$$

$$\sum_i \phi_r^2(T_i) \neq 0$$

を満足することはよく知られている。

この直交多項式 ϕ_j ($j=0, 1, 2, \dots, k$) を用いて、

代表点 v_i ($i=1, 2, \dots, N$) で求められた $\log_{10}(E_4(v_i/\sqrt{3}))$

の値から k 次式を当てはめて

$$(18) \quad \log_{10}(E_4(v/\sqrt{3})) = \lambda_0 + \lambda_1 \phi_1(T) + \lambda_2 \phi_2(T) + \dots + \lambda_k \phi_k(T)$$

が得られる。

そこで,

$$(9) \quad \lambda_j = \sum_i Y_i \phi_j(\tau_i) / \sum_i \phi_j^2(\tau_i), \quad j = 0, 1, \dots, k$$

である.

このようにして, N の代表値により $\log_{10}(E_4(v/\sqrt{3}))$ を最小二乗近似した (8) 式が求められたことになる.

ここで問題は N の代表値の選び方で, さき程の (8), (9), (10), (11) のどれがよいのか? は次の節の数値例で示す.

5. $\overline{\log_{10}(E_4(v/\sqrt{3}))}$ で $g_4(v/\sqrt{3})$ を補正して得られた

$f_4(v/\sqrt{3})$ の近似式 $\overline{f_4(v/\sqrt{3})}$

$2^{-17} \leq v \leq 2$ の範囲内における v に対して (7) より

$$E_4(v/\sqrt{3}) = f_4(v/\sqrt{3}) - g_4(v/\sqrt{3})$$

であった. そこで N 個の適当な代表値 v_i における $f_4(v_i/\sqrt{3})$ の精密な値を用意しておけば, $\log_{10}(E_4(v/\sqrt{3}))$ のよい近似として $\overline{\log_{10}(E_4(v/\sqrt{3}))}$ が前節の (18) で求められた.

そこで, $\overline{E_4(v/\sqrt{3})}$ が $E_4(v/\sqrt{3})$ の近似として

$$(20) \quad \overline{E_4(v/\sqrt{3})} = \overline{\log_{10}(E_4(v/\sqrt{3}))}$$

で求められるので, $f_4(v/\sqrt{3})$ の近似 $\overline{f_4(v/\sqrt{3})}$ は,

$$(21) \quad \overline{f_4(v/\sqrt{3})} = g_4(v/\sqrt{3}) + \overline{E_4(v/\sqrt{3})}$$

で表わすことができる. また,

$\overline{f_4(v/\sqrt{3})}$ の誤差 $\overline{f_4(v/\sqrt{3})} - f_4(v/\sqrt{3})$ は, Geary の展開の誤差 $E_4(v/\sqrt{3})$ と, $\delta = \overline{\log_{10}(E_4(v/\sqrt{3}))} - \log_{10}(E_4(v/\sqrt{3}))$

が命かれば、次の式から見つめることができる。

$$(22) \quad \overline{f_4(v/\sqrt{3})} - f_4(v/\sqrt{3}) \doteq E_4(v/\sqrt{3}) \times 2.3\delta \quad \text{注3)}$$

たとえば、(22)で誤差を 0.1_{10}^{-10} 程度にしたいならば、

$$E_4(v/\sqrt{3}) = 0.2_{10}^{-1} \quad (v=1.7) \quad \text{では} \quad \delta = 0.2_{10}^{-9},$$

$$E_4(v/\sqrt{3}) = 0.1_{10}^{-1} \quad (v=1) \quad \text{では} \quad \delta = 0.4_{10}^{-9},$$

$$E_4(v/\sqrt{3}) = 0.1_{10}^{-2} \quad (v=0.25) \quad \text{では} \quad \delta = 0.4_{10}^{-8},$$

$$E_4(v/\sqrt{3}) = 0.1_{10}^{-5} \quad (v=0.4_{10}^{-2}) \quad \text{では} \quad \delta = 0.4_{10}^{-5}$$

程度になつていけばよい。つまり v が小さい所では、Garry
の展開式 $g_4(v/\sqrt{3})$ の誤差 $E_4(v/\sqrt{3})$ が既に小さいので、補正に
用いる $\log_{10}(E_4(v/\sqrt{3}))$ の近似値の誤差 δ は、それ程小さく
しなくてもよいことになる。

(注3) (22) 式の証明

$$\begin{aligned} \overline{\log_{10}(E_4(v/\sqrt{3}))} - \log_{10}(E_4(v/\sqrt{3})) &= \delta \quad \text{なので,} \\ {}_{10}\overline{\log_{10}(E_4(v/\sqrt{3}))} &= {}_{10}(\log_{10}(E_4(v/\sqrt{3})) + \delta) \\ &= E_4(v/\sqrt{3}) \times 10^\delta \end{aligned}$$

ここで、

$$10^\delta = (e^{\log_{10}})^\delta = (e^\delta)^{\log_{10}}$$

$$\delta \ll 1 \text{ のとき, } \doteq (1+\delta)^{\log_{10}} \doteq 1 + \delta \times \log_{10} \doteq 1 + 2.3\delta$$

$$\begin{aligned} (21) \text{ か } \delta, \quad \overline{f_4(v/\sqrt{3})} &= g_4(v/\sqrt{3}) + \overline{E_4(v/\sqrt{3})} \\ &= g_4(v/\sqrt{3}) + {}_{10}\overline{\log_{10}(E_4(v/\sqrt{3}))} \\ &\doteq g_4(v/\sqrt{3}) + E_4(v/\sqrt{3}) \times \{1 + 2.3\delta\} \end{aligned}$$

$$= \left\{ f_4(v/\sqrt{3}) - E_4(v/\sqrt{3}) \right\} + E_4(v/\sqrt{3}) \cdot \{ 1 + 2.3\delta \}$$

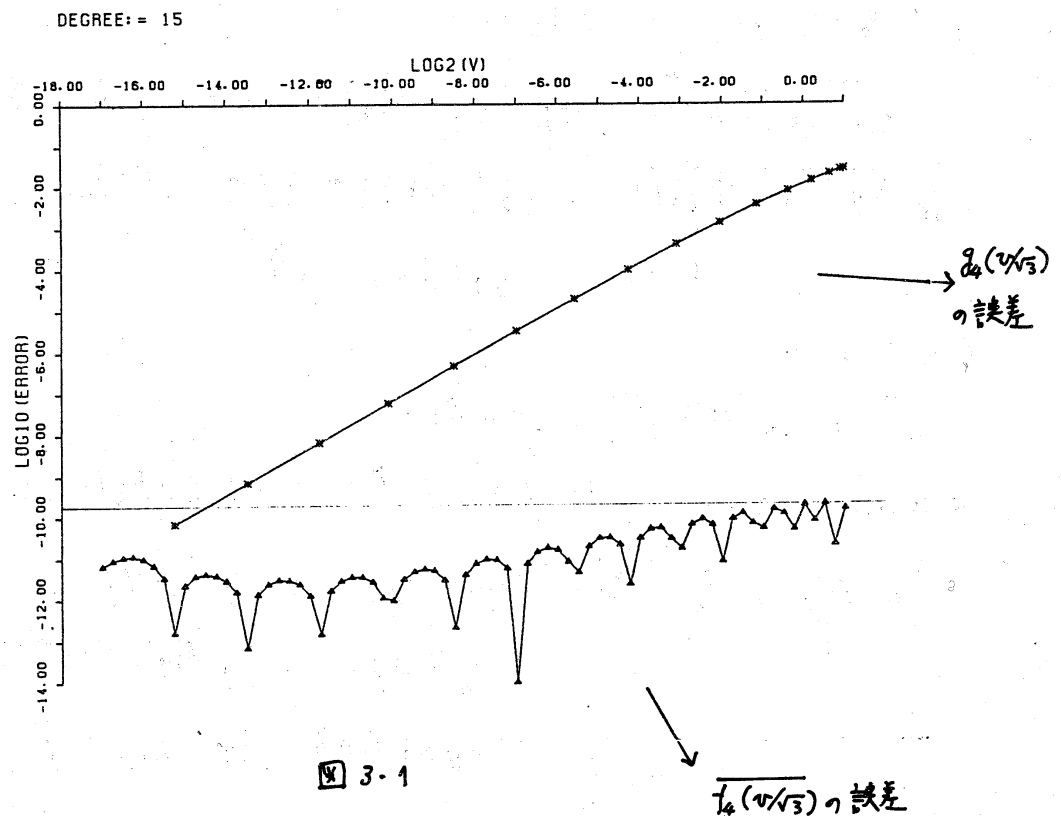
したがって、

$$\overline{f_4(v/\sqrt{3})} - f_4(v/\sqrt{3}) \approx E_4(v/\sqrt{3}) \cdot 2.3\delta$$

6. 数値例

代表真 v_i の選び方とその数 N と近似式の次数 k とを変えて $f_4(v/\sqrt{3})$ を作成し、その誤差の様子を試みた。

代表真 v_i の選び方	真の数	次数	$f_4(v/\sqrt{3})$ の最大誤差	図番号
(9)式	$N=16$	$k=15$	$10^{-9.8}$	3-1
(10)式	$N=16$	$k=15$	10^{-8}	3-2
(11)式	$N=14$	$k=11$	$10^{-9.3}$	3-3



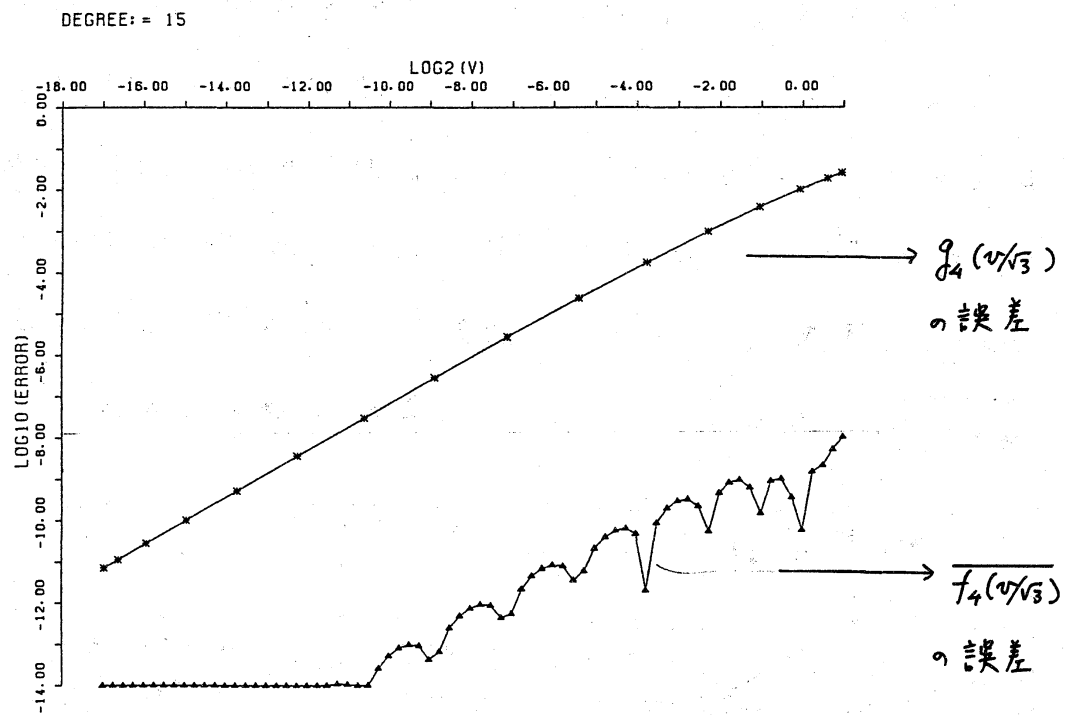


図 3-2

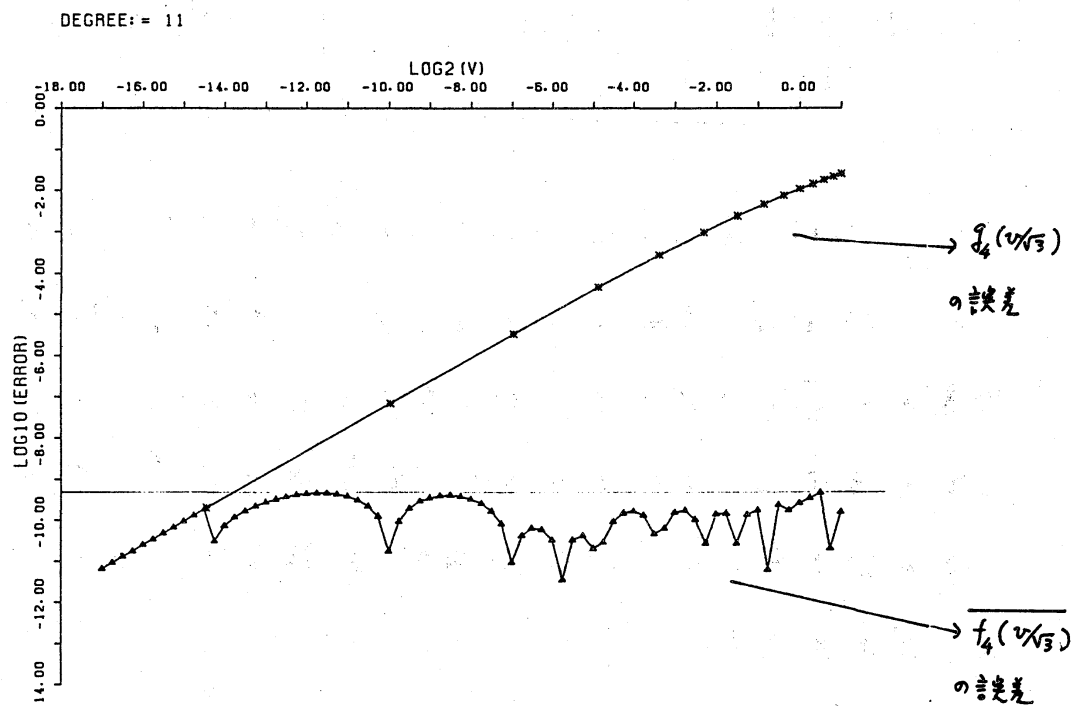


図 3-3

これらの図3.1, 3.2, 3.3 により近似式 $\overline{f_4}(v/\sqrt{3})$ の誤差のふるまいを見ると, 代表値 v_i の選び方には (9) 式を用いるのがよいと思われる。そこで, 1) $N=24$, $k=20$ の場合と 2) $N=32$, $k=22$ の場合, v の値を次のように変えて, Geary の展開式 $g_4(v/\sqrt{3})$ の誤差と, (21) 式で求めた近似式 $\overline{f_4}(v/\sqrt{3})$ の誤差とを, 上下に並べて図示する。

独立変数 (9) 式 の v_i	$\log_2 v$	v
	-17 から 1 まで	0 から 2 まで
1) $N=24$ 点 $k=20$ 次	図 4.1	図 4.2
2) $N=32$ 点 $k=22$ 次	図 5.1	図 5.2

なお, 次数 k は $(N-1)$ 次まで上げることができるとは,

1) の場合は $k=20$ のとき, 各データ値 v_i における残差が一番小さくなるので ここで止めた。

2) の場合も $k=22$ のとき残差が一番小さくなる。

この結果からは 2) の近似式が誤差のふるまいから見て優れている。

SAMPLE SIZE: = 24

DEGREE: = 20

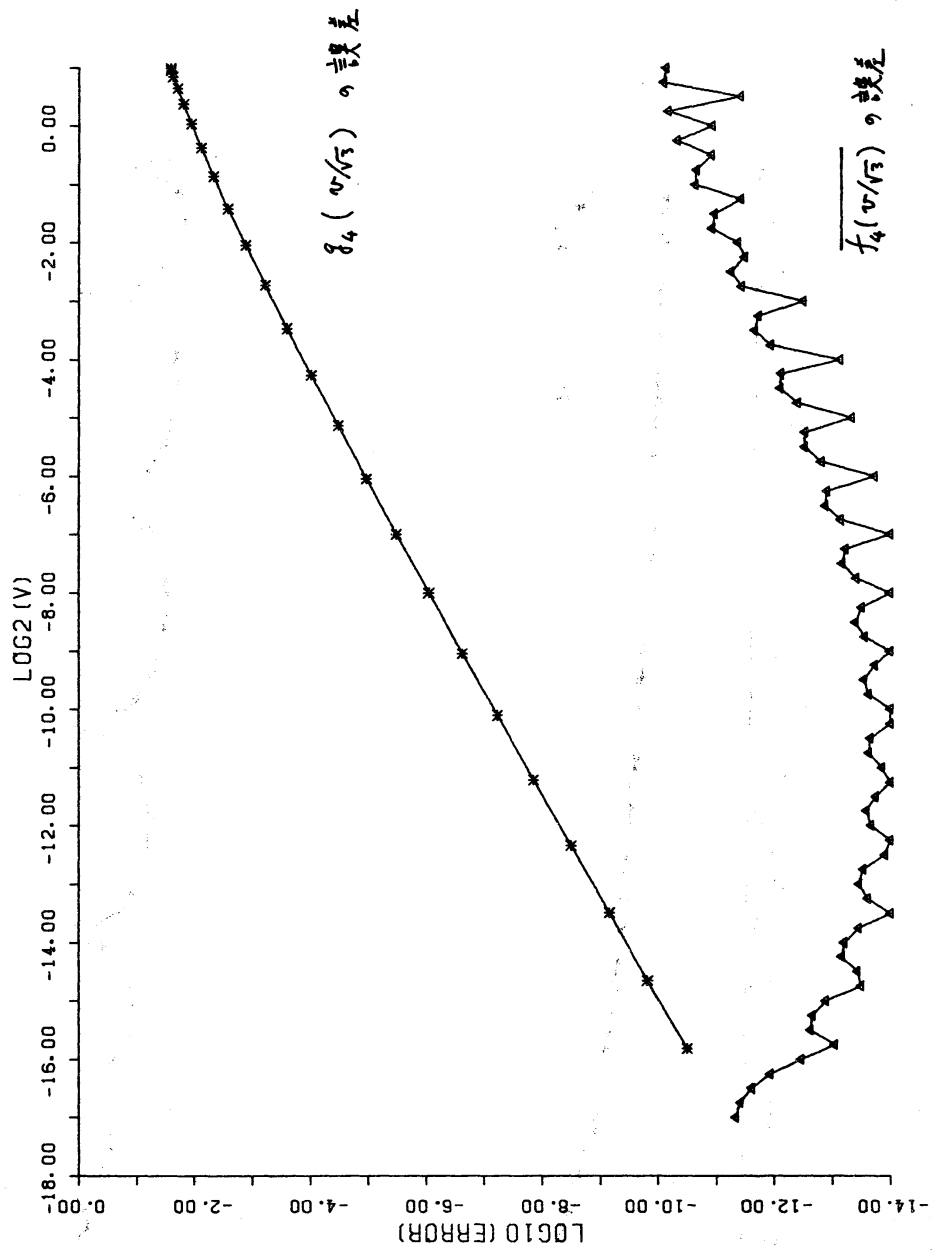
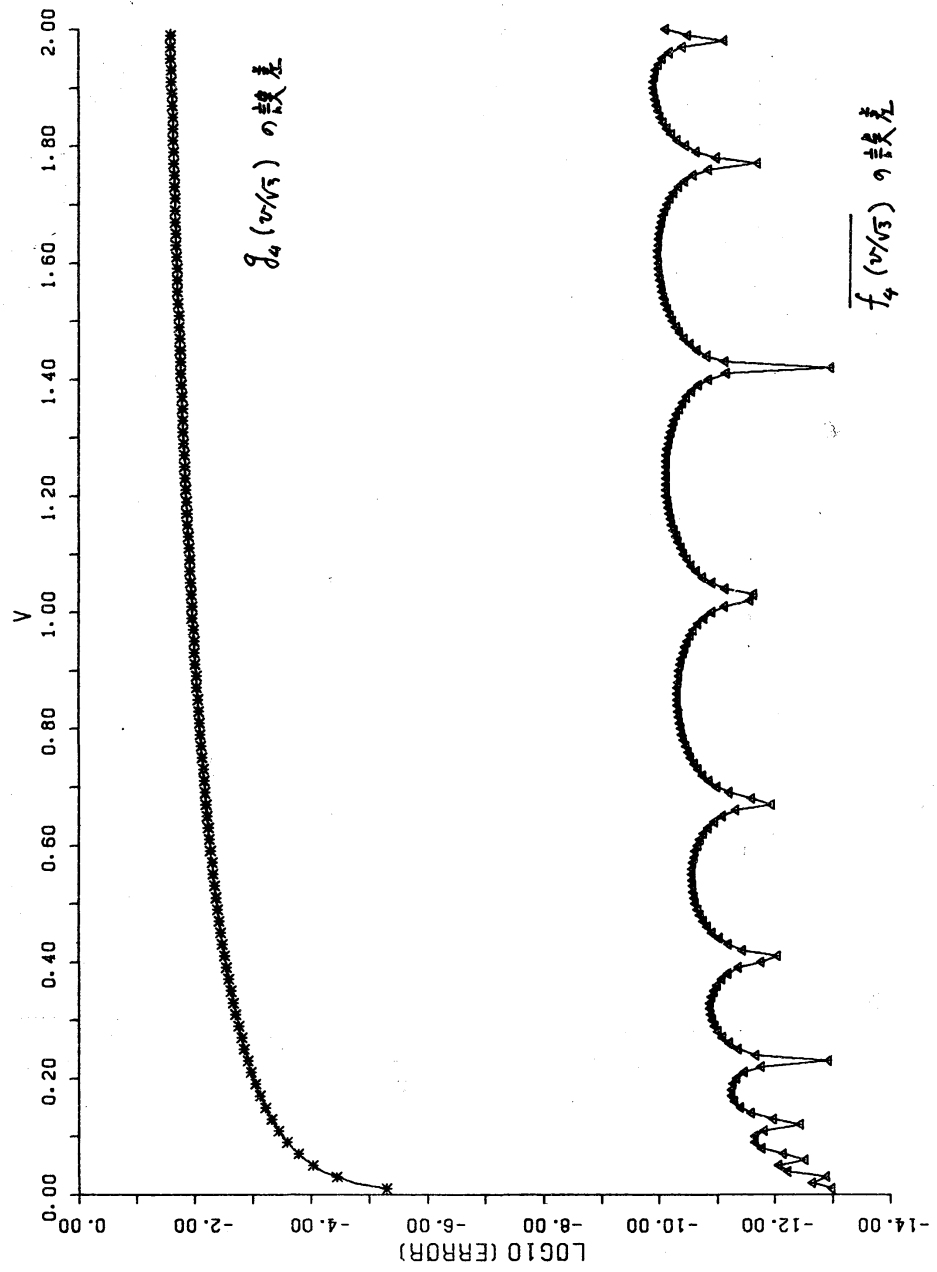


図 4.1

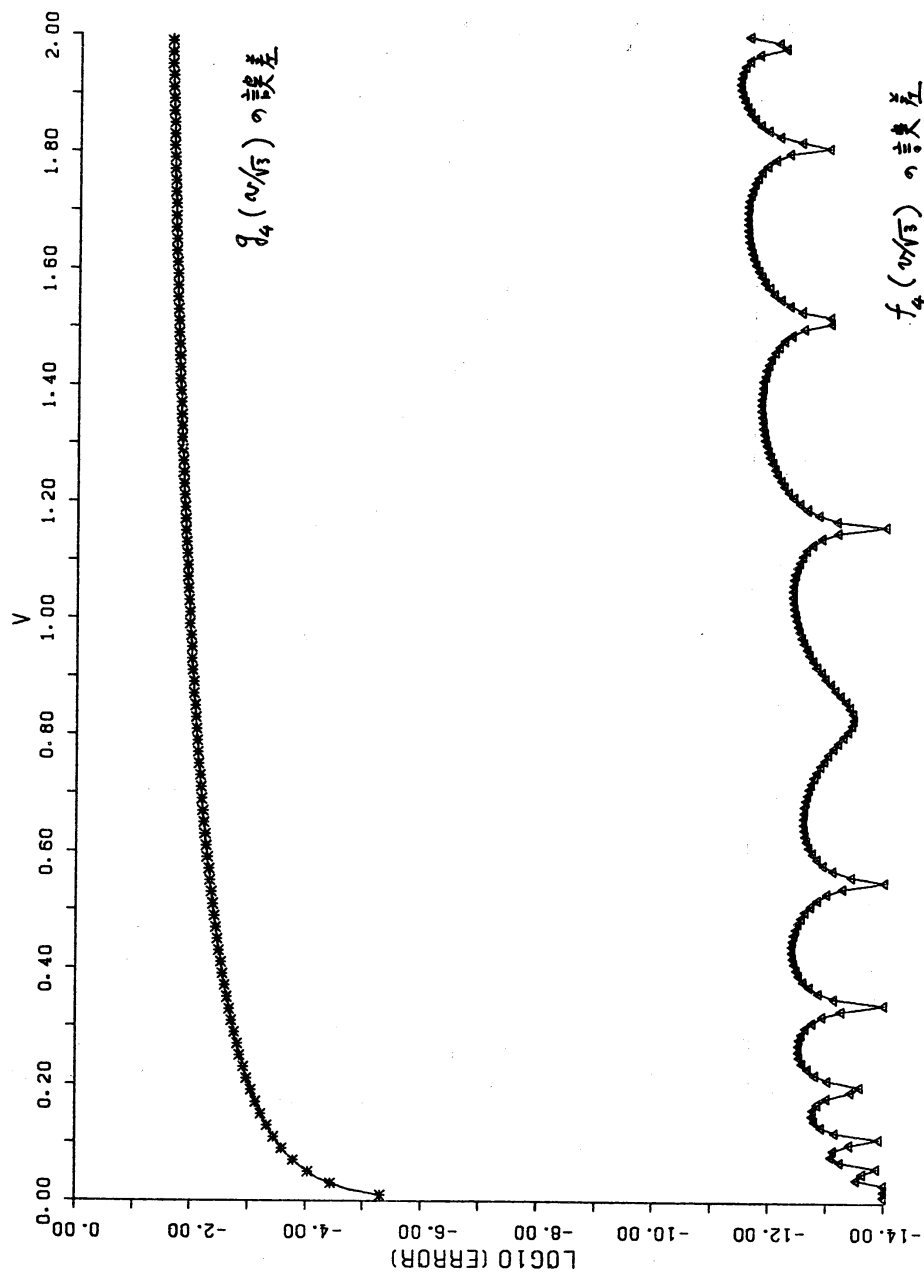
SAMPLE SIZE: = 24

DEGREE: = 20



SAMPLE SIZE: = 32

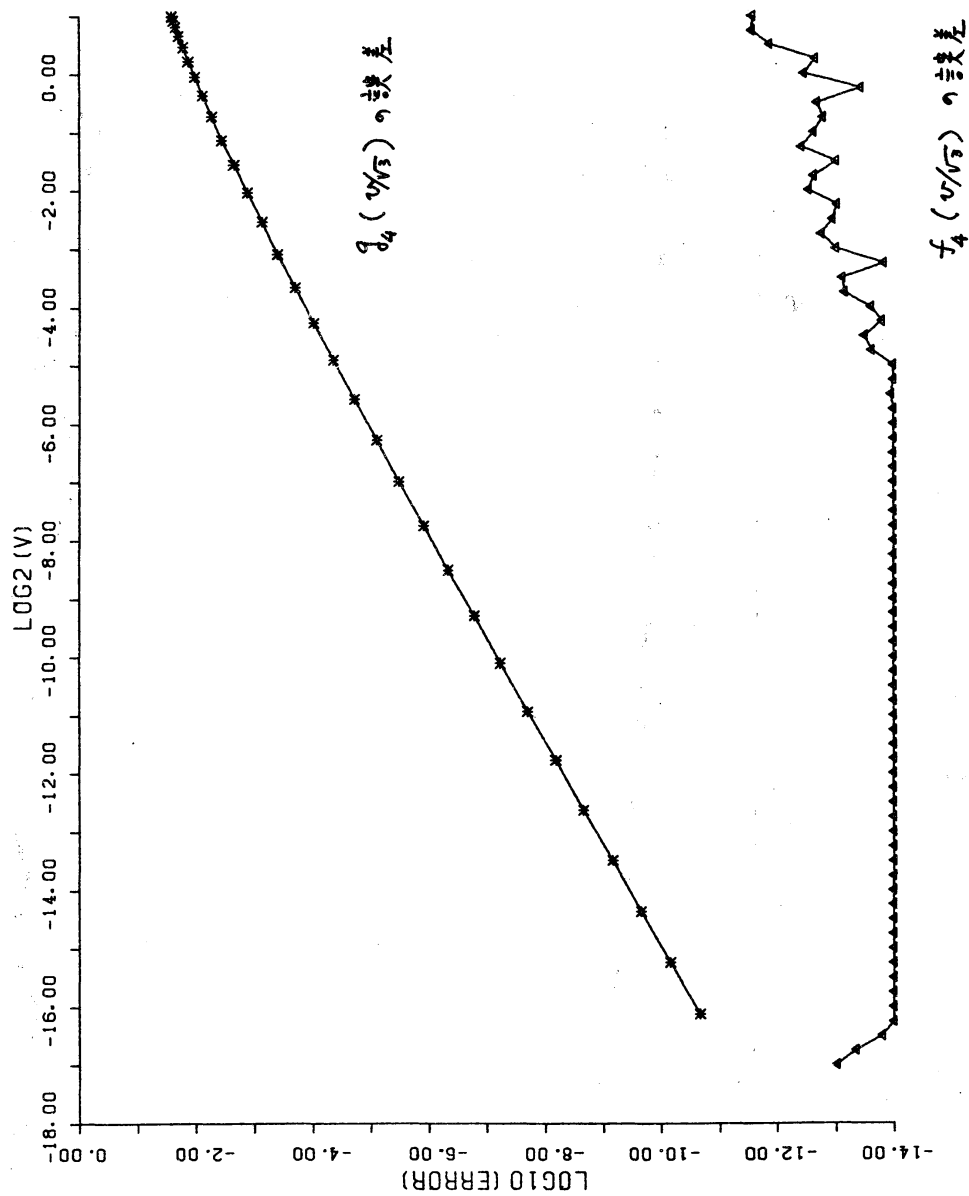
DEGREE: = 22



17 5-1

SAMPLE SIZE: = 32

DEGREE: = 22



おわりに

ある関数の積分表示の式が与えられて、それと数値積分法により 関数値は精度よく求めることが出来るが、かなり手回がかかる場合、いくつかの代表真正関数の性質に依じて "適当に" 選んで関数値を求めたおえ、選定直交多項式を用いた最小二乗法で、その関数のよい近似式を導く 一つの数値例を示したことになる。

代表真正の選び方による 近似式の最大誤差の見積りは、たとえば (18) 式において $\max_i (T_i)$ における直交多項式 $|\phi_{k+1}| \times \lambda_{k+1}$ で できると思われる。

また、代表真正の選び方は (9) 式が 今の所では 一番よかったが、これが最高であるかどうかは分らない。これらは 今後の課題とする。

謝辞

情報処理研修センター (IIT) には、コンピュータ M340S の使用に大へんお世話になった。ここに感謝する次第である。

参考文献

- 1) R.C. Geary The frequency distribution of $\sqrt{b_1}$ for samples of all sizes drawn at random from a normal population, Biometrika, 34 (1947) 68-97.