

強単峰性を中心とした更なる単峰分布についての議論
(More on Unimodal Distributions, particularly on Strong Unimodality)

近畿大学経営学部 林 芳男(Yoshio Hayashi)

Faculty of Business Administration,

Kinki University

(本文) 1. 始めに 本研究の目的は確率分布の単峰性、強単峰性の過去の理論を見直して整理することである。去年の発表(林(2012))では任意の単峰分布は区間 $(0, 1)$ 上の一様な分布と適当な確率分布を持つ独立な確率変数 U と X の積 $Z \equiv UX$ の確率分布として表現できるというヒンチンの表現定理を初等的に証明したと述べた。が、一般の離散分布に関しては成功していなかった。その修正は § 2 で述べる。本稿では強単峰性の理論を中心に Ibragimov の定理の初等的な証明を展開し、それを離散分布の場合へと拡張する。

単峰性の議論の中で導入された重要な概念の一つに強単峰性がある。「強単峰分布とはすべての単峰分布との畳み込みが単峰になる分布のことである」は単峰性の古典的な定義の下でも Keilson の離散的な場合でも同じで、興味深いことはその対象の分布に対しては事前に単峰性は要求していなかったけれど畳み込みを取った結果が単峰になることを要求することでそのような分布は単峰分布になることであった。しかし、始めから単峰分布の部分集合で畳み込みに関して閉じているものと定義してくれた方が分かりが良かったと思う。Ibragimov (1956) は連続な分布関数のクラスの中でその確率密度関数 $f(x)$ が強単峰である為の必要十分条件は $\log f(x)$ が $(-\infty, \infty)$ の上で凹であることという主張をした。その証明のために彼はまず正規分布が強単峰であることそして単峰分布列が弱収束で閉じていることを示した。しかし、私達は彼のその証明がよく分からない。後に書かれた他者による文献の中でもその証明の成否については言及しないままここでは別の証明を与えるという論の進め方(Dharmadhikari and Joag-Dev (1988))をしていて、私達にとってはそういう彼等の証明もまた分かりづらいものでした。私達は § 4 でその台が有限である場合も含めた場合のその主張の初等的証明を与える。しかし、その確率密度関数はその定義域で微分可能であることは仮定する。また § 5 で一般の離散分布に対しても対数凹性を定義し、それがまた離散分布のクラスの中で強単峰である為の必要十分条件がその類似事項であることを証明する。二つの離散分布のその独立な畳み込みが $\pm\infty$ 両方向に広がる時連続分布になることが有ることを指摘しておく。私達は本稿の最初の成果として先ず § 3 で単峰分布のクラスは畳み込みに関してはほぼ閉じていることを示す。

その前に基礎的事項と幾つかの疑問点を整理しておく。確率分布関数とは実直線 $(-\infty, \infty)$ 全体で定義された単調非減少な右連続な関数 $F(\cdot)$ で $F(-\infty) = 0$, $F(\infty) = 1$ なるものである。 $\{x : 0 < F(x) < 1\}$ なる点集合は $F(\cdot)$ の台と呼ばれ $\text{supp}(F)$ で表される。二つの分布関数 $F_1(\cdot)$, $F_2(\cdot)$ が与えられたときそれらをそれぞれ確率 α (但し、 $0 < \alpha < 1$), $1 - \alpha$ で混合したもの(mixture) $F(\cdot) \equiv \alpha F_1(\cdot) + (1 - \alpha) F_2(\cdot)$ の台は $\text{supp}(F) = \text{supp}(F_1) \cup \text{supp}(F_2)$ になる。ところで、絶対連続分布の場合、 0 をモードとする単峰分布関数 $F(x)$ は一様分布(族)の一般化混合(a generalized mixture)で表すことができるという主張がある(Dharmadhikari and Joag-Dev (1988; 4 頁、定理 1.2) 参照)。即ち、 W_a は $a > 0$ であるか $a < 0$ であるかに応じてそれぞれ区間 $(0, a)$ 又は $(a, 0)$ 上の一様分布の分布関数、更に、 W_0 は原点だけに確率質量 1 を持つ縮退した確率分布の分布関数、つまり、ヘビサイド関数であるとする。このとき $(-\infty, \infty)$ 上の確率分布関数 $\bar{H}(x)$ が存在して

$$F(x) = \int_{-\infty}^{\infty} W_z(x) d\bar{H}(z)$$

と表すことができるというのであるが、その $\overline{H}(\cdot)$ が具体的に何であるかは与えられていない。この主張の後彼等はヒンチンの表現定理(1988; 6 頁、定理1.3参照)を証明している。私が去年の発表会で示したのはこの部分の初等的証明である(林(2012))。Keilsonの離散単峰分布に関しては対応する特徴付けは行われなかったし、私が最も気にしていることは、そもそも非可算個の分布関数の混合は余り研究されておらず(Feller(1971;53頁))そういうことが数学的に無条件に許されるのであろうか?ということであり、またその $\overline{H}(\cdot)$ が具体的に与えられずにその表現が成立していることなのである。この理論は私の方法では使っていない。

確率分布関数 $F(\cdot)$ はすべて任意の実数 x に対して

$$(1.1) \quad F(x) = \alpha F_d(x) + (1 - \alpha) F_c(x)$$

という形に一意に分解することができる(Chung(1968)参照)。ここに、 α は $0 \leq \alpha \leq 1$ なる実数で $F_d(\cdot)$, $F_c(\cdot)$ も伴に確率分布関数で $F_d(\cdot)$ は階段関数で $F_c(\cdot)$ は連続関数である。これが単峰分布のものであるときのその性質を論じるのがこの論文の目的である。 $\alpha > 0$ の場合、この $F_d(\cdot)$ が $F(\cdot)$ の離散部分で、その跳躍点の集合は可算集合でそれを $\{x_n : n = 0, \pm 1, \pm 2, \dots\}$ 、 x_n での確率質量を $p_n \geq 0$ とする。ここに、 $-\infty < \dots < x_n < x_{n+1} < \dots < \infty$ で一般性を失うことなく $x_0 \equiv 0$ であるとする(もし $x_0 \equiv 0$ が本来の跳躍点でなかったならば $p_0 = 0$ とおくことにする。また、その台が(片方向もしくは両方向に)有限である場合、(その有限である方向の)無限に遠くの p_n はすべて0であると見なすことにする)。言うまでもなく

$$(1.2) \quad \sum_{n=-\infty}^{\infty} p_n = 1; \quad p_n \geq 0 \quad (0, \pm 1, \pm 2, \dots)$$

であって、これが(離散)単峰分布であるとき $x_0 = 0$ をモードとすれば

$$(1.3) \quad \dots \leq p_{-2} \leq p_{-1} \leq p_0 \geq p_1 \geq p_2 \geq \dots$$

である。これが私達の単峰な離散分布の定義である。

連続な部分 $F_c(x)$ は絶対連続である場合だけを考える。原点にモードが有る場合を考える為にその台は単連結集合で A, B を正数として $(-A, B)$ であるとする(つまり、原点は台に含まれる。無限区間の台を持つ場合はその無限に広がる方向に応じて $A = \infty$ またはかつ $B = \infty$ であるとする。その端点 $-A, B$ が有限な場合そのどちらか又は両方が定義域に含まれても構わないとする)。単峰な絶対連続分布の確率密度関数 $f(\cdot)$ の台が $(-A, B)$ であるとき

$$(1.4) \quad \int_{-A}^B f(x) dx = 1, \quad f(x) \geq 0 \quad (x \in (-A, B));$$

$f(x)$ は $x \in (-A, 0)$ で単調非減少、 $x \in (0, B)$ で単調非増加

となる。これが私達の絶対連続な単峰分布の定義である。 $f(\cdot)$ の第1階の導関数 $f'(\cdot)$ が存在する場合、その条件は $x \in (-A, 0)$ で $f'(x) \geq 0$ 、 $x \in (0, B)$ で $f'(x) \leq 0$ で置き換えることができる。ここに、“ $\geq$ ($\leq$) 0”という記号はその区間中にその左辺の関数が0でない部分が含まれるという意味である。

その $F_d(\cdot)$ の確率質量関数 $p(\cdot)$ と $F_c(\cdot)$ の(微分可能な)確率密度関数 $f(\cdot)$ との混合 $\overline{F}(\cdot)$ は数値的には $\overline{F}(\cdot) = \alpha p(\cdot) + (1 - \alpha) f(\cdot)$ でその確率質量関数 $p(x)$ は $n = 0, \pm 1, \pm 2, \dots$ の各 $x = x_n$ に対して p_n の値を取り、その他の任意の点 x では0という関数である。 $\overline{F}(\cdot)$ の形状は、区間 $(-A, B)$ 内にその離散分布の跳躍点 $x = x_n$ が入る場合、そこでの値は $\alpha p_n + (1 - \alpha) f(x_n)$ となりその周りの値 $(1 - \alpha) f(x)$ から跳躍するので、それが(その近傍で)単峰になるのは $x_{-1} = -A$ で $x_1 = B$ で $\alpha p_{-1} \leq (1 - \alpha) f(x_{-1} + 0)$ で $(1 - \alpha) f(x_1 - 0) \geq \alpha p_1$ となる場合だけである(その連続部分の端点はその定義域内に入る場合はその端点での $\overline{F}(\cdot)$ の値は $\overline{F}(-A) = \alpha p_{-1} + (1 - \alpha) f(-A)$, $\overline{F}(B) = \alpha p_1 + (1 - \alpha) f(B)$ で $\overline{F}(-A + 0) = (1 - \alpha) f(-A)$, $\overline{F}(B - 0) = (1 - \alpha) f(B)$ で単峰には成り得ない)。その絶対連続な確

率密度関数が微分可能な場合の確率分布関数を微分すると確率密度関数が導出され、確率密度関数を積分すると確率分布関数が現れるという理論をその拡大系に拡張するならばそれは

$$(1.5) \quad \overline{f}(x) = \sum_{n=-\infty}^{\infty} \alpha p_n \delta(x - x_n) + (1 - \alpha) f(x)$$

と記述されるべきのものである。ここに、 $\delta(\cdot)$ はディラック(Dirac)のデルタ関数である。超関数としては $\delta(0)$ の値は不定なのであるが、数値関数としての整合性を持たせるためには $\delta(0) = 1$ であるとしておくのが都合が良い。つまり、 $\delta(\cdot)$ は

$$(1.6) \quad \delta(0) = 1, \delta(x) = 0 \ (x \neq 0), \text{ 原点の近傍で定義された任意の関数 } g(x) \text{ に対して} \\ \int_{-\infty}^{\infty} g(x) \delta(x) dx = g(0)$$

なる擬似関数である。こういった擬似関数を使った拡張された密度関数を定義するならば、それらを数値関数と見て、その形状がその定義域上で単峰であることで以て単峰性を定義することが自然であると思う。そういう意味でその最も一般的な形の単峰関数は、拡張された密度関数 $\overline{f}(\cdot)$ で考える場合、それは $f(\cdot)$ が(1.4)の条件に加えて区間 (x_n, x_{n+1}) で微分可能で $f(x_n - 0) = p_n, f(x_{n+1} + 0) = p_{n+1} \ (n = 0, \pm 1, \pm 2, \dots)$ を満足する場合である。

この一般の確率分布関数 $F(\cdot)$ が単峰であるための必要十分条件は、一般性を失うことなくそのモードが原点に有るとして $F_d(\cdot)$ も $F_c(\cdot)$ も原点にモードを以て単峰であることであると私達は主張したい。

与えられた分布関数 $F(\cdot)$ に対して(定義域の、普通は $(-\infty, \infty)$ の)各点 x に対して

$$(1.7) \quad F'_{\pm}(x) \equiv \lim_{h \rightarrow 0 \pm 0} \frac{F(x+h) - F(x)}{h} \quad (\text{複号同順})$$

とおく(この極限が定義されるためにはその点 x の近傍、即ち、 $a < x < b$ なる区間 (a, b) で $F(x)$ の値が定義されてなければならない)。ここに、 $F(\cdot)$ は右連続な関数であるから、各点 x では $F(x) = F(x+0)$ が成り立っている。 $F(\cdot)$ は任意の点で単調非減少である、つまり、 $x < y$ ならば $F(x) \leq F(y)$ となる。 $F(\cdot)$ の不連続点は、存在すれば第一種のものしか存在せず、その不連続点 x では $F(x) \equiv F(x+0) > F(x-0)$ となる。 $F'_{\pm}(x)$ というその極限値は存在すれば、 $F'_{\pm}(x) \geq 0$ である(不連続点 x では $F'_{-}(x) = +\infty$)。 $F(\cdot)$ が微分可能である点 x では $F'_{+}(x) = F'_{-}(x)$ である。 $E \equiv \{x : F'_{+}(x) F'_{-}(x) > 0\}$ とおく。Ibragimov(1956;255頁)の中にこの集合が定義されていて E は(古典的な定義での)単峰分布 $F(\cdot)$ に対しては区間になると主張されている。私達はそれが彼の論文の中で果たした役割が分からない。 $F(\cdot)$ の左右の微分係数が異なるのはその不連続点、つまり、跳躍点 $\{x_n : n = 0, \pm 1, \pm 2, \dots\}$ の所だけである。(離散分布の場合)どの跳躍点 x_n でも $F'_{+}(x_n) = 0, F'_{-}(x_n) = +\infty$ でそれらは E に含まれない。Ibragimov(1956;258頁↓8行目)の主張は結局「 $F(x)$ が集合 $E = (-\infty, \infty)$ の上で連続な二階導関数を持つ強単峰分布であるならばその密度の対数は $(-\infty, \infty)$ 上で凹である」ということなので E は、多分、その確率分布関数の台を表すものであると思う。以下はそうであるとして論を展開していく。私達はこの一般の分布関数 $F(\cdot)$ が強単峰であるための必要十分条件は(その定義域 $(-A, B)$ 上で) $\log \overline{f}(\cdot)$ が凹であることであると主張するが、その証明は連続部分と離散部分に分けて個別に行う。

さて、林(2012)の中で離散分布に対するヒンチンの表現定理に相当するものを初等的に証明しようとして私達は少し勘違いをしてしまった。その部分の修正は以下のようにしておくがそのときの主張は、結局、離散分布に対しては達成できていない。

2. 去年の発表結果の修正 X は(絶対)連続で Y が離散分布である場合(、但し、 X と Y は独立): $p_n \equiv P(Y=y_n)$ (但し、 $n=0, \pm 1, \pm 2, \dots$; $y_n < y_{n+1}$ で $y_0 \equiv 0$)であるとする(繰り返しになるが、もし元を与えられていた Y の分布が $y_0 \equiv 0$ の位置のものに確率質量がなかったとすると $p_0 \equiv 0$ と置く)。

$$(2.1) \quad F_Z(z) \equiv P(XY \leq z) = \sum_{n=-\infty}^{\infty} P(Xy_n \leq z) P(Y=y_n) = \sum_{n=-\infty}^{\infty} p_n P(Xy_n \leq z)$$

である。ここに、この Σ の中の確率 $P(Xy_n \leq z)$ は $n > 0$ のときは $F_X(z/y_n)$ で、 $n=0$ に対しては $z < 0$ のときはそれは不可能事象の確率で0であり、 $z \geq 0$ のときはそれは確実な事象の確率で1であり、 $n < 0$ のときは $y_n < 0$ であるから

$$(2.2) \quad P(Xy_n \leq z) = P(X \geq z/y_n) = 1 - P(X < z/y_n) \\ = 1 - F_X(z/y_n) + P(X = z/y_n) = 1 - F_X(z/y_n) \quad (\because F_X(\cdot) \text{は絶対連続})$$

である。よって、

$$(2.3) \quad F_Z(z) = p_0 H(z) + \sum_{n=1}^{\infty} \{p_{-n}(1 - F_X(z/y_{-n})) + p_n F_X(z/y_n)\}$$

となる。ここに、 $H(\cdot)$ はヘビサイド(Heaviside)関数で $z < 0$ では $H(z) = 0$ 、 $z \geq 0$ では $H(z) = 1$ なる関数である。計算により、 $F_Z(0) = (1 - F_X(0))F_Y(0) + F_X(0)(1 - F_Y(0) + p_0)$ であることは直ぐ分かる(このことから、興味深いことに、 $F_Z(0) = F_X(0)$ であったとするとそれは $F_X(0) = 0$ (これは X が非負の数値しか取らない確率変数である場合)か又は $F_Z(0) = (1 + p_0)/2$ の場合だけであることが分かる。後者の場合 $F_Z(0) \geq 1/2$ である)。そうするとその密度関数 $f_Z(\cdot)$ は拡張された関数で

$$(2.4) \quad f_Z(z) = p_0 \delta(z) + \sum_{n=1}^{\infty} \{-(p_{-n}/y_{-n}) f_X(z/y_{-n}) + (p_n/y_n) f_X(z/y_n)\}$$

となる。ここに、 $\delta(\cdot)$ は(先に定義した)ディラックのデルタ関数である。ここでは $f_Z(0) = p_0$ となる。

X が区間 $(0, 1)$ 上の一様分布であるとき、 $f_X(z/y_n)$ は $n > 0$ のとき $0 < z < y_n$ のとき値1を取り、その他の z の所で0であり、 $f_X(z/y_{-n})$ は $y_{-n} < z < 0$ のとき値1を取り、その他の z の所で0となるから、よって、 $f_Z(z)$ は各 $k=1, 2, \dots$ に対して以下の各区間で

$$z < 0 \text{ のとき} \qquad z > 0 \text{ のとき}$$

$$y_{-k} < z < y_{-k+1} \text{ では } \sum_{n=k}^{\infty} \frac{p_{-n}}{(-y_{-n})}; \quad y_{k-1} < z < y_k \text{ では } \sum_{n=k}^{\infty} \frac{p_n}{y_n}$$

という値を取る階段関数である。この分布は原点に確率 p_0 の確率質点を持つ区分的一様分布とも呼ぶべきものである。 $f_Z(0) = p_0$ であるから、これが $z=0$ をモードとする離散的な単峰関数であるためには、

$$(2.4) \quad \sum_{n=1}^{\infty} \frac{p_{-n}}{(-y_{-n})} \leq p_0 \quad \text{そして} \quad \sum_{n=1}^{\infty} \frac{p_n}{y_n} \leq p_0$$

でなければならないのでここではそのような場合を考える。これは、既に指摘したように、拡張された密度関数なのであって、それが真正(proper)であるためには、その和の順序を変えても良いとして

$$(2.5) \quad \sum_{n=1}^{\infty} \frac{n p_{-n}}{(-y_{-n})} + p_0 + \sum_{n=1}^{\infty} \frac{n p_n}{y_n} = 1$$

でなければならないことが分かる。(2.5)の左辺は確率変数 Y のほぼ (-1) 次の絶対モーメントである。 $f_Z(z)$ が $z=0$ で右連続になるのは原点に確率質量が有って(、つまり、 $p_0 > 0$)で

$$(2.6) \quad p_0 = \sum_{n=1}^{\infty} \frac{n p_n}{y_n}$$

のときだけである。それで Y が非負の確率変数だとすると、 $p_{-n} = 0$ ($n = 1, 2, \dots$) で、それは $p_0 = 1/2$ の場合に限ることになる。

逆に、以上の論理を逆に辿って、任意に与えられた $z = 0$ をモードとして確率質量 q_0 を持ち $y_{k-1} < z < y_k$ の区間で一様な密度 q_k を持つ区分的に単峰な分布を持つ確率変数 Z から対応する積表示の Y の分布 $\{p_n\}$ を (2.5) を満たすものであるとして $p_0 = q_0$, $q_n = (n p_n)/y_n$, $q_{-n} = (n p_{-n})/(-y_{-n})$ ($n = 1, 2, \dots$) から決めることでそのヒンチンの積表示が実現する。但し、上述の下線部分は微妙なのである。もっと精密には (2.5) を満たす確率分布 $\{p_n\}$ が存在するときということになるべきであろう。

3. その定義域で第1階導関数 $g'(\cdot)$ が存在する単峰な確率密度関数 $g(\cdot)$ と一般の確率分布関数 $F(\cdot)$ との畳み込み $h(\cdot)$ の形状：一般性を失うことなく $g(\cdot)$ のモードは原点に在るとする。

$$(3.1) \quad h(x) = \int_{-\infty}^{\infty} g(x-\tau) dF(\tau)$$

である。その積分記号の下その定義域のすべての点 x で項別微分可能であるとする

$$(3.2) \quad h'(x) = \int_{-\infty}^{\infty} g'(x-\tau) dF(\tau)$$

である。その積分区間を $(-\infty, x)$ と (x, ∞) に分けて書くと

$$(3.3) \quad h'(x) = H_1(x) + H_2(x)$$

となる。ここに、

$$(3.4) \quad H_1(x) \equiv \int_{-\infty}^x g'(x-\tau) dF(\tau); \quad H_2(x) \equiv \int_x^{\infty} g'(x-\tau) dF(\tau)$$

である(以下の議論は $H_1(x)$ も $H_2(x)$ も共に空ではないと仮定する; 例えば、 $g(x)$ が非負の数値しか取らない確率変数の確率密度関数であるとする、 $x < 0$ では $g(x) = 0$ であるから $H_2(x) = 0$ となって $H_2(x)$ が空になる)。そうすると $g(x)$ は原点がモードであるから ($g'(0) = 0$ で $x < 0$ では $g'(x) \geq 0$ で $x > 0$ では $g'(x) \leq 0$ となって) $-\infty < \tau < x$ で $g'(x-\tau) \leq 0$ そして $x < \tau < \infty$ で $g'(x-\tau) \geq 0$ となり(その定義域内の)任意の有限な x に対しては $H_1(x) < 0$ そして $H_2(x) > 0$ となる。負の値を取る関数 $H_1(x)$ は変数 x が大きくなるのに伴いその定義式の積分範囲が広がるから単調減少な関数であり、また $H_1(-\infty) = 0$, $H_2(-\infty) > 0$ は明らかである。それに対して $H_2(x)$ は正の値を取る関数でその定義式の積分範囲は x が大きくなるのに伴い狭まって行くのでこれもまた単調減少関数である。つまり、 $h'(\cdot)$ は二つの単調減少な関数の和になるのであるからその第二階の導関数が存在したならば任意の x に対して $h''(x) \leq 0$ でなければならない(この仮定は $h'(\cdot)$ の単調減少性にまで緩めることができるから本質的ではない)。 $H_1(\infty) < 0$, $H_2(\infty) = 0$ であることも明らかである。よって、 $h'(-\infty) > 0 > h'(\infty)$ となるからどこか1点 x_0 ($-\infty < x_0 < \infty$) で $h'(x_0) = 0$ となる。つまり、そういう点 x_0 が密度関数 $h(x)$ のモードになる(一意とは限らないことに注意)。以上の論法に誤りは無いであろう。そうすると微分可能な単峰分布との畳み込みで表される任意の分布は単峰ということになって強単峰性の議論が空しくなってしまう。しかし、単峰性が畳み込みに関して閉じてないことは知られているので、多分、上記の仮定が満たされないことの方が多いのに違いない。もう暫くその仮定の隙間を埋めてみる。

$H_2(\cdot)$ が空である場合(、これは $g(\cdot)$ が非負の確率変数の確率密度関数であるときや $g(x)$ が $x > 0$ で区分的に一様な分布であるときである): 定義域内の任意の点 x で $h'(x) < 0$ である。これは $h(\cdot)$ がその定義域内で単調減少であることを意味するから、その定義域の区間の左

端をモードとして(広義に)単峰であると言える(そのモードが定義域に含まれない可能性は有る)。

$H_1(\cdot)$ が空である場合(、これは $g(\cdot)$ が非正の確率変数の確率密度関数であるときや $g(x)$ が $x < 0$ で区分的に一様な分布であるときである): 定義域内の任意の点 x で $h'(x) > 0$ である。これは $h(\cdot)$ がその定義域内で単調増大であることを意味するから、その定義域の区間の右端をモードとして(広義に)単峰であると言える(そのモードが定義域に含まれない可能性は有る)。

この最後の二つの場合はその畳み込みの結果が狭義に単調減少または増大する確率密度関数になる場合でモードは無いと言っても良さそうな場合である。その定義域が有限であれば、その定義域の端点が定義域に含まれていない場合でも、モードは前者の時は左端、後者の時は右端であると言っても良いであろう。

$H_1(\cdot)$ も $H_2(\cdot)$ も共に空になる場合や(3.2)の項別微分が成立しない場合が未考察なのであるが、 $h(\cdot)$ が階段関数である場合 $h'(x) = 0$ a. e. であるのでこれは前者の場合に相当する。例えば、 $u(\cdot)$ を区間 $(0, 1)$ 上の一様密度であるとして

$$u_{(2)}(x) \equiv \int_{-\infty}^{\infty} u(x-\tau) u(\tau) d\tau$$

はそのような場合で $(0, 1)$ 上で $u'(x) = 0$ であるが $u'_{(2)}(x) = 0$ とはなっていない。 $u_{(2)}(x)$ は区間 $(0, 2)$ 上の三角確率密度で $x \in (0, 1)$ では $u_{(2)}(x) = x$; $x \in (1, 2)$ では $u_{(2)}(x) = 2 - x$ となつて $(0, 2)$ 上では連続であるが $x = 1$ では微分可能ではない。つまり、前節の主張の条件(積分記号下の項別微分可能性)を満足していない。

実は単峰性は畳み込み演算で閉じていない。古典的な単峰性の定義に対してChungによる英語版のGnedenko and Kolmogorov(1954)の254-255頁にその反例が与えられていると言われている。またFeller(Vol. II 1971版;168頁問題25)もその反例であるという指摘があるが、私達の計算に依ればその主張は怪しい。以上の注意はDharmadhikari and Joag-Dev(1988)の中に有るものであるが、次に説明するのはその本の11頁の例題1.1は拡大確率密度関数の場合である。

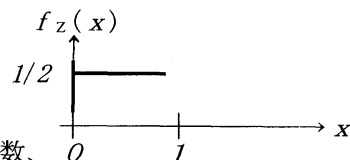
O は確率1で値0を取る確率変数、というか意味的には確実変数、 U は区間 $(0, 1)$ 上の値を一様な確率密度で取る確率変数、 Z は確率1/2で O 、確率1/2で U となる確率変数であるとする。その確率変数 Z は次のように構成することができる。 X は値0、1を伴に確率1/2で取る確率変数とし U とは独立であるとする。それは O とは当然独立である。そうすると

$$Z = XU + (1 - X)O$$

である。この確率変数の確率分布関数 $F_Z(\cdot)$ は

$$F_Z(x) = \frac{1}{2}H(x) + \frac{1}{2}U(x)$$

である、ここに、 $H(x)$ はヘビサイド関数、 $U(x)$ は同じ記号を使ってしまったが U の確率分布関数である。その確率密度関数 $f_Z(\cdot)$ は拡大確率密度関数で

$$f_Z(x) = \frac{1}{2}\delta(x) + \frac{1}{2}u(x)$$


で与えられる、ここに、 $\delta(x)$ はディラックのデルタ関数、 $u(x)$ は区間 $0 < x < 1$ 上の一様な確率密度関数である(これは原点で確率質量1/2を持ち、区間 $(0, 1)$ で一様な確率密度1/2を持つ)。これは原点ばかりではなく $(0, 1)$ 上のすべての点をモードに持つ単峰な関数である(このグラフ表示では区間 $[0, 1)$ 上の一様な密度関数とは

区別しにくい)。この Z と同じ分布を持つ独立な二つの確率変数 Z_1 と Z_2 の和 $Z_1 + Z_2$ の確率密度関数は $f_Z(x)$ 同志の畳み込み $(f_Z \star f_Z)(x)$ で、それはそのDharmadhikari and Joag-Dev(1988)の本に従えば、原点に確率質量 $1/4$ を持ち、区間 $(0, 2)$ 上で点 1 で一意な最大値 1 を持つ三角密度を持つから単峰ではないと言われているのであるが、私達が実際にその畳み込みを計算すると

$$(f_Z \star f_Z)(x) = \frac{1}{4} \delta(x) + \frac{1}{2} u(x) + \frac{1}{4} u_{(2)}(x)$$

で、ここに、 $u_{(2)}(x) = (u \star u)(x)$ はその三角密度である。これは上式の右辺の第2項の働きが有って、原点に確率質量 $1/4$ を持ち、区間 $(0, 1)$ 上で $(1/2) + (1/4)x$ 、区間 $[1, 2)$ 上で $(1/2) - (1/4)x$ となる密度で $x=1$ では値 $1/4$ を取るが、 $x=1-0$ では値 $3/4$ を取る。これはやはり単峰である。Dharmadhikari and Joag-Dev(1988;12頁)の例題1.2(Wolfe(1971)の例)はその反例になっていることは確認した。

4. Ibragimovの定理の初等的証明：私達は対象となる確率密度関数 $f(\cdot)$ はその定義域で微分可能であると仮定する。それが対数凹であるとは $g(\cdot) \equiv \log f(\cdot)$ がその定義域で凹関数であることである。

対数凹ならば単峰であることの証明：先ず $f(\cdot)$ が十分に滑らかでその第二階の導関数まで存在すると仮定し、 $g(\cdot)$ が(狭義に)凹であるとする。 $g(\cdot)$ を微分すると

$$g'(x) = \frac{f'(x)}{f(x)}; \quad g''(x) = \frac{f''(x)f(x) - \{f'(x)\}^2}{f^2(x)}$$

であるから(二階まで微分可能な)確率密度関数 $f(\cdot)$ が(狭義に)対数凹であるための必要十分条件は

$$g''(x) < 0 \Leftrightarrow f''(x)f(x) < \{f'(x)\}^2 \quad (\Leftrightarrow \frac{f''(x)}{f(x)} < \{\frac{f'(x)}{f(x)}\}^2)$$

である。その $f(\cdot)$ にモード(が存在するとしてそれ)を x_0 とする。そうすると $f'(x_0) = 0$ である。 $f(x_0) \neq 0$ なのであるから、 $g''(x_0) = f''(x_0)/f(x_0) < 0$ となって(モードの位置では) $f''(x_0) < 0$ となる(つまり、 $f(\cdot)$ も凹)。そのモードが一意であるとする、 $x < x_0$ では $f'(x) > 0$ で、そして $x > x_0$ では $f'(x) < 0$ でなければならない。ということで(モードを持つ)(二階まで微分可能な)(狭義に)対数凹な確率密度関数 $f(\cdot)$ は単峰であることが分かる。

その条件を少し緩めてその $f(\cdot)$ の第二階の微分係数が存在しないときであっても、 $f'(x)/f(x)$ が x と共に単調非増加であるという性質を指してその $f(\cdot)$ が(広義に)対数凹であると呼ぶことにする。つまり、その確率密度関数 $f(\cdot)$ はその台の任意の二点 $x < y$ に対して

$$(4.1) \quad \frac{f'(x)}{f(x)} \geq \frac{f'(y)}{f(y)}$$

という性質を持つと仮定する。そのとき $f'(x) \leq 0 \Rightarrow f'(y) \leq 0$ であり、 $f'(y) \geq 0 \Rightarrow f'(x) \geq 0$ である。これは $f'(x_0) = 0$ なる点 x_0 が存在すればそれがモードであるし、そうでなければその台の上の点 x でずっと $f'(x) \leq 0$ であるかまたは $f'(x) \geq 0$ であることを意味する。ずっと $f'(x) = 0$ であればそれはそれがその(有限な)台の上で一様分布するときである。つまり、その台の上で第一階の導関数が存在する確率密度関数 $f(\cdot)$ が単峰であるという性質は、モードが存在しないときも含めて、それが(広義に)対数凹であると呼ぶことができるのと同等である。その台の上の点 x でずっと $f'(x) \leq 0$ である場合 $f(\cdot)$ はその定義域全域で単調

非増加であり、 $f'(x) \geq 0$ の場合 $f(\cdot)$ はその定義域全域で単調非減少でモードが存在しない場合も含めた広義の単峰である。

最大値(=モード)が存在しないとき：一般性を失うことなく $f(x_0 \pm 0) = \infty$ とする。つまり、点 x_0 はその定義域に含まれない。それは x_0 より大きな値又は x_0 より小さな値しか取らない確率変数の確率密度である。 $x > x_0$ では $f'(x) \leq 0$ で、 $x < x_0$ では $f'(x) \geq 0$ であるからそういう場合は広義に単峰であると言える。

対数凹ならば強単峰であることの証明：次に対数凹な確率密度関数と任意の単峰分布との畳み込みの形状を調べる。さて、 $f(\cdot)$ は(広義に)対数凹な確率密度関数でその定義域は $(-A, B)$ であるとする、但し、 $A > 0, B > 0$ 。そのときその定義域内で第一階の導関数 $f'(\cdot)$ が存在してその定義域内の任意の二点 $x < y$ で(4.1)が成り立つ。つまり、単調非増加な関数 $\xi(\cdot)$ が存在して $f'(\cdot) \equiv \xi(\cdot) f(\cdot)$ と書くことができる。つまり、区間 $(-A, B)$ の中の任意の点 x で

$$(4.2) \quad f(x) = C \exp(\Xi(x))$$

と書くことができる、ここに、 $\Xi(x) \equiv \int_{-\infty}^x \xi(t) dt$ で C は $\int_{-\infty}^{\infty} \exp(\Xi(x)) dx$ の逆数である。

その $\xi(\cdot)$ は、勿論、そのような積分が存在するようなものであると仮定する。

一般性を失うことなくモードは $x = 0$ の位置に在るとすると $f'(0) = 0$ で $f(0) > 0$ であるから $\xi(0) = 0$ で $\xi(x)$ の形状から $x < 0$ では $\xi(x) \geq 0$ で $x > 0$ では $\xi(x) \leq 0$ となる。

単峰どころか任意の確率密度関数 $g(\cdot)$ との畳み込み $h(\cdot)$ は

$$(4.3) \quad h(x) = \int_{-\infty}^{\infty} f(x-\tau) g(\tau) d\tau$$

で、その定義域内では

$$(4.4) \quad h'(x) = \int_{-\infty}^{\infty} f'(x-\tau) g(\tau) d\tau = \int_{-\infty}^{\infty} \xi(x-\tau) f(x-\tau) g(\tau) d\tau$$

が成り立つ(この積分記号下での微分の正当性は気になる)。

$h(\cdot)$ が単峰であることを示すには或る点 x_0 で $h'(x_0) = 0$ であるならば $y > x_0$ なる任意の点 y で $h'(y) \leq 0$ であることを示せば十分である。(4.4)の積分区間を $(-\infty, x)$ と (x, ∞) に分けると前者の区間で $\xi(x) \leq 0$ 、後者の区間で $\xi(x) \geq 0$ であることを利用すると § 3 で展開したのと同じ論法で $h'(\cdot)$ がその定義域で単調非増加であることが分かるからである。

強単峰ならば対数凹であることの証明： $f(\cdot)$ は微分可能で強単峰なのに対数凹でないとする。つまり、(4.2)の形に書けるのだがその源の関数 $\xi(\cdot)$ が単調非減少ではないとする。しかし、強単峰であるから任意の単峰確率密度 $g(\cdot)$ に対して(4.3)で定義される $h(\cdot)$ は単峰になる。或る点 x_0 で $h'(x_0) = 0$ であったのに $x_1 > x_0$ なる点 x_1 で $h'(x_1) > 0$ となったら $h(\cdot)$ の単峰性に反するからである。

5. 離散分布の対数凹性の定義 確率質量関数 $p(x)$ 、 $p_n \equiv p(x_n)$ が対数凹であるとはどの $n = 0, \pm 1, \pm 2, \dots$ に対しても

$$(5.1) \quad \frac{(x_{n+1} - x_n) \log p_{n-1} + (x_n - x_{n-1}) \log p_{n+1}}{x_{n+1} - x_{n-1}} \leq \log p_n$$

が成り立つことであると定義する(Keilson and Gerber(1971)の強単峰性の定義はこの特殊な場合になっている)。実際、これを満足する数列 $\{p_n\}$ は単峰であることが示せる。添字の位置をずらせることにより、一般性を失うことなく数列 $\{p_n\}$ の最大値が $n = 0$ で生じていると仮定する。その仮定により $p_{-1} \leq p_0 \geq p_1$ である。(5.1)で $n = -1$ とおくと

$$\frac{(x_0 - x_{-1}) \log p_{-2} + (x_{-1} - x_{-2}) \log p_0}{x_0 - x_{-2}} \leq \log p_{-1}$$

となる。さて、 $p_{-2} \leq p_0$ であるから $\log p_{-2} \leq \log p_0 (< 0)$ となって

$$\frac{(x_0 - x_{-1}) \log p_{-2} + (x_{-1} - x_{-2}) \log p_{-2}}{x_0 - x_{-2}} \leq \log p_{-1}$$

であるから $\log p_{-2} \leq \log p_{-1}$ 、即ち、 $p_{-2} \leq p_{-1}$ であることが分かる。以下同様にして $n = 2, 3, \dots$ に対して順番に $p_{-(n+1)} \leq p_{-n}$ であることが示せる。(5.1)で $n = 1$ とおくと

$$\frac{(x_2 - x_1) \log p_0 + (x_1 - x_0) \log p_2}{x_2 - x_0} \leq \log p_1$$

となる。さて、 $p_2 \leq p_0$ であるから $\log p_2 \leq \log p_0 (< 0)$ となって

$$\frac{(x_2 - x_1) \log p_2 + (x_1 - x_0) \log p_2}{x_2 - x_0} \leq \log p_1$$

であるから $\log p_2 \leq \log p_1$ 、即ち、 $p_2 \leq p_1$ であることが分かる。以下同様にして $n = 2, 3, \dots$ に対して順番に $p_n \geq p_{n+1}$ であることが示せる。次の別証明を見るとこの性質が数列 $\{p_n\}$ が単調増大または単調減少である場合も含んでいることが分かる。

$$\text{別証明: } (5.1) \Leftrightarrow (x_{n+1} - x_n) \log p_{n-1} + (x_n - x_{n-1}) \log p_{n+1} \leq (x_{n+1} - x_n + x_n - x_{n-1}) \log p_n \\ \Leftrightarrow \frac{\log p_n - \log p_{n-1}}{x_n - x_{n-1}} \geq \frac{\log p_{n+1} - \log p_n}{x_{n+1} - x_n}$$

であることに注目しよう。これは、(4.2)と類似な表現をするならば、単調非増加な数列 ξ_n を使って

$$(5.2) \quad p_{n+1} = e^{\xi_n(x_{n+1} - x_n)} p_n \quad (n = 0, \pm 1, \pm 2, \dots)$$

と表すことができる(原点にモードが有るならば、 $n \leq 0$ で $\xi_n \geq 0$ 、そして $n > 0$ で $\xi_n \leq 0$ となる)。勿論、こういう数列が確率分布であるためには

$$(5.3) \quad p_n \geq 0 \text{ で } (n = 0, \pm 1, \pm 2, \dots) \sum_{n=-\infty}^{\infty} p_n = 1$$

でなければならない。(5.1)の性質を言葉で説明するならば、**対数確率の平均変化率は決して増加しない**、ということにでもなるだろうか。ところで、(5.1)から

$$(5.4) \quad \frac{p_{n+1}}{p_n} \geq 1 \text{ ならば } \frac{p_n}{p_{n-1}} \geq 1; \quad \frac{p_n}{p_{n-1}} \leq 1 \text{ ならば } \frac{p_{n+1}}{p_n} \leq 1$$

という性質が導ける。つまり、(片側方向への分布で単調増大、減少である場合も含めた)単峰であることが分かる(そのように拡大解釈するならば(5.4)は離散単峰性の定義(1.3)と同等であると言える)。

対数凹ならば強単峰であることの証明：そのためには二つの独立な離散分布 $p(x)$ 、 $p_n \equiv p(x_n)$ と $q(x)$ 、 $q_n \equiv q(y_n)$ との畳み込み $r(x)$ 、 $r_n \equiv r(z_n)$ の式を算出しなければならない。形式的には任意の実数 z に対して

$$(5.5) \quad r(z) = \sum_{p(x)q(z-x)>0 \text{ である } x \text{ に渡っての和}} p(x)q(z-x) = \sum_{p(z-x)q(x)>0 \text{ である } x \text{ に渡っての和}} p(z-x)q(x)$$

と与えられるが、一般的な離散分布の場合、その畳み込みの確率質点の位置を確定することは簡単ではない。難儀なことはその和が連続分布になる場合が有る。その連続分布が単峰になるのはどういうときかという研究はまだ進んでいない。

独立な離散分布の和(畳み込み)が連続分布になる場合： $p_n \equiv p(x_n) \equiv P(X=x_n)$ 、 $q_n \equiv p$

$(y_n) \equiv P(Y=y_n)$ で、ここに、 $x_n = nh$ 、 $y_n = nk$ ($n=0, \pm 1, \pm 2, \dots$)であるとする。また、この確率変数 X 、 Y は独立であるとする。そのとき $Z \equiv X+Y$ の分布はどうなるかを調べる。例えば、 $h=\sqrt{2}$ 、 $k=1$ のとき $X+Y$ は $m\sqrt{2}+n$ (m, n は整数) の形の数値を取る。この形の数値は一意に定まる。つまり、 $m_1\sqrt{2}+n_1=m_2\sqrt{2}+n_2$ ならば $m_1=m_2$ 、 $n_1=n_2$ である ($\because \sqrt{2}$ は無理数であるから $m\sqrt{2}+n=0 \Leftrightarrow m=0$ かつ $n=0$)。よって、

$$P(X+Y=m\sqrt{2}+n)=P(X=m\sqrt{2}, Y=n)=P(X=x_m, Y=y_n)=p_m q_n$$

である。 $\{m\sqrt{2}+n : m, n \text{ は整数}\}$ は実数空間 $\mathbf{R}$ で稠密である。すなわち、任意の実数 α に対して任意の正数 ε 毎に適当な整数 m, n で $|m\sqrt{2}+n-\alpha| < \varepsilon$ なるものが存在する(この命題は $\alpha=0$ のときは $\sqrt{2}$ が無理数であるという主張に他ならない)。

X, Y の特性関数 $\phi_X(t) \equiv E(e^{itX})$ 、 $\phi_Y(t)$ はそれぞれ

$$\phi_X(t) = \sum p_n e^{itx_n}; \quad \phi_Y(t) = \sum q_n e^{ity_n}$$

でこれらは任意の実数 $t \in (-\infty, \infty)$ に対して存在する。ここに、 $q \equiv 1-p$ 、 $p_n \equiv p q^{|n|-1}/2$ 、 $q_n \equiv p q^{|n|-1}/2$ ($n=\pm 1, \pm 2, \dots$)、 $x_0 \equiv 0$ 、 $p_0 \equiv 0$ 、 $y_0 \equiv 0$ 、 $q_0 \equiv 0$ とする(これらの確率分布が意味を持つためには $0 < p < 1$ でなければならない)。そうすると

$$\begin{aligned} \phi_X(t) &= \sum_{n=1}^{\infty} \frac{p q^{n-1}}{2} (e^{inht} + e^{-inht}) = \frac{p}{2} \left(\frac{e^{iht}}{1-qe^{it}} + \frac{e^{-iht}}{1-qe^{-it}} \right) \\ &= \sum_{n=1}^{\infty} p q^{n-1} \cos nht \\ &= \frac{p(\cos ht - q)}{1+q^2-2q\cos ht} \end{aligned}$$

を得る(これが実関数であることは注意しておこう)。ここで $h=1$ と置いた場合が $\phi_Y(t)$ であるから

$$\phi_Y(t) = \frac{p(\cos t - q)}{1+q^2-2q\cos t}$$

を得る(注意: この分母 $=(\cos t - q)^2 + \sin^2 t$ が 0 に成り得るのは $t=0 \pmod{2\pi}$ かつ $q=1$ のときだけである。そのときは分子も 0 になって不定形なのであるが、与えられた確率分布が意味の有る時は $q < 1$ であることに注意)。この分布が離散分布、取り分け、格子分布 (lattice distribution) であるということは $\phi_Y(t)=1$ なる $t \neq 0$ が存在していることから分かることである。実際、 $t \neq 0$ に対して $\phi_Y(t)=1 \Leftrightarrow \cos t = (2q)/(1+q)$ で任意に与えられた $0 < q < 1$ に対してその右辺の数は $0 < (2q)/(1+q) < 1$ であるから、そのような $t \neq 0$ は確かに存在する。 $\phi_X(t)$ に関しても同様である。さて、 $Z \equiv X+Y$ の特性関数 $\phi_Z(t) = \phi_X(t) \phi_Y(t)$ は

$$\phi_Z(t) = \frac{p^2(\cos t - q)(\cos ht - q)}{(1+q^2-2q\cos t)(1+q^2-2q\cos ht)}$$

であることが分かる。 $|\phi_Z(t)|=1$ なる $t \neq 0$ は存在しないし、

$$\limsup_{|t| \rightarrow \infty} |\phi_Z(t)| = 0$$

であるから $\phi_Z(t)$ は連続分布に属する (Luckacs(1970;20頁))。

その二つの離散分布 $p(\cdot)$ と $q(\cdot)$ の畳み込みがまた離散分布であると仮定して、形式的に

$$\triangle r(z) = \sum_{p(x) \triangle q(z-x) \neq 0} p(x) \triangle q(z-x)$$

を得る。ここに、例えば $p(\cdot)$ に対して $\triangle p(\cdot)$ はその質点 x_n で $\triangle p(x_n) \equiv p_{n+1} - p_n$ なる質

量を持つ関数でその値が0であってもその質点と質量は明示されるものとする。また $q(\cdot)$ は原点をモードとする単峰分布であるとする。その x についての和を $z \geq x$ の部分 $R_1(z)$ と $z < x$ の部分 $R_2(z)$ に分けて書くと $\triangleleft r(z) = R_1(z) + R_2(z)$ でその確率質点の位置で $R_1(z) \leq 0$ で $R_2(z) \geq 0$ 、つまり、それらが存在すれば、 $R_1(\cdot)$ は非正で $R_2(\cdot)$ は非負の関数である。定義から、 $R_1(-\infty) = 0$, $R_1(\infty) < 0$, $R_2(\infty) = 0$, $R_2(-\infty) > 0$ である。離散分布の場合 $R_1(\cdot)$ も $R_2(\cdot)$ 両方とも空となることはない。 $r(\cdot)$ は確率質量関数であるから $\triangleleft r(\infty) = 0$ である。 $\triangleleft r(-\infty) = 0$ であるべきということと $R_2(-\infty) > 0$ という要件は両立しない(後者の要件に対しては或る有限な z_* に対して $R_2(z_*) > 0$ であるとする)。 $R_2(z)$ は確率質点で正の値を取る関数でその定義式の和の範囲は z が大きくなるのに伴い狭まって行くので単調減少関数である。 $R_1(z)$ は確率質点で負の値を取る関数でその定義式の和の範囲は z が大きくなるのに伴い広がって行くので単調減少関数である。つまり、 $\triangleleft r(\cdot)$ は二つの単調減少関数の和であるからその第二階の差分は任意の確率質点 z に対して $\triangleleft^2 r(z) \leq 0$ でなければならない。よって、 $\triangleleft r(z_*) > 0 > \triangleleft r(\infty)$ となるからどこか1点 z_0 ($-\infty < z_0 < \infty$) で $\triangleleft r(z_0) = 0$ となる。つまり、そういう点 z_0 が確率質量関数 $r(\cdot)$ のモードになる(それは一意とは限らないことに注意)。 $R_1(\cdot)$ が空のとき $r(\cdot)$ は非正の値を取る確率変数の確率質量関数で単調非減少で原点がモードになる。 $R_2(\cdot)$ が空のとき $r(\cdot)$ は非負の値を取る確率変数の確率質量関数で単調非増大で原点がモードになる(以上の議論は $p(\cdot)$ が対数凹でない任意の確率質量関数であっても成立しているから怖い)。

強単峰ならば対数凹であることの証明: $p(\cdot)$ は強単峰な確率質量関数なのに対数凹でないとする。つまり、(5.2)の形に書けるのだがその源の数列 ξ_n が単調非減少ではないとする。しかし、強単峰であるから任意の単峰確率質量関数 $q(\cdot)$ に対して(5.5)で定義される $r(\cdot)$ は単峰になる。或る点 x_0 で $\triangleleft r(x_0) = 0$ であつたのに $x_1 > x_0$ なる点 x_1 で $\triangleleft r(x_1) > 0$ となつたら $r(\cdot)$ の単峰性に反するからである。

参考文献:

- Chung, K.L., A Course in Probability Theory, Harcourt, New York, 1968
 Feller, An Introduction to Probability Theory and Its Applications, Vol. II, 1971
 B.V. Gnedenko and A.N. Kolmogorov, Limit Distributions for Sums of Independent Random Variables, Moscow-Leningrad, 1949 (In Russian)
 林 芳男、「単峰分布についての一考察」京都大学数理解析研究所講究録2012年7月、207-213(発表会での原題: Keilsonの離散単峰性理論について)
 I.A. Ibragimov, "On the Composition of Unimodal Distributions," Theory of Probability and Its Applications, Volume 1, 1956, 255-260.
 J. Keilson and H. Gerber, "Some Results for Discrete Unimodality," *Journal of the American Statistical Association*, June 1971, Volume 66, Number 334, Theory and Methods Section, 386-389.
 J. Keilson, Markov Chains Models—Rarity and Exponentiality, Springer-Verlag New York Heidelberg Berlin, 1979.
 E. Lukacs, Characteristic Functions, Second Edition, 1970, Griffin London,
 E. Parzen, Modern Probability Theory and Its Applications, John Wiley & Sons, Inc., 1960.
 Sudhakar Dharmadhikari and Kumar Joag-Dev, Unimodality, Convexity, and Applications, Academic Press, Inc., 1988