

メソリンガムモデルとその周辺

松井メモ

株式会社ビデオリサーチ

メソリンガムモデル式及びその変形

1. 視聴回数分布と重複視聴率の関係

- N時点からとり出した任意の k時点の重複視聴率の合計を Q_k とする。
- 丁度 i回視聴した人の割合を $f(i)$ とする。

$$\text{公式(1)} \quad Q_k = \sum_{i=k}^N {}_i C_k f(i)$$

〈証明〉

$i \geq k$ の時 $f(i)$ に数えられている視聴者は、 i 個の中から k 個とった任意の組み合わせの視聴者であるから、 Q_k の中に ${}_i Q_k$ 回、数えられている。
 $i < k$ である時は $f(i)$ に数えられている視聴者は存在しない。従って $i = k, k+1, \dots, N$ の全てについて、 ${}_i Q_k f(i)$ を合計すれば過不足なく Q_k に一致する。 Q. E. D.

2. 到達率算出の一般公式

- $R(i) = \sum_{j=i}^N f(j)$ とおく。 $R(i)$ は i 回以上視聴の到達率である。

$$\text{公式(2)} \quad f(i) = R(i) - R(i+1)$$

〈証明〉

$$\begin{aligned} R(i) &= \sum_{j=i}^N f(j) = f(i) + \sum_{j=i+1}^N f(j) = f(i) + R(i+1) \\ \therefore f(i) &= R(i) - R(i+1) \end{aligned}$$

$$\text{公式 (3)} \quad Q_k = \sum_{i=k}^N {}_{i-1}C_{k-1} R(i)$$

〈証明〉

公式(1)に公式(2)を代入すると

$$\begin{aligned} Q_k &= \sum_{i=k}^N {}_iC_k (R(i) - R(i+1)) \\ &= {}_kC_k (R(k) - R(k+1)) \\ &\quad + {}_{k+1}C_k (R(k+1) - R(k+2)) \\ &\quad + \Lambda \\ &\quad + {}_NC_k (R(N) - 0) \quad (R(N+1) = 0) \\ &= {}_kC_k R(k) \\ &\quad + ({}_{k+1}C_k - {}_kC_k) R(k+1) \\ &\quad + \Lambda \\ &\quad + ({}_iC_k - {}_{i-1}C_k) R(i) \\ &\quad + \Lambda \\ &\quad + ({}_NC_k - {}_{N-1}C_k) R(N) \quad (R(i) \text{ について整理}) \end{aligned}$$

ここで ${}_iC_k - {}_{i-1}C_k = {}_{i-1}C_{k-1}$ 及び ${}_kC_k = {}_{k-1}C_{k-1} = 1$ だから

$$\begin{aligned} Q_k &= {}_{k-1}C_{k-1} R(k) \\ &\quad + {}_kC_{k-1} R(k+1) \\ &\quad + \Lambda \\ &\quad + {}_{i-1}C_{k-1} R(i) \\ &\quad + \Lambda \\ &\quad + {}_{N-1}C_{k-1} R(N) \\ &= \sum_{i=k}^N {}_{i-1}C_{k-1} R(i) \end{aligned}$$

Q. E. D.

$Q(k)$ から $f(k)$ を求める方法

$$Q(x) = \sum_{i=x}^N {}_i C_x f(i) \quad \text{としておく。}$$

θ についての次の N 次多項式をつくる。

$$\begin{aligned} g(\theta) &= \sum_{x=0}^N \theta^x Q(x) \\ &= \sum_{x=0}^N \sum_{i=x}^N {}_i C_x \theta^x f(i) \end{aligned}$$

和の順序を変更すると

$$\begin{aligned} g(\theta) &= \sum_{i=0}^N \left(\sum_{x=0}^i {}_i C_x \theta^x \right) f(i) \\ &= \sum_{i=0}^N (\theta+1)^i f(i) \end{aligned}$$

$$\text{故に} \quad \sum_{x=0}^N \theta^x Q(x) = \sum_{x=0}^N (\theta+1)^x f(x)$$

両辺を θ について k 回微分すれば

$$\begin{aligned} \sum_{x=k}^N x(x-1)\dots(x-k+1)\theta^{x-k} Q(x) \\ = \sum_{x=k}^N x(x-1)\dots(x-k+1)(\theta+1)^{x-k} f(x) \end{aligned}$$

ここで $\theta=-1$ を代入すれば

$$\sum_{x=k}^N x(x-1)\dots(x-k+1)(-1)^{x-k} Q(x) = k! f(k)$$

$$\begin{aligned} (4) \quad \therefore f(k) &= \sum_{x=k}^N \frac{x!}{k!(x-k)!} (-1)^{x-k} Q(x) \\ &= \sum_{x=k}^N (-1)^{x-k} {}_x C_k Q(x) \end{aligned}$$

(付録)

$P_k = \int_0^1 P^k w(p) dp$ が全ての k について成立する $w(p)$ が存在する場合

$$f(k) = {}_N C_k \int_0^1 P^k w(1-p)^{N-k} w(p) dp$$

〈証明〉

$$\begin{aligned} f(k) &= \sum_{x=k}^N (-1)^{x-k} {}_x C_k Q(x) \\ &= \sum_{x=k}^N (-1)^{x-k} {}_x C_k {}_N C_x P_x \end{aligned}$$

$$\begin{aligned} {}_x C_k {}_N C_x &= \frac{x!}{k!(x-k)!} \cdot \frac{N!}{x!(N-x)!} \\ &= \frac{N!}{k!(N-k)!} \cdot \frac{(N-k)!}{(x-k)!(N-x)!} \\ &= {}_N C_k \cdot {}_{N-k} C_{x-k} \end{aligned}$$

$$\begin{aligned} \therefore f(k) &= \sum_{x=k}^N (-1)^{x-k} {}_N C_k {}_{N-k} C_{x-k} P_x \\ &= {}_N C_k \sum_{x=k}^N {}_{N-k} C_{x-k} (-1)^{x-k} \int_0^1 P^x w(p) dp \\ &= {}_N C_k \int_0^1 P^k \sum_{x=k}^N {}_{N-k} C_{x-k} (-1)^{x-k} P^{x-k} w(p) dp \\ &= {}_N C_k \int_0^1 P^k (1-P)^{N-k} w(p) dp \end{aligned}$$

Q. E. D.

$Q(k)$ から $R(k)$ を求める方法

$$Q(x) = \sum_{i=x}^N C_{x-1} R(i) \quad (1)$$

θ についての多項式 $g(\theta)$ を次式で定義する。

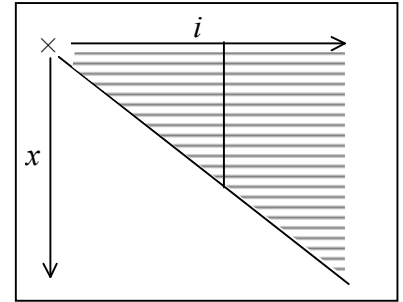
$$g(\theta) = \sum_{x=1}^N \theta^{x-1} Q(x) \quad (2)$$

(2)に(1)を代入して変形する。

$$\begin{aligned} g(\theta) &= \sum_{x=1}^N \theta^{x-1} \sum_{i=x}^N C_{x-1} R(i) \\ &= \sum_{x=1}^N \sum_{i=x}^N C_{x-1} \theta^{x-1} R(i) \end{aligned}$$

和の順序を変更すると

$$\begin{aligned} g(\theta) &= \sum_{i=1}^N \sum_{x=1}^i C_{x-1} \theta^{x-1} R(i) \\ &= \sum_{x=1}^N R(i) \left\{ \sum_{x=0}^i C_{x-1} \theta^{x-1} \right\} \\ &= \sum_{i=1}^N R(i) \cdot (Q+1)^{i-1} \cdot \theta \end{aligned}$$



故に

$$\sum_{x=1}^N \theta^{x-1} Q(x) = \sum_{x=1}^N (\theta+1)^{x-1} R(x)$$

両辺を Q について $(k-1)$ 回微分する

$$\begin{aligned} \sum_{x=k}^N (x-1) \dots (x-k+1) \theta^{x-k} Q(x) \\ = \sum_{x=k}^N (x-1)(x-2) \dots (x-k+1) (\theta+1)^{x-k} R(x) \end{aligned}$$

ここで $\theta = -1$ を代入すれば

$$\begin{aligned} \sum_{x=k}^N \frac{(x-1)!}{(x-k)!} (-1)^{x-k} Q(x) &= (k-1)! R(k) \\ \therefore R(k) &= \sum_{x=k}^N \frac{x!}{k!(x-k)!} (-1)^{x-k} Q(x) \end{aligned} \quad (4)'$$

公式(5) N時点から取った任意のk時点のあらゆる組合わせについての重複視聴率の平均を $P(k)$ とする時

$$R(a) = \sum_{k=a}^N (-1)^{k-a} {}_{k-1}C_{a-1} {}_N C_k P(k)$$

〈証明〉

$$P_k = \frac{Q_k}{{}_N C_k} \quad \text{であるから} \quad Q_k = {}_N C_k P_k$$

これを公式(4)へ代入すれば上式が得られる。

3. メソリンガムモデル

メソリンガムモデルは、公式(5)における $P(k)$ を一定の仮定の下に導く。
その仮説は

- ・視聴確率 r であるグループの大きさは r についてのベータ分布

$$\beta(r) \text{で与えられる。 } \beta(r) = \frac{1}{B(l, m)} r^{l-1} (1-r)^{m-1}$$

- ・ $P(k)$ は $\int_0^1 r^k \beta(r) dr$ で算出される。

公式(6)

$$P(k) = \frac{B(l+k, m)}{B(l, m)} = \frac{l(l+1)\dots(l+k-1)}{(l+m)(l+m+1)\dots(l+m+k-1)}$$

〈証明〉

$$\begin{aligned} P(k) &= \int_0^1 \frac{1}{B(l, m)} r^{l+k-1} (1-r)^{m-1} dr \\ &= \frac{B(l+k, m)}{B(l, m)} \end{aligned}$$

Q. E. D.

メソリンガムモデルによるフリークエンシーディストリビューションの算出はデータから l, m を定め $P(k)$ を算出し、これを公式(5)に代入し、更に公式(2)を用いて $f(i)$ を求めるという手順から構成されています。

4. メソリンガムモデルに於ける算出手順の変更

上記算出手順はかなり複雑であり、特に公式(5)の計算はめんどろです。しかし、これは到達率算出の一般公式を求めて点にその根本的な理由があります。メソリンガムモデルの仮説そのものから直接に求める簡略式を導くことが可能です。そのために仮説を次のように言い変えます。

- ・視聴確率 r であるグループの大きさは r についてのベータ分析 $\beta(r)$ で与えられる。(これは先と同じ)
- ・特定の r をとる時、そのグループにおける視聴回数の分布は二項分布 ${}_N C_i r^i (1-r)^{N-i}$ である。

この仮説から導く時全体の視聴回数の分布は

公式(7)

$$f^*(i) = {}_N C_i \frac{B(l+i, m+N-i)}{B(l, m)}$$

〈証明〉

$$\begin{aligned} f^*(i) &= \int_0^1 {}_N C_i r^i (1-r)^{N-i} \cdot \frac{r^{l-1} (1-r)^{m-1}}{B(l, m)} dr \\ &= \frac{{}_N C_i}{B(l, m)} \int_0^1 r^{l+i-1} (1-r)^{m+N-i-1} dr \\ &= {}_N C_i \frac{B(l+i, m+N-i)}{B(l, m)} \end{aligned}$$

Q. E. D.

この公式が先の公式と同値であることはその論理構成から当然ではありますが、一応平均重複視聴率 P_k^* が一致することを確認しておきます。(P_k^* さえ一致すれば、フリークエンシーディストリビューション全体が一致することは既に見たところから明らかです。)

公式(1)により

$$\begin{aligned}
Q_k^* &= \sum_{i=k}^N C_k f^*(i) \\
&= \sum_{i=k}^N C_k \int_0^1 {}_N C_i r^i (1-r)^{N-i} \cdot \frac{r^{l-1} (1-r)^{m-1}}{B(l,m)} dr \\
&= \int_0^1 \left(\sum_{i=k}^N {}_i C_k {}_N C_i \cdot r^i (1-r)^{N-i} \right) \frac{r^{l-1} (1-r)^{m-1}}{B(l,m)} dr \\
&= \int_0^1 {}_N C_k r^k \left(\sum_{i=k}^N {}_{N-k} C_{i-k} r^{i-k} (1-r)^{N-i} \right) \frac{r^{l-1} (1-r)^{m-1}}{B(l,m)} dr \\
&= {}_N C_k \int_0^1 r^k \frac{r^{l-1} (1-r)^{m-1}}{B(l,m)} dr \\
&= {}_N C_k \cdot P_k
\end{aligned}$$

一方

$$Q_k^* = {}_N C_k P_k^*$$

故に

$$P_k^* = P_k$$

公式(8)

$$f^*(0) = \frac{m(m+1)\dots(m+N-1)}{(l+m)(l+m+1)\dots(l+m+N-1)}$$

$$f^*(i+1) = f^*(i) \times \frac{N-1}{i+1} \times \frac{l+1}{m+N-i-1}$$

〈証明〉 $f^*(0)$ は公式(7) で $i=0$ を入れれば直ちに得られる。また、第二の式は、

$$\begin{aligned}
\frac{f^*(i+1)}{f^*(i)} &= \frac{{}_N C_{i+1}}{{}_N C_i} \times \frac{B(l+i+1, m+N-i-1)}{B(l+i, m+N-i)} \\
&= \frac{N-i}{i+1} \times \frac{l+1}{m+N-i-1}
\end{aligned}$$

より明らか。

従ってこの場合の計算手順は、 l, m をデータから定めることができさえすれば、 $f^*(0)$ からスタートして順次公式(8)の第2式を用いて直接フリークエンシーディストリビューションを算出するという形に簡略化される。

1. モデル式のベースとなっている仮説

視聴率（閲読率の場合も同じです。以下の文中でも同様）がPである時点N個によって得られる視聴回数の分布は、

- ・各時点の視聴率は全て等しくPである。
- ・各時点間に関連性がなく全て独立の事象である。

という二つの仮説があれば、二項分布

$$f(k) = {}_N C_k P^k (1-P)^{n-k}$$

で与えられます。

しかし通常この二つの条件は満たされませんので、二項分布により視聴回数分布の推定は精度が悪く、用いることができません。

メソリンガムモデル及び負の二項分布モデルは上の二つの仮説を修正することによって視聴回数分布の適合度を高くするわけですが、その修正に於ける基本的な発想は、

- ・視聴率Pを単一の数字と見ず、異なる視聴確率を持った複数のグループが存在し、それらのグループの視聴率から合成されたものとする。

という点にあります。そして各グループ内での視聴回数の分布は上記の二項分布に従うと考えるわけです。従って、この場合の最終的な視聴回数の分布は、二項分布を複数個合成したものになります。

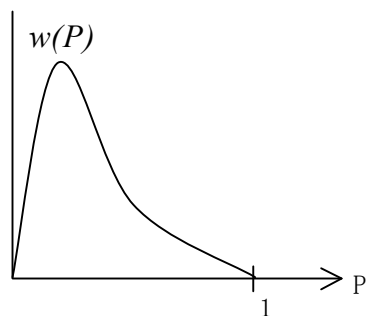
二項分布を合成するために、視聴確率P（これは0から1まで変わる連続的な変数です）に対応するグループの大きさが必要になりますが、この各グループの大きさを与える方法という点でメソリンガムモデルと負の二項分布モデルの相異が出てきます。

・ <メソリンガムモデルの場合の仮説>

視聴確率 P に対応するグループの大きさ $w_1(P)$ は β 分布で表される。

$$w_1(P) = \frac{1}{B(l, m)} P^{l-1} (1-P)^{m-1}$$

β 分布の通常形態は、下図に示されます。



負の二項分布の場合にはグループの大きさを P の関数としてではなく、 $N \times P$ つまり各グループの $G R P$ の関数と考えます。（この場合の $G R P$ は視聴確率の合計という意味です。）

・ <負の二項分布の場合の仮説>

$G R P$ が g であるグループの大きさ $w_2(g)$ は Γ 分布で表される。

$$w_2(g) = \frac{b^m}{\Gamma(m)} g^{m-1} e^{-bg}$$

Γ 分布をウエイトとして二項分布を合成する場合には、二項分布の代りにポアソン分布を用います。

メソリンガムモデル及び負の二項分布モデルが導出される仮説は以上ですから；両分布の公式は次のようになります。

(1) メソリンガムモデル

$$\begin{aligned} f(k) &= \int_0^1 w_1(P) {}_N C_k P^k (1-P)^{N-k} dP \\ &= {}_N C_k \frac{B(l+k, m+N-k)}{B(l, m)} \end{aligned}$$

(2) 負の二項分布モデル

$$\begin{aligned} f(k) &= \int_0^1 w_2(g) \frac{g^k}{k!} e^{-g} dg \\ &= \left(\frac{b}{b+1} \right)^m \frac{m(m+1)\dots(m+k-1)}{(b+1)^k \cdot k!} \\ &= {}_{m+k-1}C_{m-1} p^m (1-p)^k \end{aligned}$$

(但し、 $P = \frac{b}{b+1}$)

2. モデル式に含まれるパラメーターの決定方法

メソリンガムモデルも負の二項分布モデルも、その式の中に二つのパラメーターを含んでいます。この二つのパラメーターはデータとして与えられなければなりませんし、またもし与えられれば、数式は完全に決定されてしまいます。

パラメーターを決めるための手段として次の式を用いられます。

$${}_N C_k \cdot \overline{P}_k = \sum_{i=k}^N {}_i C_1 f(i) \quad (1 \leq k \leq N) \quad \text{<注 1>}$$

\overline{P}_k ; N時点の中から取った任意のk時点の重複視聴率
のあらゆる組合せについての平均

$f(i)$; 視聴回数*i*回の視聴者の割合

この式の左辺に実測値から算出した \overline{P}_k を入れ、右辺にパラメーターを含んだ $f(i)$ の式を入れれば、パラメーターについての方程式が得られますから、これを解いてパラメーターを決定するわけですが、パラメーターの数は二つだけですから実際には \overline{P}_1 及び \overline{P}_2 についての方程式を用いることになります。

メソリンガムモデルの場合のパラメーターを決定する方程式は、右辺の和を計算すると、

$$\overline{P}_1 = \frac{l}{l+m}$$

$$\overline{P}_2 = \frac{l}{l+m} \cdot \frac{l+1}{l+m+1}$$

になります。これを解いて、 l, m は

$$l = \frac{\overline{P}_1(\overline{P}_1 - \overline{P}_2)}{\overline{P}_2 - \overline{P}_1^2}$$

$$m = \frac{(1 - \overline{P}_1)(\overline{P}_1 - \overline{P}_2)}{\overline{P}_2 - \overline{P}_1^2}$$

負の二項分布モデルの場合のパラメーターを決定する方程式は、

$$\bar{P}_1 = m \cdot \frac{1-P}{P} \cdot \frac{1}{N}$$

$$\bar{P}_2 = m(m-1) \cdot \left(\frac{1-P}{P}\right) \cdot \frac{1}{N(N-1)}$$

になり、これを解くと、 m 、 P は

$$m = \frac{N\bar{P}_2^2}{N(\bar{P}_2 - \bar{P}_1^2) - \bar{P}_2}$$

$$P = \frac{\bar{P}_1}{N(\bar{P}_2 - \bar{P}_1^2) + \bar{P}_1 - \bar{P}_2}$$

になります。

<注1> ${}_N C_k \bar{P}_k = \sum_{i=k}^N {}_i C_k f(i)$ の証明

左辺は k 時点重複視聴率をあらゆる k 時点について合計したものと
同じです。こういう合計をした場合、 $i(\geq k)$ 回視聴者は i 時点の
中の任意の k 時点の重複視聴者ですから i 時点から k 時点をとり出
す組合せの数だけ、この合計の中に算えこまれています。従って
上の式が成立します。

3. パラメーター決定のためのデータ

パラメーター決定のためには \overline{P}_1 及び \overline{P}_2 が必要です

(1) \overline{P}_1

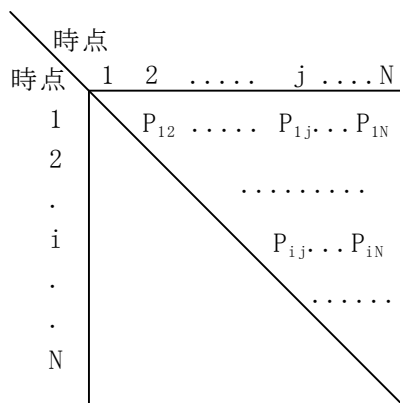
\overline{P}_1 は平均視聴率に他なりませんから、とり上げているN時点の視聴率 P_i から作成されます

$$\overline{P}_1 = \frac{1}{N} \sum_{i=1}^N P_i$$

(2) \overline{P}_2

\overline{P}_2 はN時点のうちの任意の2時点の重複視聴率の平均ですから、あらゆる2時点の組合せについて重複視聴率 P_{ij} を実測して下図のマトリックスを埋めれば作成できます。

$$\overline{P}_2 = \frac{2}{N(N-1)} \sum_{i < j} P_{ij}$$



(テレビ、ラジオの時点は膨大になり、その組合せについて P_{ij} をデータして用意しておくことは著しく不経済ですから、当社では P_{ij} を実測によらないで、要因分析によって推定する方法をとっています。)

4. 視聴回数分布推定の作業ステップ

メソリンガムモデル又は負の二項分布モデルによって視聴回数の分布を推定するための作業ステップは、以上の説明を逆にたどってゆけばよいこととなります。

- (1) 時点の設定
- (2) (i) 各時点の視聴率の設定
(ii) 二時点間の重複視聴率の設定
- (3) (i) 平均視聴率 \bar{P}_1 の算出
(ii) 二時点間平均重複視聴率 \bar{P}_2 の算出
- (4) モデルにインプットすべきパラメーターの算出
- (5) 視聴回数分布のモデル計算

メソリンガムモデル及び負の二項分布モデルにおけるパラメーターの条件及び平均視聴率、平均重複視聴率の間の制約

1. メソリンガムモデルの場合

モデルのベースとなっている β 分布のパラメーター l, m について $l > 0, m > 0$ という条件が課されている。(これは β 関数 $B(l, m)$ の収束条件に他ならない。) フリークエンシーディストリビューション $f(k)$ を求める時の積分は l が $l+k$ になり、 m が $m+N-k$ になるだけであるから収束の条件は常に満たされている。($k \geq 0, N-k \geq 0$)

以上によってパラメーター l, m に課せられる条件は

$$l = \frac{P_1(P_1 - P_2)}{P_2 - P_1^2} > 0 \quad (1)$$

$$m = \frac{(1 - P_1)(P_1 - P_2)}{P_2 - P_1^2} > 0 \quad (2)$$

の二式で表わされる。これに加えて、 P_1 は平均視聴率であるから常に

$$0 < P_1 < 1 \quad (3)$$

をみたしていることを考慮すれば、(1), (2) より

$$P_1^2 < P_2 < P_1 \quad (4)$$

が導かれることは明らかである。

2. 負の二項分布モデルの場合

モデルのベースとなっている Γ 分布のパラメーター α, β について $\alpha > 0, \beta > 0$ という条件が課されている。

フリークエンシーディストリビューションを求める時の積分は α が $\alpha + k$ になり、 β が $\beta + 1$ に変わるだけであるから収束の条件は満たされている。

従ってパラメーターに課される条件は

$$\alpha = \frac{Q_1^2}{2Q_2 - Q_1^2} > 0 \quad (5)$$

$$\beta = \frac{Q_1}{2Q_2 - Q_1^2} > 0 \quad (6)$$

の二式である。また Q_1 は常に正であるから Q_1 及び Q_2 の間に

$$2Q_2 - Q_1^2 > 0 \quad (7)$$

という関係が与えられれば、必要十分なことは明らか。

これを N, P_1, P_2 に書き直せば、

$$\begin{aligned} N(N-1)P_2 - N^2P_1^2 &> 0 \\ \therefore P_2 &> \frac{N}{N-1}P_1^2 \end{aligned} \quad (8)$$

(8)式が負の二項分布モデルが成立するための必要条件であるが、重複視聴率 P_2 の実体的な意味から考えて

$$P_2 < P_1 \quad (9)$$

という条件が課させることは当然であるから

$$\frac{N}{N-1}P_1^2 < P_2 < P_1 \quad (10)$$

によって P_2 に制約を加えておくことが妥当であろう。

3. P_2 が条件式の下限に近づく時の極限分布

i) メソリンガムモデルの場合

視聴確率 P の分布 $\beta(p)$ の平均及び分散は、

$$\bar{P} = \int_0^1 P \beta(p) dp = \frac{B(l+1, m)}{B(l, m)} = \frac{l}{l+m} = P_1$$

$$\begin{aligned} V(P) &= \int_0^1 P^2 \beta(p) dp - \bar{P}^2 = \frac{B(l+2, m)}{B(l, m)} - \bar{P}^2 \\ &= \frac{l(l+1)}{(l+m)(l+m+1)} - P_1^2 = P_2 - P_1^2 \end{aligned}$$

ここで P_1 を一定としておいて $P_2 \rightarrow P_1^2$ とすれば $\beta(p)$ は一点 P_1 に集中する。(分散が0になる。)

従って、視聴確率 P_1 の視聴者のみが存在することになり、視聴回数の分布は当然二項分布に収束する。

(これは、 $f(k)$ の極限から直接確かめることも可能)

ii) 負の二項分布モデルの場合

延視聴確率の意味での G R P の分布 $\gamma(g)$ の平均と分散は、

$$\bar{g} = \int_0^{\infty} g \gamma(g) dg = \frac{\beta^2}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} = \frac{\alpha}{\beta} = Q_1 = NP_1$$

$$\begin{aligned} V(g) &= \int_0^{\infty} g^2 \gamma(g) dg - \bar{g}^2 = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+2)}{\beta^{\alpha+2}} - N^2 P_1^2 \\ &= \frac{\alpha(\alpha+1)}{\beta^2} - n^2 P_1^2 = 2Q_2 - n^2 P_1^2 = N(N-1) \left\{ P_2 - \frac{N}{N-1} P_1^2 \right\} \end{aligned}$$

従って n , P_1 を固定して $P_2 \rightarrow \frac{N}{N-1} P_1^2$ とする場合、この分布は分散が0

に収束し、一点 $n P_1$ に集中してしまふ。したがって視聴回数の分布は当然ポアソン分布に収束する。(これも $f(k)$ の極限から直接確かめることができる。)

4. P_2 が条件式の上限に近づく時の極限分布

モデルの如何を問わず、重複視聴率が視聴率そのものに一致してくるわけであるから、この場合にはリーチが全く伸びないということと同じである。従って視聴回数の分布は、N回視聴者の割合が P_1 、非視聴者の割合が $N-P_1$ 、それ以外にはなし、という両極端に分かれた分布形をとる。このことはメソリンガム分布の場合は $f(k)$ について極限をとれば、直接確かめることができる。

$$f(k) = {}_N C_k \times \frac{B(l+k, m+N-k)}{B(l, m)} \quad (k \neq 0, N)$$

$$= \frac{N!}{(N-k)!k!} \times \frac{[l(l+1)\Lambda(l+k-1)] \cdot [m(m+1)\Lambda(m+N-k-1)]}{(l+m)(l+m+1)\Lambda(l+m+N-1)}$$

$P_2 \rightarrow P_1$ の時 $l \rightarrow 0$ $m \rightarrow 0$ 但し、 $\frac{l}{l+m} = P_1$ であるから

$$f(k) = \frac{N!}{(N-k)!k!} \times P_1 \times m \times \frac{[(l+1)\Lambda(l+k-1)] \cdot [(m+1)\Lambda(m+N-k-1)]}{(l+m+1)\Lambda(l+m+N-1)}$$

この最後項は $\frac{(k-1)!(N-k-1)!}{(N-1)!}$ に収束する。

従って $f(k)$ は0に収束する。

これに対して $k=0$ 及びNの時は、 $f(k)$ は $N-P_1$ 及び P_1 に収束することが上と全く同様にして確かめられる。

しかし負の二項分布モデルの場合は、元来本数は無限大という仮説をベースにしており、 n , P_1 , P_2 はデータとパラメーターを結びつけるために用いられているだけであるから、メソリンガムのような理論どおりの収束を示さない。これは P_2 の上限が数式そのものの内在的な条件ではなかったことから考えても当然である。 $P_2 \rightarrow P_1$ の場合の負の二項分布は最大値が $k=N-2$ に於いて現れる通常の、特殊でない、負の二項分布にすぎない。

〈 $P_2 < P_1^2$ の場合のFrequency Distribution 推定モデルについての仮説〉

視聴回数の分布は k 時点平均重複視聴率が全ての k について与えられれば定められる。(メソリンガムモデルの変形についてのレポート参照)

メソリンガムモデルは一定の仮説から全ての P_k を与えることによって視聴回数の分布を定めるものであったが、パラメーターの条件から

$$P_1^2 < P_2 < P_1$$

の範囲にその適用を制限されている。

しかし実際には複数媒体又は新聞相互などの場合に上記の制約は、その下限の条件について、必ずしもみたされていない。従ってそのような場合に使用され得るモデル式ないし仮説が必要となる。以下はそのための一つの推論である。

メソリンガムモデル式を重複視聴率の面から把えると

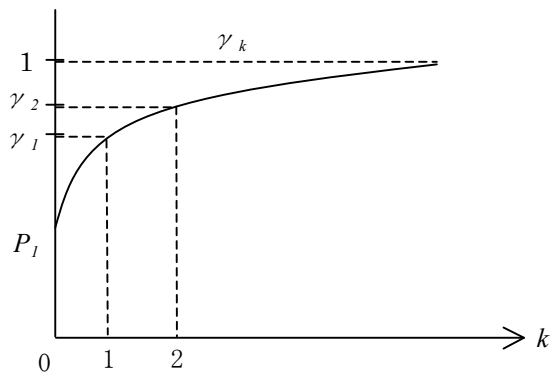
$$\gamma_k = \frac{P_{k+l}}{P_k} = \frac{l+k}{l+m+k} > \frac{l}{l+m} = P_l$$

となっており、 P_{k+l} を P_l^{k+l} と比べると常に

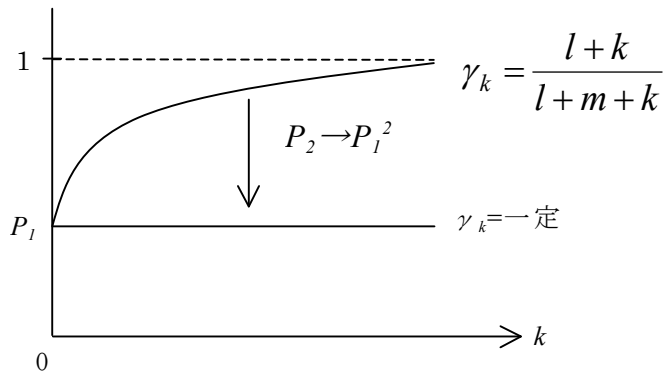
$$P_{k+l} > P_l^{k+l}$$

である。

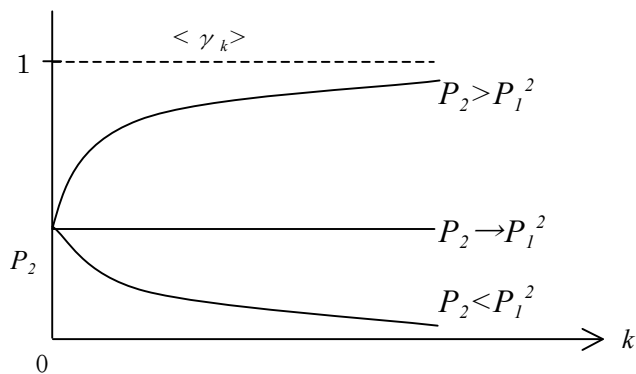
γ_k を k について調べると P_1 を通り1へ収束する分数関数の上に乗せられている。



そして $P_2 \rightarrow P_1^2$ の極限においては P_{k+1} は P_1^k に近づき γ_k については一定値 P_1 に収束する。



このパターンからさらに P_2 が P_1^2 より小になった場合 $\alpha\gamma_k$ の形態を今度は P_1 を通り0に収束する分数関数の上に乗せることを仮定するのは自然な推論であろう。



従って $P_2 < P_1^2$ の時は、

$$\gamma_k = \frac{B}{A+k} \quad , \quad \frac{B}{A} = P_1 \quad (1)$$

$$\gamma_1 = \frac{B}{A+1}$$

$$P_2 = \gamma_1 P_1 = \frac{B}{A+1} \cdot \frac{B}{A} \quad (2)$$

一般に

$$P_k = \frac{B^k}{A(A+1)\Lambda(A+k-1)}$$

となることが想定される。

ここで(1)及び(2)より

$$A = \frac{P_2}{P_1^2 - P_2} \quad B = \frac{P_1 P_2}{P_1^2 - P_2} \quad (3)$$

によってパラメーターA, Bは定められる。

A, Bを(3)で定めて P_k を算出すれば到達率についての一般公式を用いて視聴回数分布は算出可能である。

メソリンガムモデルと負の二項分布モデルの関連

メソリンガムモデルを導く前提となった仮説は

1. 視聴率が P であるグループの大きさ $w_l(P)$ はベータ分布

$$w_l(P) = \frac{1}{B(l, m)} P^{l-1} (1-P)^{m-l}$$

で表される。

2. 一定の視聴率 P であるグループの中での視聴回数の分布は二項分布

$$f_P(k) = {}_N C_k P^k (1-P)^{N-k}$$

で表される。

以上の2点であった。

全体の視聴回数の分布 $F(k)$ は P の各レベルに於ける $f_P(k)$ を足し上げることによって得られる。

$$F(k) = \int_0^1 w_l(P) f_P(k) dp$$

負の二項分布モデルは、上の仮説に加えて

3. 総時点数は無有限大である。

但し、延視聴率 Q_1 及び延二時点間重複視聴率 Q_2 は一定とする。

という仮説を加えることによってメソリンガムモデルからの一つの極限分布として導き出すことができる。

つまり負の二項分布モデルは、総時点数 N を人為的に変数 n と見做しメソリンガムモデルに於いて $n \rightarrow \infty$ とした場合の極限を想定するわけであるが、この場合メソリンガムモデルでデータとして与えられる N , Q_1 , Q_2 のうち N は無視しても、 Q_1 及び Q_2 は保持しておくわけである。

上記の極限分布が負の二項分布に一致することを証明する。

(1) パラメーター l, m と n, Q_1, Q_2 の関係

メソリンガムモデルによると

$$\frac{l}{l+m} = \bar{P}_1 = \frac{Q_1}{n}$$

$$\frac{l}{l+m} \cdot \frac{l+1}{l+m+1} = \bar{P}_2 = \frac{2}{n(n-1)} Q_2$$

この式から

$$l = \frac{Q_1^2 - \frac{Q_1(Q_1 + 2Q_2)}{n}}{2Q_2 - Q_1^2 + \frac{Q_1^2}{n}}$$

$$m = n \cdot \frac{Q_1 - \frac{Q_1^2 + Q_1 + 2Q_2}{n} + \frac{Q_1(Q_1 + 2Q_2)}{n^2}}{2Q_2 - Q_1^2 + \frac{Q_1^2}{n}}$$

が得られる。

この式から

$$l = \frac{Q_1^2}{2Q_2 - Q_1^2} - \varepsilon_1 = \alpha - \varepsilon_1$$

$$m = n \left(\frac{Q_1}{2Q_2 - Q_1^2} - \varepsilon_2 \right) - n(\beta - \varepsilon_2)$$

但し、

$$\alpha = \frac{Q_1^2}{2Q_2 - Q_1^2} \quad \beta = \frac{Q_1}{2Q_2 - Q_1^2}$$

$$\varepsilon_1 = \frac{Q_1^2 - \frac{Q_1(Q_1 + 2Q_2)}{n}}{2Q_2 - Q_1^2 + \frac{Q_1^2}{n}} - \frac{Q_1^2}{2Q_2 - Q_1^2}$$

$$\varepsilon_2 = \frac{Q_1 - \frac{Q_1^2 + Q_1 + 2Q_2}{n} + \frac{Q_1(Q_1 + 2Q_2)}{n^2}}{2Q_2 - Q_1^2 + \frac{Q_1^2}{n}} - \frac{Q_1^2}{2Q_2 - Q_1^2}$$

であり、 ε_1 、 ε_2 は $n \rightarrow \infty$ の時0に収束する。

- (2) $n \rightarrow \infty$ の時、 $nP=x$ を一定とするグループのウエイト $w_2(x)$ は Γ 分布に収束する。

$$\begin{aligned} (\Theta) \quad w_2(x) &= w_1\left(\frac{x}{n}\right) \times \frac{1}{n} \\ &= \frac{1}{B(\alpha - \varepsilon_1, n - (\beta - \varepsilon_2))} \cdot \left(\frac{x}{n}\right)^{\alpha - 1 - \varepsilon_1} \left(1 - \frac{x}{n}\right)^{n(\beta - \varepsilon_2) - 1} \cdot \frac{1}{n} \\ &= \frac{n^{-\alpha + \varepsilon_1}}{B(\alpha - \varepsilon_1, n - (\beta - \varepsilon_2))} \cdot x^{\alpha - 1 - \varepsilon_1} \left\{ \left(1 - \frac{x}{n}\right)^n \right\}^{\beta - \varepsilon_2} - \left(1 - \frac{x}{n}\right) \end{aligned}$$

$n \rightarrow \infty$ の時右辺第2項以下は

$$\begin{aligned} x^{\alpha - 1 - \varepsilon} &\rightarrow x^{\alpha - 1} \\ \left(1 - \frac{x}{n}\right)^n &\rightarrow e^{-x} \end{aligned}$$

従って

$$\left\{ \left(1 - \frac{x}{n} \right)^n \right\}^{\beta - \varepsilon_2} \rightarrow e^{-\beta x}$$
$$1 - \frac{x}{n} \rightarrow 1$$

従って $w_2(x)$ 全体は

$$C \cdot x^{\alpha-1} e^{-\beta x}$$

に収束するが、0から n までの $w_2(x)$ の積分は常に 1 であるから

$$C = \frac{\beta^\alpha}{\Gamma(\alpha)}$$
 である。

(右辺第一項の極限值の厳密な証明は略す)

<注. 一様収束性>

以上によってベータ分布の一つの極限分布として Γ 分布

$$w_2(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

が得られた。

(3) $nP=x$ を一定とするグループの中での視聴回数の分布は、二項分布であるからこれを $n \rightarrow \infty$ とする時は、ポアソン分布に収束することは明らか。

以上によってメソリンガムモデルの二つの仮説において n を Q_1 , Q_2 一定という条件で無限大にもってゆけば、負の二項分布モデルの二つの仮説

- ・GRPが x であるグループの大きさは Γ 分布であらわされる。
- ・一定のGRP・ x であるグループの中での視聴回数の分布はポアソン分布であらわされる。

が導かれることが証明された。

ここでついでに x の導出から得られたパラメーター α 、 β が負の二項分布モデルから直接得られたものと同じであることを確かめておこう。

$$\alpha = \frac{Q_1^2}{2Q_2 - Q_1^2} = \frac{N^2 P_1^2}{2 \times \frac{N(N-1)}{2} P_2 - N^2 P_1^2} = \frac{N^2 P_1^2}{N(P_2 - P_1^2) - P_2}$$

$$\begin{aligned} \frac{\beta}{\beta+1} &= \frac{\frac{Q_1}{2Q_2 - Q_1^2}}{\frac{Q_1}{2Q_2 - Q_1^2} + 1} = \frac{Q_1}{2Q_2 - Q_1^2 + Q_1} \\ &= \frac{NP_1}{2 \cdot \frac{N(N-1)}{2} + P_2 - N^2 P_1^2 + NP_1} \\ &= \frac{NP_1}{N(P_2 - P_1^2) - P_2 + P_1} \end{aligned}$$

これらは負の二項分布モデルのパラメーター m , P と完全に一致している。

(1)

$$Q_k = \sum_{i=k}^N C_k f(i)$$

(2)

$$\begin{aligned} f(k) &= \sum_{i=0}^{N-k} {}_N C_k {}_{N-k} C_i (-1)^i P_{k+i} \\ &= \sum_{i=0}^{N-k} {}_{N-k} C_i (-1)^i Q_{k+i} \end{aligned}$$

(3)

$P_k = \int_0^1 P^k w(p) dp$ をみたす $w(p)$ が存在すれば

$$f(k) = {}_N C_k \int_0^1 P^k (1-P)^{N-k} w(p) dp$$

(4) 上記と同じ条件の下で

$$R(k) = \int_0^1 \int_0^P \frac{x^{k-1} (1-x)^{N-k}}{B(k, N-k+1)} w(p) dx dp$$

(5) $P_1^2 > P_2$ の時は (3) の条件をみたす $w(p)$ は存在しない。

(⊖) $V(P) = P_2 - P_1^2 < 0$ となり矛盾

$$\textcircled{1} \quad P_2 \leq P_1$$

⊖

$$P_{ij} \leq \text{Min}(P_i, P_j) \leq \frac{1}{2}(P_i + P_j)$$

$$\sum_{i < j} P_{ij} \leq \frac{1}{2} \times (N-1) \times \sum P_i = \frac{1}{2} N(N-1) P_1$$

$$\therefore P_2 = \frac{2}{N(N-1)} \sum_{i < j} P_{ij} \leq P_1$$

Q. E. D.

等号が成立するのは $P_i = P_j$ で $P_{ij} = P_i = P_j$ の場合だけである。(N時点の全ての視聴者が同一の場合に他ならない。)

上と同様にして一般に $P_k \leq P_{k-1}$

$$\therefore P_{i_1, i_2 \wedge i_k} \leq \text{Min}(P_{j_1, j_2 \wedge j_{k-1}}) \leq \frac{1}{k} \sum P_{j_1, j_2 \wedge j_{k-1}}$$

$$\sum P_{i_1, i_2 \wedge i_k} \leq \frac{1}{k} \times k \times C$$

②

$$P_2 \cong \frac{N^2 P_1^2}{N(N-1)} \left(1 - \frac{1}{NP_1}\right) = \frac{NP_1^2}{N-1} \left(1 - \frac{1}{NP_1}\right)$$

$$\begin{aligned} V(f) &= \sum k^2 f(k) - \bar{f}^2 \\ &= \sum k(k-1)f(k) + \sum kf(k) - \bar{f}^2 \\ &\quad \quad \quad (= (\sum kf(k))^2) \\ &= 2Q_2 + Q_1 - Q_1^2 \quad (\circledast) \\ &= 2 \times \frac{N(N-1)}{2} P_2 + NP_1 - N^2 P_1^2 \\ &= N(N-1)P_2 - N^2 P_1^2 \left(1 - \frac{1}{NP}\right) \end{aligned}$$

$$V(f) \geq 0$$

故に

$$P_2 \geq \frac{N^2 P_1^2}{N(N-1)} \left(1 - \frac{1}{NP}\right)$$

(\circledast) $Q_k = \sum_{i=k}^N C_k f(i)$ を用いた)

①及び②より

$$\frac{N}{N-1}P_1^2\left(1-\frac{I}{NP_1}\right)\leq P_2 \leq P_1$$

従って $P_2 < P_1^2$ である時には

$$\frac{N}{(N-1)}P_1^2\left(1-\frac{I}{NP_1}\right)\leq P_2 \leq P_1^2$$

従って

$$\begin{aligned} P_1^2 - P_2 &< P_1^2 - \frac{N}{N-1}P_1^2\left(1-\frac{I}{NP_1}\right) \\ &= P_1^2\left(1-\frac{N}{N-1}\left(1-\frac{I}{NP_1}\right)\right) \\ &= P_1^2\left(\frac{I}{(N-1)P_1} - \frac{1}{N-1}\right) \\ &= \frac{I}{N-1}P_1(1-P_1) = \frac{I}{N-1}p \cdot q \quad (q=1-p) \end{aligned}$$

Nが充分大きい時には P_1^2 と P_2 の差額は無視できる。

◎ $P_2 = \frac{N}{N-1} P_1^2 \left(1 - \frac{1}{NP_1} \right)$ である場合について

この時 $f(k)$ の分散は0であるから、全視聴者の視聴回数は等しく NP_1 である。従って全視聴者にまんべんなく視聴が行われ、かつ、 NP_1 が整数になる場合に限られる。

上記と同じ条件の下での P_k について

$$Q_k =_{NP_1} C_k =_N C_k P_k$$

$$\therefore P_k = \frac{NP_1 C_k}{N C_k}$$

これは $k=2$ の場合については

$$P_2 = \frac{NP_1(NP_1-1)}{N(N-1)} = \frac{NP_1^2}{N-1} \left(1 - \frac{1}{NP_1} \right)$$

となり上式と一致する。

これは全ての $P_k(k \geq 2)$ が最小になる場合に他ならない。

この時の

$$\begin{aligned} \frac{P_{k+1}}{P_k} &= \frac{NP_1 C_{k+1}}{NP_1 C_k} \times \frac{N C_k}{N C_{k+1}} = \frac{NP_1 - k}{k+1} \times \frac{k+1}{N-k} \\ &= \frac{NP_1 - k}{N-k} = 1 + \frac{Nq_1}{k-N} \end{aligned}$$

$$P_2^* = \frac{2 \sum_{i < j} P_i P_j}{N(N-1)} \text{ について}$$

$$P_1 = \frac{\sum P_i}{N}$$

$$(NP_1)^2 - N(N-1)P_2^* = \sum P_i^2 \geq 0$$

$$\therefore \boxed{P_2^* \leq \frac{N}{N-1} P_1^2} \quad ; \quad \boxed{P_2^* = \frac{N}{N-1} P_1^2 - \overline{P_i^2}}$$

$$\begin{aligned} V(P_i) &= \frac{1}{N} \left(\sum_{i=1}^N (P_i - P_1)^2 \right) \\ &= \frac{1}{N} \left(\sum P_i^2 - NP_1^2 \right) \\ &= \frac{1}{N} \left\{ (NP_1)^2 - N(N-1)P_2^* - NP_1^2 \right\} \\ &= (N-1)(P_1^2 - P_2^*) \geq 0 \end{aligned}$$

$$\text{もし } \boxed{P_2^* = P_1^2 \text{ なら } V(P_i) = 0}$$

$$\underline{P_{ij} = \alpha P_{ij} + \beta}$$

$$\begin{aligned} P_2 &= \frac{2}{N(N-1)} \sum_{i < j} P_{ij} \\ &= \alpha \left(\frac{2}{N(N-1)} \sum P_i P_j \right) + \beta \\ &= \alpha P_2^* + \beta \end{aligned}$$

$$\begin{aligned} \therefore P_1^2 - P_2 &= P_1^2 - \alpha P_2^* - \beta \\ &= \alpha(P_1^2 - P_2^*) + (1-\alpha)P_1^2 - \beta \\ &= \alpha V(P_i) + (1-\alpha)P_1^2 - \beta \\ &= \alpha(V(P_i) - P_1^2) + P_1^2 - \beta \end{aligned}$$

$$\begin{aligned}\varphi_k(x) &= \frac{{}_x C_k f(x)}{Q_k} & (N \geq x \geq k) \\ &= 0 & (0 \leq x \leq k)\end{aligned}$$

を確率分布とする分布 x の分散から $Q_k \rightarrow (P_k)$ についての不等式を求める。

予備定理 1 .

$$\begin{aligned}x \cdot {}_x C_k &= (k+1) {}_x C_{k+1} + k {}_x C_k \\ \ominus \quad x \cdot {}_x C_k &= (x-k) \frac{x!}{k!(x-k)!} + k \frac{x!}{k!(x-k)!} \\ &= (k+1) \frac{x!}{(k+1)!(x-k-1)!} + k \frac{x!}{k!(x-k)!} \\ &= (k+1) {}_x C_{k+1} + k {}_x C_k\end{aligned}$$

予備定理 2 .

$$x^2 {}_x C_k = (k+1)(k+2) {}_x C_{k+2} + (k+1)(2k+1) {}_x C_{k+1} + k^2 {}_x C_k$$

\therefore 予備定理 1 を 2 回適用すれば得られる。

以上を準備として $\varphi_k(x)$ の平均 E_k の分散 V_k を求める。

$$\begin{aligned}E_k(x) &= \frac{1}{Q_k} \sum_{x=0}^N x {}_x C_k f(x) \\ &= \frac{1}{Q_k} \left\{ \sum (k+1) {}_x C_{k+1} f(x) + \sum k {}_x C_k f(x) \right\} \\ &= \frac{1}{Q_k} \left\{ (k+1) Q_{k+1} + k Q_k \right\}\end{aligned}$$

$$\begin{aligned}
E_k(x^2) &= \frac{1}{Q_k} \sum x^2 {}_x C_k f(x) \\
&= \frac{1}{Q_k} \left\{ (k+1)(k+2)Q_{k+2} + (k+1)(2k+1)Q_{k+1} + k^2 Q_k \right\}
\end{aligned}$$

∴

$$\begin{aligned}
V_k(x) &= E_k(x^2) - E_k(x)^2 \\
&= \frac{1}{Q_k^2} \left\{ (k+1)(k+2)Q_k Q_{k+2} + (k+1)Q_k Q_{k+1} - (k+1)^2 Q_{k+1}^2 \right\}
\end{aligned}$$

Q_k の間に成立する不等式

$$V_k(x) \geq 0$$

故に

$$(k+2)Q_k Q_{k+2} - (k+1)Q_{k+1}^2 + Q_k Q_{k+1} \geq 0$$

∴

$$(k+2)Q_{k+2} Q_k \geq Q_k Q_{k+1} \left((k+1) \frac{Q_{k+1}}{Q_k} - 1 \right)$$

$$(k+2) \frac{Q_{k+2}}{Q_{k+1}} \geq (k+1) \frac{Q_{k+1}}{Q_k} - 1$$

$k=0$ の場合 $\varphi_0(x) = f(x)$ についての不等式

$$2 \frac{Q_2}{Q_1} \geq \frac{Q_1}{Q_0} - 1$$

$$\frac{N(N-1)P_2}{NP_1} \geq \frac{NP_1}{1} - 1$$

$$\therefore N(N-1)P_2 \geq NP_1(NP_1 - 1)$$

$$P_2 \geq \frac{N}{N-1} P_1^2 \left(1 - \frac{1}{NP_1} \right)$$

Q_k についての不等式について等号が成立する場合の Q_k 及び P_k について

$$\begin{cases} Q_k = {}_{NP}C_k \\ P_k = \frac{{}_{NP}C_k}{{}_N C_k} \end{cases}$$

証明 (i) $k=0$ 及び 1 の場合に成立することは明らか。

(ii) $k+1$ 以下の全ての場合について成立するものと仮定すれば

$$\begin{aligned} Q_{k+2} &= \frac{Q_{k+1}}{k+2} \left(\frac{Q_{k+1}(k+1)-1}{Q_k} \right) \\ &= \frac{Q_{k+1}}{k+2} \times \{NP - (k+1)\} \\ &= {}_{NP}C_{k+2} \end{aligned}$$

従って $k+2$ の場合にも成立する。

(i) 及び(ii)によって、上式が証明された。