最尤推定法と中央値の検定

2009 年 3 月 25 日

ここでの目的は、尤度 (likelihood)、尤度比 (likelihood ratio)、最尤推定法 (maximum likelihood estimation)、事後確率、ベイズの定理 (Bayes theorem) などは慣れない場合にはわかりにくい。それらについて解説し、これらの概念から中央値の検定、対数尤度比とカイ 2 乗統計量との関係について数理的な導出を織り込みながら述べる。指数関数と対数関数との関係や展開式も補記する。

1　確率と尤度

サイコロを振った結果は $1, 2, 3, 4, 5, 6$ のそれぞれの目が出る場合に分けられる。これらの結果が表れることを事象とよぶ。同じ確からしさならば、たとえば 3 の目がでる事象の起きる確率は $1/6$ であるという。しかしこのようにある事象の「確率」という場合、通常はその事象はまだ起きていない。3 という目が出てしまうと、それは生起した事柄を指しているものであり、「$1/6$」という数値は確率とよぶ概念とは違ったものになる。もし、サイコロが少しひずんでいるとして、そのひずみのもとで現れる確率を p で表す。そのひずみにより、3 の目の出る確率が $1/6$ から少しずれるとすると、確率は p の関数となる。p のように、各事象の確率に影響を与える因子をパラメータという。一般に確率はパラメータによって定まるから、パラメータの関数となる。科学においては、我々はまずモデルを立て (仮説)、それが実際の観察データに合っているかどうか、想定したものと合致しているかが問題となる。この問題を考察するためにその指標として尤度という概念を導入する。尤度を考える場合、事象は既に起きており、観察データが得られていることから出発する。しかし、それを説明する仮説の正しさが不明である。そこで、あるモデルが正しいとして、その仮説の下での、観察データが起きる確率、あり得そうな数値化を考える。これが尤度である。尤度はしばしば、未知パラメータの推定をするために用いられる。パラメータとは、例えばサイコロのひずみの程度と考えてよい。サイコロのひずみが p であったという仮説の下で「3 の目が出る」確率を考えると、これが、「3 の目が出た」という観測データの下でのサイコロのひずみが p であるという仮説の尤度である。確率も尤度も数式や数値は同じである。「ひずみが p であるときの 3 の出る確率」も「3 が出たという観察データの下でのひずみが p であるという仮説の尤度」も数式は同じである。しかし、確率と尤度は意味するところが異なり、それに係わる法則も違う。確率が「事象の確率」であるのに尤度は「観察データの下での仮説の尤度 (likelihood for a hypothesis given a set of observations)」である点に注意する。

2　尤度比とロッド値

一般にモデルの尤もらしさ (likelihood) を比較する場合、尤度の値そのものの大小にはとくに意味をもたず、異なったモデルの比較のためには「尤度の比」を考察するほうが意味をもつ。つまり、 2 つの仮説 (H_1)

と仮説 (H_0) があるとして、どちらの仮説が尤もらしいか、尤度の比を取ることを考える。いま仮説 H_0 には今から否定したい仮説 (帰無仮説) を取るとし、一方の仮説 H_1 には肯定したい仮説 (対立仮説) が取られる。データを $x = (x_1, x_2, \cdots, x_n)$ が得られたとし、仮説 H の尤度を $L_x(H)$ と表すと、これらの尤度比 $L_x(H_0, H_1) = L_x(H_1)/L_x(H_0)$, (ここで H_0 を基準にするから分母とすることに注意) を取って、H_1 の尤もらしさが H_0 の尤もらしさよりどの程度高いかを調べる。

統計学はさまざまな分野に応用されている。ゲノム情報を実際に利用するためには様々な表現型 (疾患や形質など) がゲノムのどの位置と関係しているのかを調べることは重要である。連鎖不平衡を利用したゲノムワイド関連解析には様々な応用があり、近年特に Pharmacogenomics(PGx) への応用が活発になっています。遺伝統計学を用いた表現型とゲノムとの関係の推定では、疾患感受性遺伝子の同定、薬効・代謝多型の同定など原因の理解、リスク管理、最適な投与量の決定、副作用予測に使われています。また遺伝マーカーによる統計解析では、連鎖不平衡を利用してゲノム上の多型のある部位と表現型との相関を調べます。目的の表現型を持つ個体 (Case 群) と持たない個体 (Control 群) の間でマーカーの頻度を調べ、もし連鎖不均衡があれば、頻度に差が出ることから相関を調べられるという理屈に基づきます。

例えば、連鎖解析では H_0 は連鎖なし、H_1 は連鎖ありという仮説とします。組み換え割合 θ で表すと、$H_0 : \theta = 0.5$、$H_1 : \theta < 0.5$ であり、尤度比 $L_x(H_0, H_1) = L_x(H_1)/L_x(H_0) = L_x(\theta)/L_x(0.5)$ で表されます。尤度比の対数を取ったものが下記のロッド値 (lod score)('lod' は log odds オッズ記録) です。

$$Z_x(\theta) = \log[L_x(\theta)/L_x(0.5)]$$

H_1 も H_0 も同様に尤もらしい場合は尤度比は 1 となり、ロッド値は 0 となります。

3　最尤推定法

遺伝統計学でのパラメトリック連鎖解析で単点 (二点) 分析の場合は、多くの場合パラメータは組み換え割合、θ です。観察データとは、家系と家族員に与えられた遺伝子型と疾患のある無しのデータです。θ にいくつかの値を代入し、観察データ x の下でのそれぞれの θ の尤度を計算します。ある θ の時に最も尤度が最高になるとすると、観察データ x からはその値に依存した $\theta = \theta(x)$ が最も尤もらしいということになります。このようにパラメータを動かし変化させて、尤度が最高になるようなパラメータを捜す方法を最尤推定法 (最尤法、maximum likelihood estimation) という。そのような最大の尤度を最大尤度 (maximum likelihood) といい、それを与えるパラメータの値を最尤推定値 (maximum likelihood estimate; MLE) といい、パラメータを表す変数の上に山形記号 (hat) をつけて $\hat{\theta}$ と表す。このように最も尤もらしいパラメータを推定することができます。$L_x(\theta)/L_x(0.5)$ という尤度比では、分母は θ に関しては定数であり、関係しない。すなわち、尤度比を最大にするパラメータと尤度を最大にするパラメータ、さらにはロッド値を最大にするパラメータは同じものとなります。最尤推定値の求め方は二つ知られていて、一つは解析的に求める方法です。尤度はパラメータの関数なので、これをパラメータで微分し、0 と置いて方程式 (最尤方程式という) を解いてその解が推定値です。もう一つは数値的に求める方法です。Estimation-maximization algorithm (EM アルゴリズム) がしばしば用いられます。最尤推定の局所最適解を E ステップ と M ステップ の二つのステップの繰り返しにより求めるものですが、ここでは省略します。

4　最尤推定値のバイアス (偏り)

このように推定された最尤推定値はバイアス (偏り) がかかっていないのでしょうか。バイアスがかかっていないとは、次のような意味です。例えば組み換え割合 θ により、p_i の確率で d_i というデータが出て、d_i から θ_i という最尤推定値が推定されるとします。このとき、θ(真の値) $= \sum_i p_i\theta_i$(推定値) という等号の関係が成り立つならば、最尤推定値にバイアスがかかっていないという。組み換え割合についてはしばしば最尤推定値にバイアスがかかっていること、等号式が成り立たないことが知られている。それは、n 個の減数分裂の内、$n/2$ を越える組み換えが見られた場合、θ の推定値が $1/2$ となることによる。これにより θ は実際より低く見積もられる事になり、バイアスがかかることになる。しかし、多くの減数分裂の観測を行うことにより、このバイアスは低下する。$n/2$ を越える組み換え割合が推定される確率が減るからである。一般に、連鎖解析による θ の最尤推定値にはバイアスがあるが、その程度は小さく、観察データを増やすことによりバイアスは低下する。

5　検定と危険率

連鎖解析でパラメーターである θ を推定する場合、連鎖が無いという仮説を否定する目的である場合が多い。すなわち、$\theta = 1/2$ という仮説を否定することを意図する。このように、ある仮説 (連鎖している) を証明したい場合、それと反対の仮説 (連鎖していない) を検定により否定するという立場を取る。後者の仮説を帰無仮説とよぶ。ほとんどの場合、帰無仮説は 100% 否定することはできない。ここで、観察データから計算できる統計量を考える。もし、帰無仮説が正しければ、どのような観察データがどのような確率で生み出されるかはわかるので、その統計量の確率分布もわかる。例えば、95% の確率で、平均値というテスト統計量が 67.3 以下になるというようにわかる。即ち、何回も帰無仮説のもとで生み出された観察データの統計量のほとんど (これを、割合 $1-\alpha$ としよう) がある値以下となる、というような閾値を設ける。そして、その閾値を越えるデータが観測されたとき、帰無仮説を否定する。そして、その危険率を α とする。本当は帰無仮説が正しいのにそれが否定される事 (偽陽性) をタイプ I のエラー (過誤) と呼び、その確率をタイプ I のエラー比率と呼ぶ。帰無仮説が否定された場合、有意であり (significant) といい、否定できない場合、有意でない (non-significant) という。即ち、タイプ I のエラー比率とは、帰無仮説が正しい場合に、significant と結論づける比率である。医学では疾患の診断法などについて、「感度 (sensitivity)」「特異度 (specificity)」の概念がしばしば用いられる。疾患である (罹患) ことを仮説 H_1、疾患で無い (非罹患) ことを H_0 とすると、感度は疾患であるとき (H_1) 疾患と判定される確率なので検定力 ($1-\beta$)、特異度は疾患でないとき (H_0) 疾患で無いと判定される確率なので $1-\alpha$ と考えられる。一般に、テスト統計量の閾値は α の値により人為的に調節できる。α を小さくすると有意差が出にくくなり、大きくすると出やすくなる。例えば $\alpha = 0.05$ などと固定し、テストを行う。このようにテストすると、本来、有意があるのにテストで有意でない、と出る場合もある (偽陰性)。これをタイプ II のエラーとよび、その比率をタイプ II のエラー比率とよび、β で表す。$1-\beta$ を検定力 (パワー) とよぶ。すなわち、パワーとは対立仮説 H_1 が正しいとき、有意となる確率である。有意差なしという結果がでると、帰無仮説が正しいことが証明されたと考えるのは誤りである。単にパワーが低いためである可能性もある。連鎖解析の場合、しばしば帰無仮説 H_0 は $\theta = 0.5$ であり、H_1 は $\theta < 0.5$ である。

6　事後確率とベイズの定理

一般に確率は事象が起きる前に予測するためのものであるが、尤度はすでに起きた事象の観察データに基づいて、モデル (仮説) の尤もらしさを考えるためのものであった。事後確率は、ある観察データのもとでのモデルの確率を計算するものである。ベイズの定理は非常に有用な定理であるが、感覚的にわかりにくいところがある。ここで、RA テストという検査の例を取って、ベイズの定理を説明する。慢性関節リウマチという病気は一般に 1% 位の頻度で存在する。慢性関節リウマチ患者に RA テストを行うと 80% が陽性である。慢性関節リウマチでない人に RA テストを行うと 3% が陽性となる。検査というのは当然、その結果によって疾患の可能性が変わるから行うのであるが、RA テストが陽性という結果がでた事により、慢性関節リウマチの可能性はどの程度高くなるのであろうか。ここで、慢性関節リウマチ (RA) であるという事象を F、RA テストが陽性という事象を E で表す RA でないという事象を F^c で表すとする。上の 80%、3% という数値は、次のような概念に相当する。$P(E|F) = 80\%, P(E|F^c) = 3\%$ ここで、$P(E|F)$ は F という条件の下で E という事象の確率、すなわち条件付き確率を示す。また、慢性関節リウマチの頻度が 1% であり、今テストを行われる人につて全く情報が無いとすると、$P(F) = 1\%$ ここで、$P(F|E)$ という式について考えてみる。これは RA テストが陽性の時に、それを行われた人が RA である確率である。これは次のような順序で考えるとわかりやすい。

(1) 世の中には RA である人と RA でない人がいるが、その割合は $P(F)$、$P(F^c) = 1 - P(F)$ である。
(2) RA であって、RA テスト陽性の人の全体での割合は、$P(F)P(E|F)$。
(3) RA ではなく、RA テスト陽性の人の全体での割合は、$P(F^c)P(E|F^c)$。
(4) RA テスト陽性の人の全体での割合は (2)+(3)
(5) その中で、RA の人の全体での割合は (2)。
(6) 従って、RA テスト陽性の人の中で RA の人の割合は (2)/[(2)+(3)] 即ち、

$$P(F|E) = P(F)P(E|F)/[P(F)P(E|F) + P(F^c)P(E|F^c)]$$

最初に何の情報もなく、RA テスト陽性という結果だけがわかっているとすると、この人が RA である確率は、

$$P(F|E) = 0.01 \times 0.8/(0.01 \times 0.8 + 0.99 \times 0.03) = 0.212$$

約 21% である。$P(F)$、即ち、RA テストを行う前の RA の確率のようなものを事前確率と呼ぶ。それに対して、$P(F|E)$、つまり RA テストをやったら陽性がでたときに、RA の確率を考えるような場合を事後確率と呼ぶ。RA テストという観察データにより RA である確率が変化し、上昇した。即ち、事前確率 1% から事後確率 21% になった。

以上の例では、RA である、RA でない、の二つの事象を考えたが、もっと多くの事象にわかれる場合、n 個の事象 $E_k, k = 1, 2, \cdots, n$ においてベイズの定理は次のように表される。

$$P(E_k|F) = P(F|E_k)P(E_k)/\sum_j P(F|E_j)P(E_j), \quad k = 1, 2, \cdots, n$$

分母は定数だから、C を定数として表現すると、

$$P(E_k|F) = C \cdot P(F|E_k)P(E_k).$$

ここで $1/C = \sum_j P(F|E_j)P(E_j)$, E_j は $j = 1, 2, ..., n$ の異なった事象である。上の例では、RA の診断という概念を用いたので、事後確率という概念が比較的わかりやすく理解できたと思う。それは診断基準などの概念があり、誤診率などの概念も存在して、一般に RA である確率という概念が理解しやすいためである。しかし、RA か無いかは既に決まっていると考えると、RA の確率という概念は理解しにくいとも言える。一般に、事後確率の概念は非常に理解しにくいことが多い。研究者によっては事後確率の概念を用いることを嫌う人もいる。尤度比という概念を用いれば、事後確率の概念は必要無いと考える研究者もいる。事後確率の概念が嫌われる一つの理由は、事前確率が多くの場合、不明だからである。前述のように、尤度に事前確率をいれると、すぐに事後確率が計算できる。事前確率があいまいなままでも仮説の尤もらしさを比べることができるのが尤度比なのである。

例えば、前述の場合には事前に全く、その人の情報が無いと仮定した。従って、RA テストが陽性になっても RA の可能性は 21% なのである。しかし、一般には RA を疑うから RA テストを行うのである。その場合は、21% より更に高い確率となるであろう。なぜなら、事前確率、$P(F)$ が 1% ではないからである。この場合には、事前確率は用いずに、RA である場合と、RA で無い場合の二つに分けて、それぞれの場合の尤を計算し、その比を取るだけで表した方が良いと考える人も多いであろう。ベイズの定理は不完全データから完全データを予測するためのアルゴリズムにしばしば用いられる。例えば、EM アルゴリズムでは不完全データの尤度を計算し、それに基づいてベイズの法則でパラメータの期待値を計算する。Genehunter では観察データに基づいて、継承ベクトルの尤度を計算し、その継承ベクトルの尤度に基づいてθ の期待値をベイズの定理により計算する。そして、そのθ を用いて観察データに基づいて再び継承ベクトルの尤度を計算する。そのようにして、尤度が収束するまで繰り返す。

7　中央値の検定

尤度比検定は、よく使われる便利な検定の枠組みである。中央値検定 (メディアン検定) といえば、ノンパラメトリック検定の中でも、依存している仮定の少ない検定である。両者の関係は薄いようにも見えるかもしれないが、ノンパラメトリックな検定も何らかの確率モデルは想定している (特定の分布を想定しないだけ)。中央値検定は、尤度比検定の例であるとも考えることができる。

母中央値を境にして、大と小の 2 つに分けると、等しい確率 $(1/2)$ で得られて、しかも合計は (全体だから) 1 である。つまり、大も小もそれぞれ 0.5 の確率でおこる。ある母中央値を持つ母集団からの、サンプルサイズ n の標本で、大が x 個、小が $(n-x)$ 個となる確率は、${}_nC_x(0.5)^x(0.5)^{(}n-x)$、つまり、$\frac{n!}{x!(n-x)!}(0.5)^n$ である。

1 サンプル (サンプルサイズ n は偶数としておく) の、データを小さい方から大きい方に $y_1, y_2, \cdots$ と順にならべる。母中央値が y_j と y_{j+1} の間にあるとすると (小が j 個)、尤度は、すぐ上の項から、${}_nC_j(0.5)^n$ である。${}_nC_j$ と ${}_nC_{j+1}$ の比は $(j+1)/(n-j)$ だから、$j = n/2$ つまりサンプルのデータを二分するところに母中央値があるときに最大となり (n が奇数ときには、$(n-1)/2$ と $(n+1)/2$ の尤度は等しい)、ここに母中央値があると推定するのが最尤推定となる。

対立仮説のモデル:母中央値がちがう (対立仮説に対応) モデルでは、上記の 1 サンプルの場合をそれぞれのサンプルについて行なったものが最大尤度を与えるから、サンプルサイズを n_1, n_2 (いずれも偶数とする) とすると、最大尤度は、

$$ {}_{n_1}C_{n_1/2}(0.5)^{n_1} \times {}_{n_2}C_{n_2/2}(0.5)^{n_2} $$

となる。母中央値はちがっていいので、パラメーターは 2 つである。

母中央値が同じ (帰無仮説に対応) するモデルでは、両方を一緒にした (n_1+n_2) 個を大きい半分と小さい半分に分けることになる。第 1 のサンプルは、小さい方に x_1 個、大きい方に (n_1-x_1) 個第 2 のサンプルは、小さい方に x_2 個、大きい方に (n_2-x_2) 個と分かれたとする ($x_1+x_2=(n_1+n_2)/2$ である)。母中央値は同じなので、パラメータは 1 つである。第 1 のサンプルについてのこのモデルの最大尤度は

$$ {}_{n_1}C_{x_1}(\ 0.5\ )^{n_1} $$

第 2 のサンプルについては

$$ {}_{n_2}C_{x_2}(\ 0.5\ )^{n_2} $$

となる。最大対数尤度の差は、

$$ \begin{aligned} \log L_x(H_0,H_1) &= \log\frac{L_x(H_1)}{L_x(H_0)} = \log L_x(H_1)-\log L_x(H_0) \\ &= \log\{{}_{n_1}C_{n_1/2}\}+\log\{{}_{n_2}C_{n_2/2}\}-\log\{{}_{n_1}C_{x_1}\}-\log\{{}_{n_2}C_{x_2}\} \end{aligned} $$

で、整理すると、

$$ \log\{x_1!\}+\log\{(n_1-x_1)!\}\ +\ \log\{x_2!\}+\log\{(n_2-x_2)!\}-2\log\{(n_1/2)!\}-2\log\{(n_2/2)!\} $$

となる。スターリングの公式 $n!\sim\sqrt{2\pi n}n^ne^{-n}$, $\log n!\sim n\log n-n+\frac{1}{2}\log n+\log\sqrt{2\pi}$ で近似すると、簡単な式変形をおこなうことで

$$ \begin{aligned} & \log L_x(H_0,H_1) \\ = \ & x_1\log x_1-x_1+(n_1-x_1)\log(\ n_1-x_1\ )-(n_1-x_1) \\ & +x_2\log x_2-x_2+(n_2-x_2)\log(\ n_2-x_2\ )-(n_2-x_2) \\ & -n_1\log(\ n_1/2\ )+n_1-n_2\log(\ n_2/2\ )+n_2 \\ = \ & x_1\log x_1+(n_1-x_1)\log(n_1-x_1)+x_2\log x_2+(n_2-x_2)\log(n_2-x_2) \\ & -n_1\log(n_1/2)-n_2\log(n_2/2) \\ = \ & x_1\log\frac{x_1}{n_1/2}+(n_1-x_1)\log\frac{n_1-x_1}{n_1/2}+\ x_2\log\frac{x_2}{n_2/2}+\ (n_2-x_2)\log\frac{n_2-x_2}{n_2/2} \end{aligned} $$

となる。この 2 倍が対数尤度比統計量で、この場合、(パラメーター数の差は $2-1=1$ なので) 自由度 1 のカイ 2 乗分布と比べることになる。

観測値として $o_{11}=x_1, o_{12}=x_2, o_{21}=n_1-x_1, o_{22}=n_2-x_2$ さらに期待値として $e_{11}=n_1/2, e_{12}=n_1/2, e_{21}=n_2/2, o_{22}=n_2/2$ とおくと、したがって対数尤度比統計量は、

$$ \begin{aligned} 2\log L(H_0,H_1) &= 2\times\ (\text{実測値})\log(\text{実測値}/\text{帰無仮説のもとでの期待値})\text{の合計} \\ &= 2\textstyle\sum_{i,j}o_{ij}\log\dfrac{o_{ij}}{e_{ij}} \end{aligned} $$

という形をしている。上記の式で計算した実測値と帰無仮説での期待値とで定まる量、この分布間との距離を Kullback-Leibler ダイバージェンスといい、最尤推定は , これで測ったとき , 真の分布に最も近いモデルの分布を見つけているといえる。

中央値検定では 2 × 2 分割表を作って検定するが、その際には、Fisher の検定やいわゆるカイ 2 乗検定、G 検定などが使われる。上記の対数尤度比統計量は G 検定するときの G 統計量と同じである (G 検定は、尤度比に基づくものだから、意外性は薄いが)。

8　Pearson のカイ 2 乗統計量

分布の適合度を測る尺度として、よく用いられる Pearson のカイ 2 乗統計量を述べよう。いま k 種類の事象が、それぞれ確率 $p=(p_1,p_2,\cdots,p_k), p_1+p_2+\cdots+p_k=1$ で起こる。n 回のうち、これらの事象が起こった回数を x_i とする。すなわちパラメータ (n,p) の多項分布とする。$x=(x_1,x_2,\cdots,x_k)$ に対して、密度関数は

$$P(x_1,x_2,\cdots,x_k|p)=\frac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k},\quad x_i\in\{0,1,2,\cdots,n\},\sum_{i=1}^{k}x_i=n$$

で与えられる。パラメータの集合を Θ_0 として 2 つの仮説

$$\begin{array}{l}H_0:p_i=p_i(\theta),\ \theta\in\Theta_0\\ H_1:p_i(\text{任意})\end{array}$$

を考えると、標本比率 (割合) $\hat{p}_i=\dfrac{x_i}{n}$ と H_0 のもとでの θ の最尤推定量を $\hat{\theta}$ とおき、

$$\begin{array}{l}H_0:p_i(\hat{\theta})\\ H_1:\hat{p}_i=\dfrac{x_i}{n}\end{array}$$

に対する対数比尤度を計算する。

$$\begin{aligned}\log L_x(H_0,H_1) &= \log L_x(H_1)-\log L_x(H_0)=\textstyle\sum_i x_i\log\hat{p}_i-\sum_i x_i\log p_i(\hat{\theta})\\ &= \textstyle\sum_i x_i\log\dfrac{\hat{p}_i}{p_i(\hat{\theta})}=\sum_i x_i\log\dfrac{x_i/n}{p_i(\hat{\theta})}=\sum_i o_i\log\dfrac{o_i}{e_i}\end{aligned}$$

ここで $o_i=x_i$: 観測度数 (observed)、$e_i=np_i(\hat{\theta})$: 期待度数 (expected) とした。

ここで対数関数の展開式

$$\log(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\cdots,\qquad |x|<1$$

を用い、さらに係数 2 をつけて、変形していく。観測度数 o_i と期待度数 e_i の差を $\delta_i=o_i-e_i$ とおくと、$\sum_i\delta_i=\sum_i(o_i-e_i)=n-n=0$ であるから、

$$\begin{aligned}2\log L_x(H_0,H_1) &= 2\textstyle\sum_i o_i\log\dfrac{o_i}{e_i}=2\sum_i(\delta_i+e_i)\log\dfrac{\delta_i+e_i}{e_i}=2\sum_i(\delta_i+e_i)\log\left(1+\dfrac{\delta_i}{e_i}\right)\\ &= 2\textstyle\sum_i(\delta_i+e_i)\left(\dfrac{\delta_i}{e_i}-\dfrac{1}{2}\left(\dfrac{\delta_i}{e_i}\right)^2+\dfrac{1}{3}\left(\dfrac{\delta_i}{e_i}\right)^3-\cdots\right)\\ &= 2\textstyle\sum_i\left(\dfrac{\delta^2}{e_i}-\dfrac{\delta^3}{2e_i^2}+\dfrac{\delta^4}{3e_i^3}-\cdots\quad+\delta_i-\dfrac{\delta^2}{2e_i}+\dfrac{\delta^3}{3e_i^2}-\cdots\right)\\ &= 2\textstyle\sum_i\left(\delta_i+\dfrac{\delta^2}{2e_i}-\dfrac{\delta^3}{6e_i^2}+\cdots\right)\\ &= \textstyle\sum_i\left(\dfrac{\delta^2}{e_i}-\dfrac{\delta^3}{3e_i^2}+\cdots\right)\\ &\fallingdotseq \textstyle\sum_i\dfrac{\delta^2}{e_i}=\sum_i\dfrac{(o_i-e_i)^2}{e_i}=\sum_i\dfrac{o_i^2}{e_i}-n\end{aligned}$$

補記

(1) 一般二項展開式について一般の実数 s に対し、変数 x における関係式

$$(1+x)^s = 1 + \binom{s}{1}x + \binom{s}{2}x^2 + \binom{s}{3}x^3 + \cdots = \sum_{k=0}^{\infty}\binom{s}{k}x^k$$

を一般二項展開とよぶ。ここで二項係数 $\binom{s}{k} = \dfrac{(s)_k}{k!} = \dfrac{\overbrace{s(s-1)\cdots(s-k+1)}^{k}}{k!}$

問１．もし $s = n$(正の整数) のとき、$\binom{s}{k} = \binom{n}{k} = \begin{cases} {}_nC_k & k = 0,1,2,\cdots,n \\ 0 & k = n+1, n+2, \cdots \end{cases}$

問２．$\dfrac{1}{1+x} = (1+x))^{-1} = 1 - x + x^2 - x^3 + \cdots$

つまり $\binom{-1}{k} = (-1)^k = \begin{cases} +1 & k = 0,2,4,\cdots(even) \\ -1 & k = 1,3,5,\cdots(odd) \end{cases}$

(2) 対数の展開式について

1. 不定形極限 (オイラー) 1^∞ の形: $\lim_{s\to 0}(1+s)^{1/s} = e = 2.71828\cdots$
2. 「指数関数の展開式」実数 x について $\lim_{t\to\infty}\left(1+\dfrac{x}{t}\right)^t = e^x = \exp(x) = \sum_k \dfrac{x^k}{k!}$ が成立。

   なぜなら、$\left(1+\dfrac{x}{t}\right)^t = \sum_k \binom{t}{k}\left(\dfrac{x}{t}\right)^k = \sum_k (1-1/t)(1-2/t)\cdots(1-(k-1)/t)\dfrac{x^k}{k!}$ において、k を固定して、$t\to\infty$ とすると、$(1-1/t)(1-2/t)\cdots(1-(k-1)/t) \to 1$ より、

$$\begin{aligned} e^x &= \lim_{t\to\infty}\left(1+\frac{x}{t}\right)^t = \lim_t \sum_k \binom{t}{k}\left(\frac{x}{t}\right)^k \\ &= \sum_k \lim_t \left\{(1-1/t)(1-2/t)\cdots(1-(k-1)/t)\right\}\frac{x^k}{k!} \\ &= \sum_k \frac{x^k}{k!} \end{aligned}$$

   とくに $x = 1$ ならば、e はネピア数あるいはオイラー数とよばれる超越数で、

$$e = 2.7182818284\cdots = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

3. 「対数関数の展開式」任意の実数 s について、$a^s = 1 + sb$ という a, b の関係を調べる。いま二項係数の展開式を用いるために、$a = 1+x$, $b = y$ とおく。極限を $s\to 0$ とすれば、

$$\begin{aligned} y &= \frac{(1+x)^s - 1}{s} = \sum_{k=1}^{\infty}\binom{s}{k}\frac{x^k}{s} \\ &= \sum_{k=1}^{\infty}\frac{\overbrace{(s-1)(s-2)\cdots(s-(k-1))}^{k-1}}{k!}x^k \\ &\longrightarrow \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}x^k = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots \end{aligned}$$

   ここで $\lim_{s\to 0} y = \lim_{s\to 0}\dfrac{(1+x)^s - 1}{s} = \log(1+x)$ より、

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots, \quad |x| < 1$$

   が得られる。

問

$e = \lim_{t \to 0}(1+t)^{1/t}$ と $\log(1+x) = \lim_{s \to 0} \dfrac{(1+x)^s - 1}{s}, \forall x$ とは同値である。