

【第1章】確率・統計の基礎

1.1.3 最尤推定

やっと本命の最尤推定という言葉が出てきました。お待たせしました。この節はいままでの中で最も長く、少し難しい内容も出てきます。がんばってください。これが終わるといよいよ本命のMLEM、OSEMの章です。

ところで“尤”なる字はあまり見かけませんね。“ゆう”と読みます。“いぬ”ではありません！！この意味は「もっともらしい」という意味合いで、ある理論により得られた数字が、最も確からしい推定によって得られる理論を最尤推定法と呼んでいます。なんのことやら？ですが、皆さんよく使うもので、あるデータをグラフ上にプロットして“相関直線”を引いて、Rが幾つだから有意だ！とかの結果を良く見ますよね。この直線は最小二乗法という方法で求めるのですが、これも最尤推定法の1つです。最も近似できる直線近似式で、結果を予想することを行っているからです。最初にも書きましたが、我々が普通に計算している“平均値”も最尤推定です。

もう1つ、「ニュートン法」という概念を簡単に解説しましょう。これは逐次近似の概念です。関数電卓で、解を得る方法はこのニュートン法を使っています。

(ニュートン法)

これは単純な計算を繰返して行い、解を近似値として求める方法です。MLEMと同じ概念ですね！！高校の数学で、ある曲線の接線の方程式ってのがありました。これを使います。思い出してください。解説は省略します。

最初にある関数 $f(x)$ は微分可能であるとします。方程式

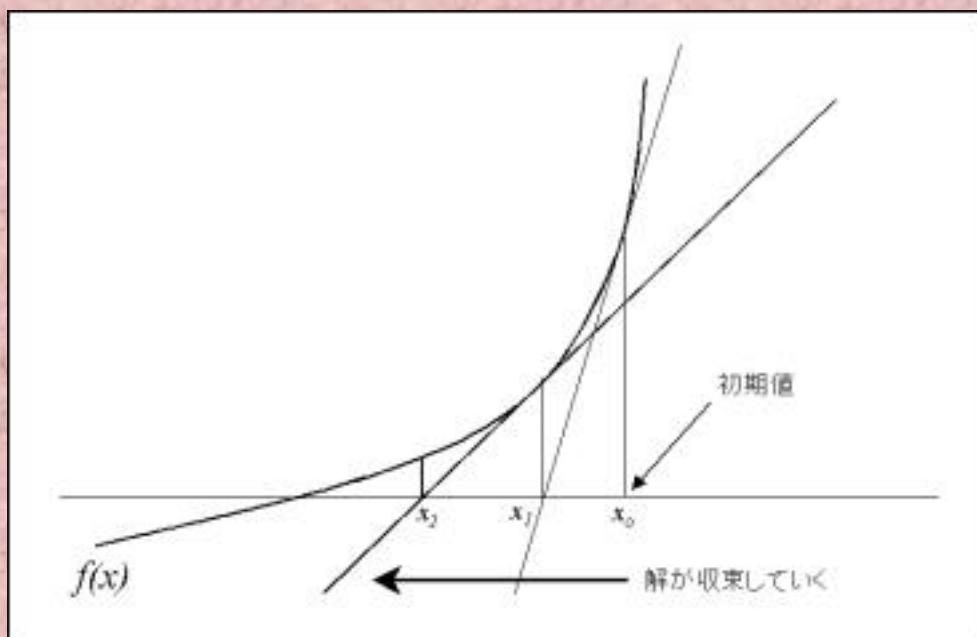
$$f(x)=0$$

を解くニュートン法は、

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

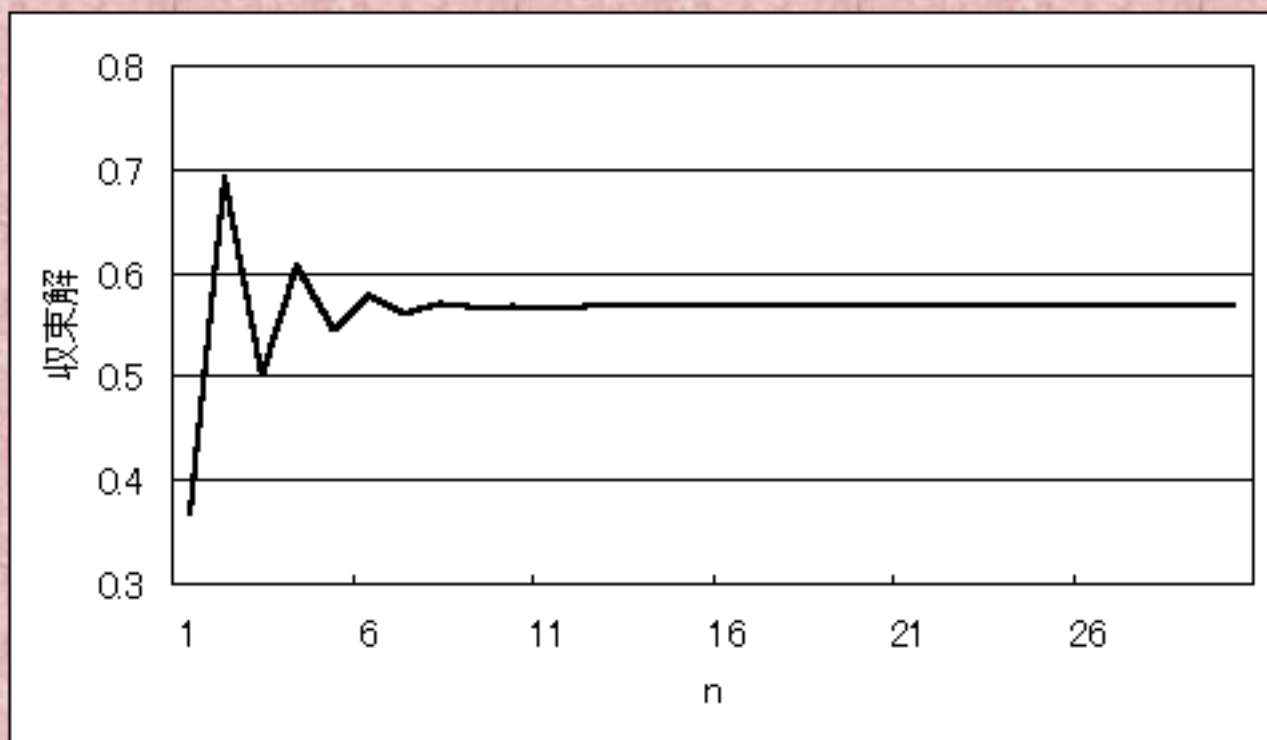
と書けます。ここで $n+1$ とか n は、 i 番目の計算回数を示します。最初、つまり初期値を x_0 としましょう。式の意味を考えますと、上式の右辺は点 $(x_n, f(x_n))$ における $f(x)$ の接線と x 軸との交点の x 座標を表しています。最初に x_0 を右辺に代入し、計算結果を x_1 とします。 x_1 をまた右辺に代入して計算し、結果を x_2 とします。同様に計算を繰返すと次第に解へと近づいてい

くのです。図に示してみましょう。



ニュートン法による解の収束の様子

下のグラフは式； “ $x-2e^{-x}=0$ ” 計算結果の収束の様子を示したものです。横軸はくり返し演算の回数で、初期値は $x_0=1$ としました。縦軸は x_{n+1} の値であり、 x_1 から示してあります。



$x-2e^{-x}=0$ のニュートン法による解の収束

このように、繰り返し演算を行い、近似解を求める方法を逐次近似法といいます。

さて、最尤推定に話を戻しましょう。最初にいくつかの定義を書いておきます。

【定義 1】

標本 $X = (X_1, X_2, \dots, X_n)$ の分布がパラメトリックモデル、 $P = \{f(x; \theta) : \theta \in \Theta \subset R^k\}$ に属するとする。 $f(x; \theta)$ は確率関数または確率密度関数とする。与えられた観測値 $x = (x_1, x_2, \dots, x_n)$ について $f(x; \theta)$ を、 θ の関数とみなしたものを尤度関数と呼び、とくに $L(\theta; x)$ と表す。つまり、

$$L(\theta; x) = f(x; \theta)$$

とくに $x = (x_1, x_2, \dots, x_n)$ がランダム標本の場合はその尤度関数は、

$$L(\theta; x) = \prod_{i=1}^n f(x_i; \theta)$$

で与えられる。 $f_i(x; \theta)$ は母集団の確率関数または確率密度関数である。

【定義 2】

尤度関数を最大にする値、 $\hat{\theta} = \hat{\theta}(x)$ を最尤推定値といい、

$$L(\hat{\theta}; x) = \sup\{L(\theta; x); \theta \in \Theta\} = \sup\{f_i(x; \theta); \theta \in \Theta\}$$

$\hat{\theta} = \hat{\theta}(X)$ を θ の最尤推定量と呼ぶ。

【定義 3】

尤度関数の対数をとったものを対数尤度関数といい、

$$\log\{L(\theta; x)\}$$

と表す。また、 $\Theta = (\theta_1, \dots, \theta_k)$ が R^k の開集合で、それぞれの θ_i についての 1 次偏導関数が

存在するとき、最尤推定値 $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$ は以下の方程式を満たす。

$$\frac{\partial}{\partial \theta_i} \log L(\hat{\theta}; x) = 0 \quad i=1, \dots, k$$

これを尤度方程式という。

さて定義1から考えてみましょう。変な記号が出てきますが、要は独立した事象を複数個考えた場合、その尤度関数というのは、ここの確率（密度）関数をかけたものであるといっています。サイコロを2回投げて、2の目が続けて出るとは、確率は $1/6 \times 1/6$ のように、各々の確率を掛けているのです。そのために記号として \prod を使っているのです。

定義2はなんでしょうか？ これは定義1でのべた尤度関数を最大にする値を求めることを最尤推定と呼びますといっています。“Sup”とは“最大の”という意味です。ちなみに「最小の」は「Inf」と書きます。^の記号は、母集団に対して標本集団のパラメータに使用する記号で、区別の為に用いているにすぎません。

定義3は、ただ単に尤度関数の対数を取っただけなのですが、何のためかといいますが、これにより、記号 \prod が \sum に置き換えることができるのです。なんのため？ あとでやりますが、計算を簡単にするためのテクニックなのです。そして後半は、関数を1階微分したものが0になることを示しています。1階微分するとは、極値を求めることになります。例えば上に凸の2次関数なら、曲線が一番上のところに来るところを示しています。高校の数学では、関数のグラフを書く時に極小点や極大点また変曲点を探すために微分を使いました。記号「 ∂ 」は微分を表す記号で、「ディー」とか「デル」とか「ラウンド」と読みます。多くの場合、「ディー」と読みます。余談ですが、私が大学で習ったある中国人の数学の講師は、「ロンダ？」と発音していました。この先生は大学も大学院も東大卒なので日本語はペラペラなのですが、「ロンダ」とは何語かわかりませんでした？

これらのことから、平均値が本当に最尤推定値、つまり最も確からしい値は平均値であること、正規分布とポアソン分布の2つの分布について証明しましょう。少し難しいかもしれませんが、いままでのことを用いればできるはずですよ。

(正規分布の場合)

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

さて、尤度関数はどのように書けるでしょうか。以下のように書けます。

$$L(\mu, \sigma^2; \mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x_i - \mu)^2\right]$$

次に対数尤度関数を考えましょう。これは計算をしやすくするためでした。

$$\text{Log}L(\mu, \sigma^2; \mathbf{x}) = \text{Log}\left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x_i - \mu)^2\right]\right)$$

さて、思い出しましょう。Logの計算です。Log $ab = \text{Log} a + \text{Log} b$ でしたね。つまりLogの中での掛け算は、バラバラにして足算に直せるのです。また割算は引き算にできました。すると

次のように変形できます。

$$\begin{aligned}
 & \log \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (x_i - \mu)^2 \right] \right) \\
 &= \log \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} + \log \prod_{i=1}^n \exp \left[-\frac{1}{2\sigma^2} (x_i - \mu)^2 \right] \\
 &= \log \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} + \log \prod_{i=1}^n e^{\left[\frac{1}{2\sigma^2} (x_i - \mu)^2 \right]} \\
 &= \log \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} + \log \prod_{i=1}^n e^{\left[\frac{1}{2\sigma^2} \right]} e^{\left[(x_i - \mu)^2 \right]} \\
 &= -\frac{1}{2} \log \prod_{i=1}^n (2\pi\sigma^2) - \frac{1}{2\sigma^2} \log \prod_{i=1}^n e^{\left[(x_i - \mu)^2 \right]} \\
 &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2
 \end{aligned}$$

次にこの尤度方程式はどうなるでしょう。微分したものが0でしたね。パラメータを平均値としましたので、 μ で微分します。すると μ に対して右辺第一項と第二項は定数ですから微分すると0になります。よって尤度方程式からなくなります。従って第三項のみを考えます。すると、

$$\frac{\partial}{\partial \mu} \log L(\hat{\mu}, \hat{\sigma}^2) = \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n (x_i - \hat{\mu}) = 0$$

つまりこの式が成立するのは、 $\hat{\mu} = \bar{x}$ のとき、つまり標本集団の平均値のパラメータは母集団の平均値に一致することになりますので、最尤推定値は平均値に一致しました。分散に関しては同様に σ^2 で微分すると求まります。最終の式だけ書くと次のようになります。確かめてみてください。

$$\begin{aligned}
 \frac{\partial}{\partial \sigma^2} \log L(\hat{\mu}, \hat{\sigma}^2) &= -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (x_i - \hat{\mu})^2 = 0 \\
 \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2
 \end{aligned}$$

ここで微分は、偏微分となるのですが、細かいことは省略しまして、微分としていますのでご了承ください。

(ポアソン分布の場合)

$$f_X(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

同様に考えてみましょう。こちらはパラメータが1つ()しかないのです、少し楽です。さて手順は、尤度関数を作り、その対数をとった対数尤度関数を作成し、 を に変更し整理します。それをパラメータで微分したものが0となる尤度方程式を考えてあげれば最尤推定値を求められます。ではやってみましょう。

$$L(\lambda; \mathbf{x}) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

$$\begin{aligned} \log L(\lambda; \mathbf{x}) &= \log \left[\prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \right] \\ &= \log \prod_{i=1}^n [e^{-\lambda} \lambda^{x_i} (x_i!)^{-1}] \\ &= \sum_{i=1}^n (-\lambda + x_i \log \lambda - \log x_i!) \\ &= -n\lambda + \log \lambda \sum_{i=1}^n x_i - \sum_{i=1}^n \log x_i! \end{aligned}$$

で微分します。

$$\frac{\partial}{\partial \lambda} \log L(\lambda; \mathbf{x}) = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i = 0$$

$$\therefore \lambda = \frac{1}{n} \sum_{i=1}^n x_i$$

最後の式は、 n 個のデータを全部足して n で割っている、つまり平均を計算しています。つまり平均を求めることがポアソン分布のパラメータである平均と一致しました。

MLEMでよく言われる最尤推定によって何が求められるのか・・・これはつまり平均値を求めることが、最もそれらしい数値ですよ！ ってことを、理論的に示したのです。ですから胸を張って平均値が、最も信用がおける数値として採用できるのです。平均値の信頼性は、データが多いほうがその精度が高くなることは経験的にもわかりますし、統計学的にもそのとおりです。



これで第1章は終わりにします。次は臨床に使われているMLEMの概念に入ります。よく見れば簡単なことをやっているのに気がつくと思います。演算は「足す・掛ける・割る」の3つを繰返しているだけです。本当です。引き算は行いません！！

最後に、私的な都合により、
発行が遅れましたこと深くお詫びいたします。
これからもよろしく願いいたします。

