尤度関数

情報ソース：Wikipedia, 統計学実践ワークブック

【尤度】

データのもっともらしさ。

【尤度関数】

ある前提条件に従って結果が出現する場合に、逆に**観察結果からみて前提条件が「何々であった」と推測する尤もらしさ**（もっともらしさ）を表す数値を、「何々」を変数とする関数として捉えたものである。また単に尤度ともいう。

得られた標本が確率分布F_θによってどれだけ生成されやすいかの指標と解釈できる。
確率分布F_θの確率密度関数$f(x;\theta)$としたとき、標本の独立同一性より同時確率密度関数は、

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

パラメータを推定する際に、想定するパラメータ領域内において**尤度を最大化**するパラメータ値をパラメータの推定量として用いる事ができる。
これを**最尤推定量**と呼ぶ。
尤度の対数を取ると計算しやすくなることが多く、以下のような対数を取り(対数尤度)、対数尤度の最大化によって最尤推定量も得られる。

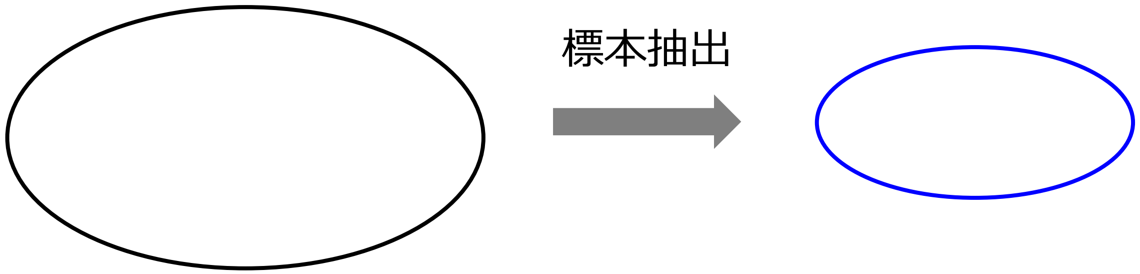
$$l(\theta) = \sum_{i=1}^{n} f(x_i;\theta)$$

尤度関数

【尤度のイメージ例】

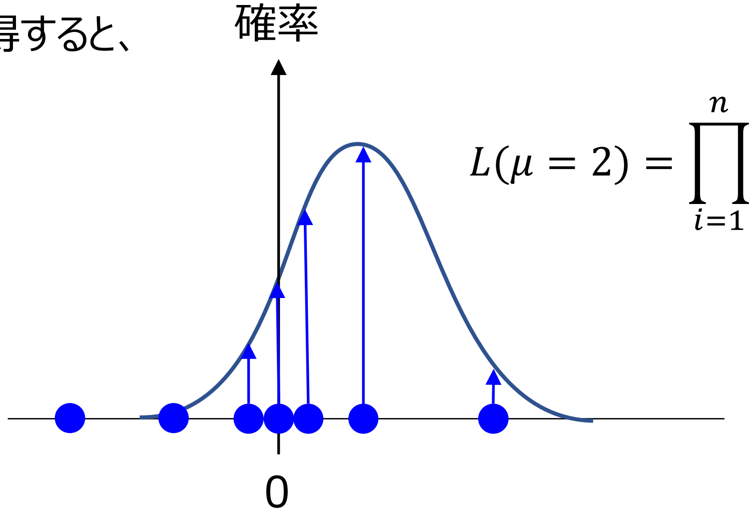
母集団が正規分布$N(0,22)$に従う。
標本データを取得するとして、その時点では平均μは未知、分散$\sigma = 2$が分かっているとし μ を推定するとする。

母集団$N(0,2^2)$

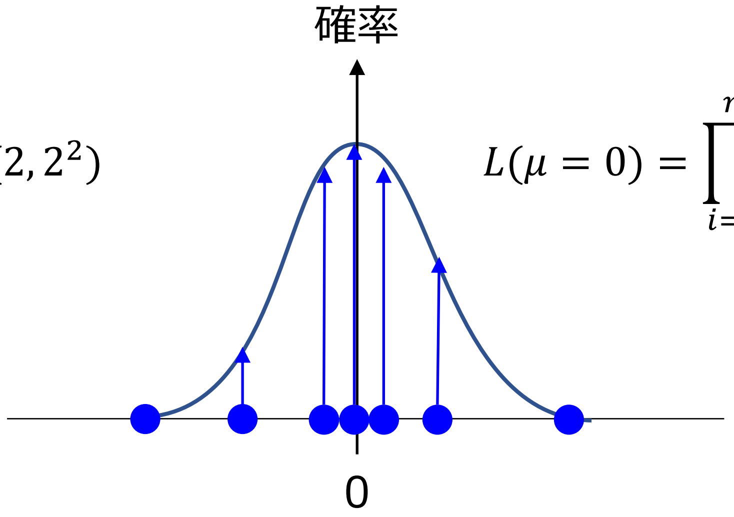


推定したμが母集団分布に近いほど
尤度関数$L(\mu)$が大きくなる。
つまり平均μは、**2よりも0の方が尤もらしい**。

例えば2つのμのパターンを考える。
そこにサンプル●を母集団から取得すると、
●は母集団の確率分布に従う



$$L(\mu = 2) = \prod_{i=1}^{n} N(2,2^2)$$



$$L(\mu = 0) = \prod_{i=1}^{n} N(0,2^2)$$

条件付尤度

【条件付尤度】

ある確率分布関数$f(X|\theta,\varphi)$を考えたときに推定したいパラメータθと、興味が無い局外パラメータφがあるとする。
この時にφをある統計量で条件付けをした際の、θの尤度計算を行う。

φの推定量をTとして以下のように分解できる。

$$f(X|\theta,\varphi) = f(X,T|\theta,\varphi) = \underbrace{f(X|\theta,t)}_{\varphi\text{に依存しない}}\underbrace{f(T|\varphi)}_{\varphi\text{を条件付け}}$$

この確率分布関数を尤度化すると、

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta,\varphi) = \prod_{i=1}^{n} f(x_i|\theta,t)f(T|\varphi)$$

条件付尤度

【条件付対数尤度】

パラメータθを以下の尤度関数から推定する場合、解析的には解けない事が多いためコンピュータを使って計算する。
しかし、確率は(0,1)の数値であるため、例えば0に近い量で0.00000000000*のようになってくると表現できなくなりアンダーフローしてしまう。
そのため、対数を取って計算するようにする。

尤度関数に対数を取ると、以下のように書ける。

$$l(\theta) = \log L(\theta) = \log \prod_{i=1}^{n} f(x_i|\theta,\varphi) = \sum_{i=!}^{n} \log f(x_i|\theta,\varphi) = \sum_{i=1}^{n} \log f(x_i|\theta,t) f(T|\varphi) = \sum_{i=1}^{n} (\log f(x_i|\theta,t) + \log f(T|\varphi))$$

この式よりθを推定するとする。その場合は、対数尤度関数を微分して極値を求める。
上式の最後の部分の$\log f(T|\varphi)$は、θには関係が無いため微分するとゼロになるため、最終的には以下の条件でθを求める。

$$\frac{d}{d\theta} l(\theta) = \frac{d}{d\theta} \sum_{i=1}^{n} (\log f(x_i|\theta,t) + \log f(T|\varphi)) = \frac{d}{d\theta} \sum_{i=1}^{n} (\log f(x_i|\theta,t)) = 0$$