最尤法と十分統計量

第１回講義：資料３

土居正明

注) 最尤法と不偏推定量（UMVU まで) については仮定しています。

1　最尤法の復習

まずは最尤法の復習として、いくつかの分布・パラメータについて最尤推定量を求めてみましょう。

1.1　最尤推定量の具体例

「例１：二項分布」

二項分布の確率関数は

$$f(x|n,p) = {}_nC_x p^x (1-p)^{n-x} \quad (0 < p < 1)$$

と表せます (n は固定された定数とします)。データ $x = m$ が与えられた時の p の尤度関数は

$$L(p|x=m,n) = {}_nC_m p^m (1-p)^{n-m}$$

となります。対数尤度は

$$\begin{aligned} l(p|x=m,n) &= \log\left(L(p|x=m,n)\right) \\ &= \log\left({}_nC_m p^m (1-p)^{n-m}\right) \\ &= \log\left({}_nC_m\right) + m\log p + (n-m)\log(1-p) \end{aligned}$$

ここで、$l(p|x=m,n)$ を p で微分すると、

$$\frac{\partial l(p|x=m,n)}{\partial p} = \frac{m}{p} - \frac{n-m}{1-p}$$

これより、

$$\frac{\partial l(p|x=m,n)}{\partial p} = 0$$

を解くと

$$\begin{aligned} \frac{m}{p} - \frac{n-m}{1-p} = 0 &\Longleftrightarrow (1-p)m - (n-m)p = 0 \\ &\Longleftrightarrow np = m \end{aligned}$$

従って、p の最尤推定量は

$$\widehat{p} = \frac{m}{n}$$

となります。

「例２：Poisson 分布」

Poisson 分布の確率関数は

$$f(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} \quad (\lambda > 0)$$

です。n 個の独立なデータ $x = (x_1, \cdots, x_n)$ が得られたときの λ の尤度関数は

$$\begin{aligned} L(\lambda|x) &= \prod_{i=1}^{n}\Big(\frac{\lambda^{x_i}}{(x_i)!}e^{-\lambda}\Big) \\ &= \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n (x_i)!}e^{-n\lambda} \end{aligned}$$

対数尤度関数は

$$\begin{aligned} l(\lambda|x) &= \log\Big(\frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n (x_i)!}e^{-n\lambda}\Big) \\ &= \Big(\sum_{i=1}^{n} x_i\Big)\log\lambda - \sum_{i=1}^{n}\log\left((x_i)!\right) - n\lambda \end{aligned}$$

これを λ で微分すると、

$$\frac{\partial l(\lambda|x)}{\partial\lambda} = \frac{\sum_{i=1}^n x_i}{\lambda} - n$$

よって、

$$\frac{\partial l(\lambda|x)}{\partial\lambda} = 0$$

を解くと、λ の最尤推定量

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

が得られます。

「例３：正規分布」

正規分布の確率密度関数は

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) \ \ (-\infty < \mu < \infty, 0 < \sigma < \infty)$$

です。今、$\underline{\text{平均 } \mu, \text{分散 } \sigma^2 \text{ はともに未知と仮定}}$ します。

n 個のデータ $x = (x_1, \cdot, x_n)$ が与えられた時の尤度関数は

$$\begin{aligned} L(\mu,\sigma^2|x) &= \prod_{i=1}^{n}\Big(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{(x_i-\mu)^2}{2\sigma^2}\Big)\Big) \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^n}\exp\Big(-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\Big) \end{aligned}$$

対数尤度関数は、

$$l(\mu,\sigma^2|x) = -n\log\sqrt{2\pi\sigma^2} - \frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}$$

となります。これを μ, σ^2 でそれぞれ微分します (σ^2 は「２乗」込みで一つの文字と考えます)。
(μ で微分)

$$
\begin{aligned}
\frac{\partial l(\mu, \sigma^2|x)}{\partial \mu} &= -\frac{1}{2\sigma^2} \cdot (-2\sum_{i=1}^{n}(x_i - \mu)) \\
&= \frac{1}{\sigma^2}(\sum_{i=1}^{n} x_i - n\mu)
\end{aligned}
$$

(σ^2 で微分)

$$
\frac{\partial l(\mu, \sigma^2|x)}{\partial (\sigma^2)} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2(\sigma^2)^2}
$$

いま、

$$
\begin{aligned}
\frac{\partial l(\mu, \sigma^2|x)}{\partial \mu} &= 0 \\
\frac{\partial l(\mu, \sigma^2|x)}{\partial (\sigma^2)} &= 0
\end{aligned}
$$

つまり、

$$
\frac{1}{\sigma^2}(\sum_{i=1}^{n} x_i - n\mu) = 0 \tag{1}
$$

$$
-\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2(\sigma^2)^2} = 0 \tag{2}
$$

を同時に満たす $(\mu, \sigma^2) = (\widehat{\mu}, \widehat{\sigma^2})$ が最尤推定量となります。まず (1) より、

$$
\frac{1}{\sigma^2}(\sum_{i=1}^{n} x_i - n\mu) = 0
$$

より、μ の最尤推定量は

$$
\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{(のちのため、これを } \bar{x} \text{ とも書きます)}
$$

です。次に、これを (2) に代入して、

$$
\begin{aligned}
-\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{\sum_{i=1}^{n}(x_i - \widehat{\mu})^2}{2(\sigma^2)^2} &= 0 \\
-n\sigma^2 + \sum_{i=1}^{n}(x_i - \widehat{\mu})^2 &= 0
\end{aligned}
$$

これより、σ^2 の最尤推定量は

$$
\begin{aligned}
\widehat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{\mu})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2
\end{aligned}
$$

です。

1.2　注意

最尤推定量は結構良い推定量です[*1]。しかし、上の例の中では最後の $\widehat{\sigma^2}$ だけは不偏推定量になりません。このように「もう少し工夫の余地がある」場合もいくらか見られます。

2　十分統計量とは

2.1　はじめに

今回十分統計量を扱うのは、**「完備十分統計量」**を用いる定理（下の**「定理２」**や**「系３」**）を使って、**ある種の不偏推定量が UMVU であることを確かめる**ことができるようになるからです。十分統計量「それ自身」の理解もそれはそれで大事なのですが、このような事情から今回は「使い方」（の１つ）に重点を置いたご説明をします。

2.2　イメージから

十分統計量の定義は大変分かりにくいです。したがって、定義より先にイメージからお伝えしましょう。

「イメージ：十分統計量」

十分統計量とは、最尤推定量を作るのに十分なデータの固まり である。

この表現でもよく分かりませんね。では、先ほど計算した例を用いてご説明しましょう。
「例２：Poisson 分布」の場合、最尤推定量は

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

です。このときに、**「$\widehat{\lambda}$ を求めるには、何があれば十分か？」**という風に考えます。$x_1, x_2, \cdots, x_n$ の１個１個のデータは必要ではなく、全体の和をとった

$$T(x) = \left(\sum_{i=1}^{n} x_i\right) \quad \left(\text{つまり、} T_1(x) = \sum_{i=1}^{n} x_i\right)$$

が分かれば、最尤推定量は計算できますね[*2]。この $T(x)$ が十分統計量の例です。

しかしこれだけではありません。たとえば、極端な例では全部のデータがあってももちろん「十分」と言えるわけです。つまり、

$$T'(x) = (x_1, \cdots, x_n) \quad (\text{つまり、} T'_1(x) = x_1, \cdots, T'_n(x) = x_n)$$

もまた十分統計量です。さらに、

$$T''(x) = \left(x_1, \sum_{i=2}^{n} x_i\right) \quad \left(\text{つまり、} T''_1(x) = x_1,\ T''_2(x) = \sum_{i=2}^{n} x_i\right)$$

$$T'''(x) = \left(x_1, x_2, \sum_{i=2}^{n} x_i\right) \quad \left(\text{つまり、} T'''_1(x) = x_1,\ T'''_2(x) = x_2,\ T'''_3(x) = \sum_{i=2}^{n} x_i\right)$$

も全て十分統計量です[*3]。たくさんありますので、**「十分統計量の中でさらによい性質を持ったもの」**が欲しくなります。今考えているのは「最尤推定量を求めるのに何があれば十分なのか？」という問いですので、**「できれば少ない方がいい」**

[*1] 具体的には、$n \to \infty$ のとき UMVU と同程度の性質を持ちます。
[*2] つまり例えば、「Poisson 分布に従うデータを 100 個とりました。１個１個のデータは教えてあげないけれど、合計は 300 でした」と言われたら、最尤推定量は $\widehat{\lambda} = \frac{300}{100} = 3$ と計算できますね。同様に、「1 個目は 10 で、残り 99 個の合計は 290 でした」という言い方でも $\widehat{\lambda}$ は求まりますね。
[*3] $T'''_2(x)$ と $T'''_3(x)$ は x_2 が重複していますが、最尤推定量を求めるのに「十分」であることには変わりありません。

[4]という風に考えるのは自然でしょう。そこで 十分統計量の中でより良いもの（情報がきちんと縮約されているもの）として**「完備十分統計量」**を考えるのですが、その前にいくつか準備をします。

まずは具体例についてもう少し触れておきましょう。

「例３：正規分布」

最尤推定量は

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

$$\begin{aligned}\widehat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ &= \frac{1}{n}\Big(\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2\Big) \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2\end{aligned}$$

これより、十分統計量としては

$$T(x) = \Big(\bar{x}, \sum_{i=1}^{n} x_i^2\Big)$$

$$T'(x) = (x_1, \cdots, x_n)$$

$$T''(x) = \Big(x_1, \bar{x}, \sum_{i=2}^{n} x_i^2\Big)$$

のようにやはり色々と考えられます。

逆に、

$$\widetilde{T'} = (x_1, \cdots, x_{n-1})$$

$$\widetilde{T''} = (x_1, \bar{x}, \sum_{i=3}^{n} x_i^2)$$

のようになものは十分統計量にはなりません。

2.3　定義：十分統計量

十分統計量の求め方・具体例についてイメージがいくらかは沸いてきたのではないかと思いますので、定義に入ります。まず一つ記号の説明をしておきます。

2.3.1　十分統計量の関数

データ $X = (X_1, X_2, X_3)$ が与えられ、$T(X) = (T_1(X), T_2(X))$ とおきます。ここで例えば、

$$T_1(X) = \sum_{i=1}^{3} X_i$$

$$T_2(X) = \sum_{i=1}^{3} X_i^2$$

[4] 実は、この観点を推し進めていったときに**最小十分統計量**という別の概念も出てくるのですが、今回は触れないことにします。

とします。ここで、$g(T(X))$ の形の関数 とは例えば、

$$g_1(T(X)) = T_1(X) + T_2(X) = \sum_{i=1}^{3} X_i + \sum_{i=1}^{3} X_i^2$$
$$g_2(T(X)) = \exp\left(T_1(X) \cdot T_2(X)\right) = \exp\left(\left(\sum_{i=1}^{3} X_i\right) \cdot \left(\sum_{i=1}^{3} X_i^2\right)\right)$$

のように、$T_1(X), T_2(X)$ の形以外では X_1, X_2, X_3 が出てこない ものを指します。このとき、$g(T(X))$ の形の関数を**統計量** $T(X)$ **の関数** と呼びます。

逆に

$$T_1(X) + T_2(X) + X_2$$
$$X_1 \exp(T_1(X))$$

のようなものは $g(T(X))$ とは表せません。

2.3.2 十分統計量の定義

では、定義です。

「定義：十分統計量」

データ $X = (X_1, \cdots, X_n)$ が与えられた時の Θ の尤度関数を $L(\Theta|X)$ とおく。このとき、$T(X) = (T_1(X), \cdots, T_k(X))$ が $\Theta = (\theta_1, \cdots, \theta_m)$ の**十分統計量**であるとは、

$$L(\Theta|X) = g_\Theta(T(X)) \cdot h(X) \tag{3}$$

と表せることである。ここで $g_\Theta(T(X))$ は、統計量 $T(X) = (T_1(X), \cdots, T_k(X))$ の関数であり、さらに $\Theta = (\theta_1, \cdots, \theta_m)$ を含む。一方 $h(X)$ は、$X = (X_1, \cdots, X_n)$ の要素はいくらでも含んでよいが $\Theta = (\theta_1, \cdots, \theta_m)$ の要素は含まない[*5]。

上で計算した具体例とこの式の形がどう結びつくんだ、という風に思われる方も多いと思いますので、ご説明します。かなり抽象的な式ですが、(3) から最尤推定量を計算しようとしてみます。まずは、対数尤度関数ですが、

$$\begin{aligned} l(\Theta|X) &= \log\left(g_\Theta(T(X)) \cdot h(X)\right) \\ &= \log\left(g_\Theta(T(X)\right) + \log\left(h(X)\right) \end{aligned}$$

となります。ここで、ある θ_i でこの $l(\Theta|x)$ を微分してみます。すると、

$$\frac{\partial l(\Theta|X)}{\partial \theta_i} = \frac{\partial g_\Theta(T(X))}{\partial \theta_i} + 0$$

ここで、$+0$ の部分には $\log\left(h(X)\right)$ がありました。しかしこの部分は Θ を含まないので、θ_i で微分すると消えてしまいます。つまり $h(X)$ **の部分は、最尤推定量を求めるのに関係のない部分**ということになります。

そして、最尤推定量を求めるのに使うのは $T(X) = (T_1(X), \cdots, T_k(X))$ と $\Theta = (\theta_1, \cdots \theta_m)$ と、あとは定数のみ となります[*6]。

具体的に先の例で見てみましょう。**「例 2：Poisson 分布」**において、対数尤度関数は

$$l(\lambda|x) = \left(\sum_{i=1}^{n} x_i\right) \log\lambda - \sum_{i=1}^{n} \log\left((x_i)!\right) - n\lambda$$

です。これを λ で微分すると、真ん中の $\sum_{i=1}^{n} \log\left((x_i)!\right)$ の部分が消えてしまいますね。従って、「最尤推定量にはどうでもいい部分」として十分統計量とは関係ないものと考えることができます。つまり $\log(h(x)) = \sum_{i=1}^{n} \log\left((x_i)!\right)$ としてよ

[*5] ここで、$h(X)$ は $T(X) = (T_1(X), \cdots, T_k(X))$ を含んでもよいです。下でご説明するように、最尤推定量を計算するときには消えてしまうので関係ないのです。

[*6] つまり、**イメージ**で書きました「十分統計量とは、最尤推定量を作るのに十分なデータの固まりである」ということですね。

いことになります。一方で、残りの部分が $g_\Theta(T(X)$ に対応する部分です。今、パラメータは λ 1つだけあり、データは $x=(x_1,\cdots,x_n)$ なので、

$$\log\left(g_\lambda(T(x))\right) = \Bigl(\sum_{i=1}^{n} x_i\Bigr)\log\lambda - n\lambda$$

とおいてやりますと、例えば

$$T(x) = \sum_{i=1}^{n} x_i$$

とおけば、

$$\log g_\lambda(T(x)) = T(x)\log\lambda - n\lambda$$

となり、$T(x)$ 以外に $x=(x_1,\cdots,x_n)$ のデータは現れませんので、$T(x)$ は十分統計量ということになります（n は定数です)*7。

2.4 指数型分布族

十分統計量の定義の式について、もう少し見ていきましょう。$h(X)$ の部分は「消えてしまう部分」なのであまり興味がありません。一方、$g_\Theta(T(X))$ の部分は最尤推定量を求めるのに本質的に重要な部分です。ですので、ここの部分が「分かりやすい」ものをグループとしてまとめてやりましょう。対数尤度関数を考えるときに考えやすくなるような形のものをまとめます。

「定義：指数型分布族」

ある確率分布において、データ $X=(X_1,\cdots,X_n)$ を与えた時の Θ の尤度関数が、

$$L(\Theta|X) = h(X)\exp\Bigl(\sum_{i=1}^{m} a_i(\Theta)\cdot T_i(X) - \psi(\Theta)\Bigr)$$

$$(a_i(\Theta)\ (1\le i\le m), \psi(\Theta) \text{ はともに}\Theta\text{は含んでよいが、}X \text{ は含まない})$$

とあらわされるとき、この確率分布は**指数型分布族**という $\Bigl(T(X)=(T_1(X),\cdots T_m(X))$ は明らかに Θ の十分統計量である$\Bigr)$。

なぜわざわざ指数 (exp) の形にしたかというと、対数をとると消えてしまうからです。実際、対数尤度関数は

$$l(\Theta|X) = \log\left(h(X)\right) + \sum_{i=1}^{m} a_i(\Theta)\cdot T_i(X) - \psi(\Theta)$$

となり、exp(　) の中身がそのまま出てきます。最初の $\log\left(h(X)\right)$ の部分は、最尤推定量の計算にとってはないのと同じですから、要は**対数尤度関数が、本質的には十分統計量の要素と Θ の関数の１次結合**という大変簡単な形になります。

実は、先に例としてあげました二項分布・Poisson 分布・正規分布は全て指数型分布族に含まれます。一つずつ確認しましょう。なお、$\log\left(h(x)\right)$ の部分は本質的ではないので、放っておきます。

「例１：二項分布」

$$l(p|x=m,n) = \log\left({}_nC_m\right) + m\log p + (n-m)\log(1-p)$$

*7 こちらの正確な定義からも、十分統計量はたくさんありうることにご注意ください。先にあげました $T_1''(x)=x_1,\ T_2''(x)=\sum_{i=2}^{n} x_i$ のときの $T''(x)=(T_1''(x),\ T_2''(x))$ なども十分統計量です。

なので、p のついていない $\log({}_nC_m)$ はおいておいて、

$$\begin{aligned}
&T_1(X) = X,\ T_2(X) = n - X \\
&a_1(p) = \log p,\ a_2(p) = \log(1-p)
\end{aligned}$$

とおけば、

$$l(p|x = m, n) = \log({}_nC_m) + a_1(p)\cdot T_1(m) + a_2(p)\cdot T_2(m)$$

となり、二項分布が指数型分布族であり、$T(X) = (T_1(X), T_2(X))$ が十分統計量であることも分かりました。

しかし、実はこれは最も簡単な表現ではありません。m は一見２カ所にあるように見えますが、１カ所にまとめられるからです。まとめるとこのようになります。

$$\begin{aligned}
l(p|x = m, n) &= \log({}_nC_m) + m(\log p - \log(1-p)) + n\log(1-p) \\
&= \log({}_nC_m) + m\log\left(\frac{p}{1-p}\right) + n\log(1-p)
\end{aligned}$$

こうすれば、

$$T_1'(X) = X,\ a_1'(p) = \log\left(\frac{p}{1-p}\right),\ \psi(p) = -n\log(1-p)$$

とおくことで (n はデータ X ではない定数なので $\psi(p)$ に入っても問題ありません。)、

$$l(p|x = m, n) = \log({}_nC_m) + a_1'(p)\cdot T_1'(m) - \psi(p)$$

となり、十分統計量は $T'(X) = (T_1'(X)) = (X)$ １つですみます。

「例２：Poisson 分布」

$$l(\lambda|x) = \left(\sum_{i=1}^{n} x_i\right)\log\lambda - \sum_{i=1}^{n}\log((x_i)!) - n\lambda$$

より、

$$T_1(X) = \sum_{i=1}^{n} X_i,\ a_1(\lambda) = \log\lambda,\ \psi(\lambda) = n\lambda$$

とおくと、

$$l(\lambda|x) = -\sum_{i=1}^{n}\log((x_i)!) + a_1(\lambda)\cdot T_1(x) - \psi(\lambda)$$

となり、λ の十分統計量 (のうちの１つ) は $T(X) = (T_1(X)) = (\sum_{i=1}^{n} X_i)$ となります。

「例題３：正規分布」

$$\begin{aligned}
l(\mu, \sigma^2|x) &= -n\log\sqrt{2\pi\sigma^2} - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2} \\
&= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{\sum_{i=1}^{n} x_i^2 - 2\mu\sum_{i=1}^{n} x_i + n\mu^2}{2\sigma^2} \\
&= -\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{n\mu^2}{2\sigma^2} - \frac{n}{2}\log(2\pi\sigma^2)
\end{aligned}$$

より、

$$T_1(X)=\sum_{i=1}^n X_i^2,\ T_2(X)=\sum_{i=1}^n X_i$$
$$a_1(\mu,\sigma^2)=-\frac{1}{2\sigma^2},\ a_2(\mu,\sigma^2)=\frac{\mu}{\sigma^2},\ \psi(\mu,\sigma^2)=\frac{n\mu^2}{2\sigma^2}+\frac{n}{2}\log(2\pi\sigma^2)$$

とおくと、

$$l(\mu,\sigma^2|x)=a_1(\mu,\sigma^2)\cdot T_1(x)+a_2(\mu,\sigma^2)\cdot T_2(X)-\psi(\mu,\sigma^2)$$

となります。これより、(μ,σ^2) の十分統計量 (のうちの 1 つ) は $T(X)=(T_1(X),T_2(X))=\left(\sum_{i=1}^n X_i^2,\sum_{i=1}^n X_i\right)$ となります。

「ポイント」

指数型分布族において、十分統計量は対数尤度関数から簡単に求まる。

3 完備十分統計量と重要な定理

さて、「十分統計量」というだけでは様々なものが含まれていることは先に述べました。そこで、**「よりよい十分統計量」**を考えたいと思います。それが**完備十分統計量**です。しかし、実はこの定義は今回の一連の講義では実にどうでもよいのです。ですので、定義は最後の補足に回しまして、重要な定理を見ていくことにしましょう。

3.1 重要な２定理

十分統計量の使い方についてはいくつかあるのですが、本稿で最も重要なのは次の 2 つの定理です。

「定理 1：完備十分統計量の判定条件」

指数型分布族

$$L(\Theta|X)=h(X)\exp\left(\sum_{i=1}^m T_i(X)a_i(\Theta)-\psi(\Theta)\right)$$

において、集合 $\{a_1(\Theta),\cdots,a_m(\Theta)\}$ が R^m 内で開集合を含む とする。このとき、統計量 $T(X)=(T_1(X),\cdots,T_m(X))$ は Θ の完備十分統計量である。

ここで、「R^m 内で開集合を含む」とは例えば **十分統計量が２つ のとき、**$(x,y)=(a_1(\Theta),a_2(\Theta))$ **の占める領域が、**R^2 **において小さい円盤を含む**、という意味です。また、十分統計量が 3 つ のとき、$(x,y,z)=(a_1(\Theta),a_2(\Theta),a_3(\Theta))$ の占める領域が R^3 において小さい球を含むことです。十分統計量が 1 つのときが、少し難しいのですが、このときは $y=a_1(\Theta)$ のとりうる範囲（値域）が R において「(a,b) の形の開区間を含む」ということがこれに対応します。

「定理２：完備十分統計量と UMVU」

完備十分統計量の関数である不偏推定量が存在すれば、それは一意的であり UMVU である。

3.2 定理の使い方

さて、ここで上の定理と最尤法とのつながりを見ていきましょう。**最尤推定量は、十分統計量のみで構成される**ということは、**イメージ**や**定義**からお分かりいただけると思います。そこで、もう一歩押し進めて、**「最尤推定量は完備十分統計量のみで構成される（完備十分統計量の関数である）」**ということが 言えたとします[*8]。このとき、**「定理２」**を使う準備が

[*8] この点は指数型分布族に限れば常に成り立つような気がしなくもないのですが、きちんと証明していないので不明です。ご存知の方がいらっしゃれば教えてください。

できます。すなわち、次の**「系」**が成り立ちます。

「系３」

ある最尤推定量が完備十分統計量の関数であるとする。このとき、この最尤推定量の関数として作られた不偏推定量はUMVU である。

「系３」は、条件として、

(i) ある最尤推定量が完備十分統計量である

(ii) 不偏推定量が最尤推定量の関数である

の２つを確認して、ようやく使えるものであることに注意してください。

では、**「定理１」**から順に使い方を見ていきましょう。これは、「十分統計量が完備十分統計量かを調べる」定理です。**「例１」**では２種類の十分統計量を見ましたので、それらが完備かどうかを調べてみましょう。まず最初の十分統計量は、

$$T_1(X) = X,\ T_2(X) = n - X$$

でした。そして p を含む部分は

$$a_1(p) = \log p,\ a_2(p) = \log{(1-p)}$$

となっていました。ここで、$(x, y) = (a_1(p),\ a_2(p))$ が $0 < p < 1$ のときに描く図形が R^2 で開集合を含めば**「定理１」**が使えます。

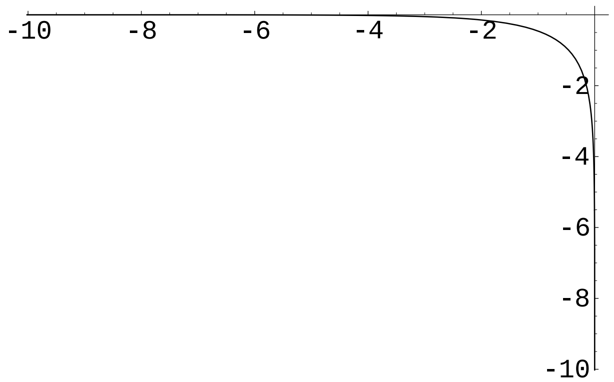

図１　$(x, y) = (\log p, \log{(1-p)})$ のグラフ $(0 < p < 1)$

上図のように**グラフは１本の曲線となるので、R^2 の開集合（小さな円盤）を含みません。**したがって、上の $T(X) = (T_1(X),\ T_2(X))$ が完備十分統計量であるとはいえませんでした[*9]。

次に２つ目の十分統計量ですが、

$$T'_1(X) = X$$

だけでした。そして p を含む部分は

$$a'_1(p) = \log\left(\frac{p}{1-p}\right)$$

でした。今度は $y = a'_1(p)$ の $0 < p < 1$ における値域が R の開集合を含んでいればよいことになります。

[*9] 完備十分統計量で「ないことが証明された」わけではありませんのでご注意ください。**「定理１」**の方法で完備十分統計量であることがいえなかっただけです。

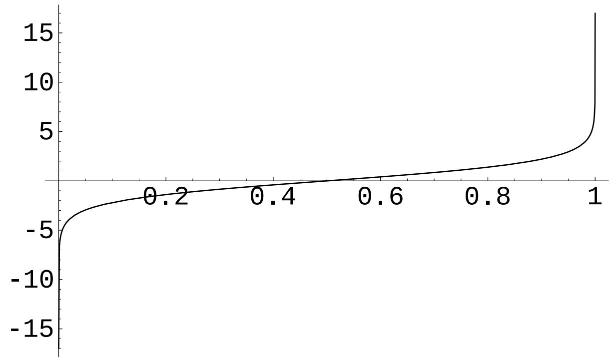

図 2　$y = \log \frac{p}{1-p}$ のグラフ $(0 < p < 1)$

上のグラフから明らかなように、y の値域は、たとえば R の開集合 $(-5,\ 5)$ を含んでいます。
このため、**「定理１」**より、この $\underline{T'(X) = (T_1'(X))}$ は p の完備十分統計量である ことがいえました。

では、**「例２：Poisson 分布」**に移りましょう。今度は λ の十分統計量として

$$T_1(X) = \sum_{i=1}^{n} X_i$$

が、そして λ を含む部分として

$$a_1(\lambda) = \log \lambda$$

となり、$y = a_1(\lambda)$ の値域が R の開集合を含めばよいことになります。$0 < \lambda < \infty$ のとき、$\log \lambda$ は全ての実数をとりえますので、明らかに開区間を含みます。したがって、**「定理１」**より $T(X) = (T_1(X))$ は λ の完備十分統計量であることが分かりました。

最後に**「例３：正規分布」**です。今度は十分統計量として

$$T_1(X) = \sum_{i=1}^{n} X_i^2,\ T_2(X) = \sum_{i=1}^{n} X_i$$

を、(μ, σ^2) を含む部分として

$$a_1(\mu, \sigma^2) = -\frac{1}{2\sigma^2},\ a_2(\mu, \sigma^2) = \frac{\mu}{\sigma^2}$$

を考えました。このとき、

$$(x, y) = (a_1(\mu, \sigma^2),\ a_2(\mu, \sigma^2))$$

が R^2 の開集合を含めばよいのですが、例えば $(\frac{1}{2} \leq \sigma^2 \leq 1,\ -1 \leq \mu \leq 0)$ のときに占める領域は

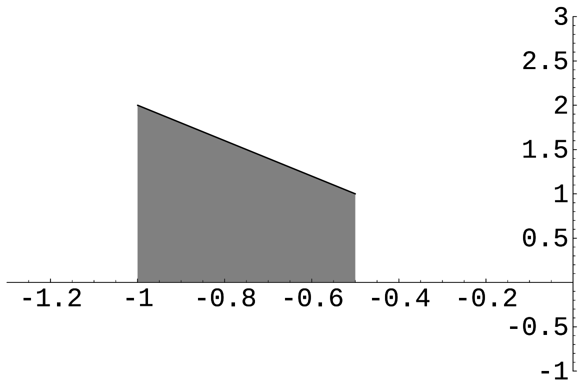

図３　$(x,y)=(-\frac{1}{\sigma^2},\frac{2\mu}{\sigma^2})$ が $(\frac{1}{2}\le\sigma^2\le 1,\ -1\le\mu\le 0)$ のときに占める領域

のようになりますので、R^2 での開集合（小さい円盤）を十分に含みます。したがって、**「定理１」**より

$$T(X)=(T_1(X),\ T_2(X))=\left(\sum_{i=1}^{n}X_i^2,\ \sum_{i=1}^{n}X_i\right)$$

は (μ,σ^2) の完備十分統計量になります。

さて、それでは次に**「定理２」**と、それからほとんど同じ**「系３」**は一度にご説明します。
まず**「例１：二項分布」**からです。今、$T'(X)=(T_1'(X))=(X)$ が p の完備十分統計量であることは分かりました。ここで、p の最尤推定量は

$$\widehat{p}=\frac{m}{n}=\frac{T_1'(m)}{n}$$

と p の完備十分統計量で表されるのですが、これがさらに不偏推定量になっています。したがって、**「定理２」**から**「完備十分統計量でできた不偏推定量だから UMVU」**となります。また、**「系３」**から、(i)「最尤推定量が完備十分統計量の関数」であり、(ii)「不偏推定量＝最尤推定量」という形で「不偏推定量は最尤推定量の関数」と考えることができるので、(i)・(ii) よりこの最尤推定量は UMVU、という風にも言えます。

「例２：Poisson 分布」では、

$$T(X)=(T_1(X))=\left(\sum_{i=1}^{n}X_i\right)$$

５が λ の完備十分統計量であり、λ の最尤推定量は

$$\widehat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{T_1(x)}{n}$$

となり完備十分統計量の関数で表されます。さらに、これも**「例１」**と同じく不偏推定量になっていますので、**「定理２」**や**「系３」**より UMVU であることが分かります。

最後に**「例３：正規分布」**です。今度は不偏推定量と最尤推定量が一致しない場合です。(μ,σ^2) の完備十分統計量は

$$T(X)=(T_1(X),\ T_2(X))=\left(\sum_{i=1}^{n}X_i^2,\ \sum_{i=1}^{n}X_i\right)$$

でした。ここで、σ^2 の最尤推定量は

$$\begin{aligned}\widehat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \frac{1}{n}(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2) \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\end{aligned}$$

ですので、完備十分統計量を用いて、

$$\widehat{\sigma^2} = \frac{1}{n}T_1(X) - \frac{1}{n^2}(T_2(X))^2$$

と書けます。したがって、**「系３」**の仮定 (i)**「最尤推定量は完備十分統計量の関数」**を満たしますね。よって、この最尤推定量を工夫して不偏推定量を作れば UMVU ということになります。ご存知のように、

$$\widehat{\sigma_u^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

は σ^2 の不偏推定量ですが、これは最尤推定量と

$$\widehat{\sigma_u^2} = \frac{n}{n-1}\cdot\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2) = \frac{n}{n-1}\widehat{\sigma^2}$$

という関係になり、**「系３」**(ii)**「この不偏推定量は最尤推定量の関数」**と仮定を満たしますので、**この不偏推定量は UMVU** となります。

4　まとめ

今回の講義（の特に前半）では、分散分析に出てくる色々な推定量が「よい推定量なのだ」ということを、UMVU などを用いて表現していきます。ですので、今回の**「定理１」**、**「定理２」**、**「系３」**は今後大事な場面で出てきます。しっかり使いこなせるようにしておいてください。

5 補足

今回の講義ではそれほど重要でない、もしくは本質的には重要ですが表立っては使わない事項をまとめておきます。

5.1 補足１：十分統計量の「普通によく使われる定義」

5.1.1 「普通によくつかわれる定義」

今回の十分統計量の定義は、通常なされるものとは表現が違います[*10]。一応、そちらの方を定理の形で書いておくことにします。

「定理４:十分統計量の通常なされる定義」

$T(X) = (T_1(X), \cdots, T_k(X))$ が $\Theta = (\theta_1, \cdots, \theta_m)$ の十分統計量であるための必要十分条件は、$T(X)$ を与えたときの、$X = (X_1, \cdots, X_n)$ の条件付分布が Θ に依存しないことである。

通常は、この定理の「十分統計量であるための必要十分条件は」と最後の「依存しないことである。」の部分を「十分統計量である、とは」と「依存しないときにいう。」に直して定義にします。しかし、今回の話の流れではこの性質を全く使わないのでこちらで定義するのはやめました。

5.1.2 使い方の例

例えば、ロジスティック回帰分析でマッチングなどをする際に、未知パラメータが「知りたいパラメータ」と「知らなくてもよいパラメータ」に分かれる場合があります。こういう際に、「知らなくてもよいパラメータ」に対して十分統計量を与えてやって、条件付分布を考えることで**知らなくてもよいパラメータを消すことができます。**そうしておいて、残りのパラメータを効率よく推定するという使い方がよくなされます。

こちらの使い方については、今回は用いませんのでこの程度にしておきます。

5.2 補足２：完備十分統計量の定義

最後に、完備十分統計量の定義を書いておきます。[**定理１**] の証明を理解されたい方などは、ご参考にしてください[*11]。

「定義:完備十分統計量」

十分統計量 $T(X)$ の任意の関数 $g(T)$ において、

$$E_\theta(g(T)) = 0 \ \ (\forall\theta) \Longrightarrow g(T) \equiv 0$$

が成り立つとき、$T(X)$ は**完備十分統計量**であるという。

[*10] **「定理４」**は必要十分条件を言っているので、定義として同値であることにご注意ください。

[*11] 十分統計量が完備であることを示すために、**「定理１」**が使えず、定義から直接求めなければならない場合もまれにありますので、知っておくとたまに役に立つかもしれません。