2012 年度秋学期　解析応用　第１２回

推定量と最尤原理

統計量と推定量

前々回の講義で説明した統計的推測では，母集団の度数分布（母集団分布）を特徴づける母数を，標本を用いて推定します。このとき，複数のデータからなる標本を取り出す必要があるので，それらのデータをひとつの値にまとめる計算を行ないます。このとき，標本から計算によって得られる量を**統計量 (statistic)** といいます。例えば，「母平均」という母数を推測するために，標本から「標本平均」という統計量を求めます。

何らかの統計量をもちいて，ある母数を一つの数値で推定することを**点推定**といい，そのときに用いた統計量をその母数の**推定量**といいます。例えば，「大数の法則」の説明で，標本の数が多ければ標本平均は「たいてい，ほぼ」母平均に等しいということを説明しました。これは，「標本平均」という統計量をもって，「母平均」という母数の推定量としていることになります。なお，推定の対象となる母数が θ であるとき，それに対応する推定量を $\hat{\theta}$ で表します。また，推定量の値を具体的数値で求めたものを**推定値**といいます。

一致性と不偏性

上で，標本平均は母平均の推定量である，と述べましたが，なぜそう言えるのでしょうか？ すなわち，ある母数を「推定している」統計量とは，どういう統計量をさしているのでしょうか？ このような「推定量のもつべき性質」の中で，とくに重要なのが次に述べる**一致性**と**不偏性**です。

一致性とは，標本サイズが十分大きければ，推定量が母数とわずかでもへだたっている確率が 0 に近づく，つまり確率収束することを意味します。この性質をもつ推定量を**一致推定量**といいます。「大数の法則」の説明で述べたように，標本平均は母平均に確率収束するので，標本平均は母平均の一致推定量です。このことは，標本平均を計算するために抽出する標本のサイズが大きくなればなるほど，標本平均が母集団平均と隔たった値になることはほとんどなく，「たいてい」標本平均＝母集団平均となることを意味します。

また，不偏性とは，推定量の期待値が母数に等しい，すなわち $E(\hat{\theta}) = \theta$ であることを意味し，不偏性のある推定量を**不偏推定量**といいます。前々回の説明で，標本平均 $\bar{X}$ と母平均 μ の間には $E(\bar{X}) = \mu$ という関係があることを示しました。したがって，標本平均は母平均の不偏推定量です。

「不偏」とは「えこひいきをしない」という意味です。標本平均の値は，そのときに偶然どんな標本がとりだされるかによって異なるので，何度も標本の組を取り出しては標本平均を計算することを繰り返すと，それらの標本平均はばらつきます。つまり，ある標本抽出で得られる標本平均は，母平均よりも大きいこともあれば小さいこともあります。しかし，「標本平均の期待値」，つまり長い目で見たときのそれらの平均が母平均に等しいということは，長い目で見れば，標本平均には母平均よりも大きいものと小さいものとが均等に現れることを意味します。すなわち，標本平均は母平均と「ほぼ」等しく，大きめになることも小さめになることもない，ということになります。

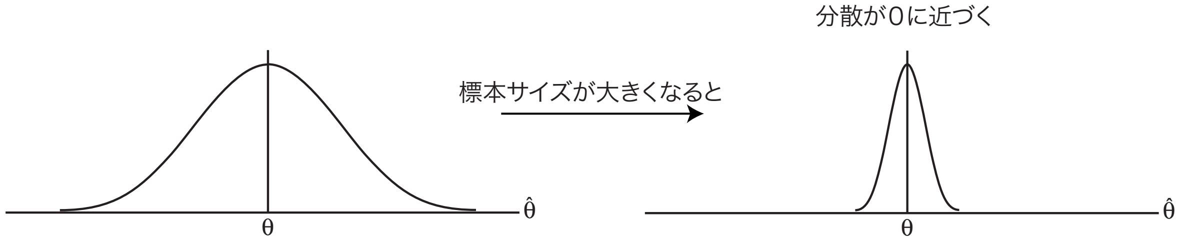

図 1: 一致性のある推定量の確率分布

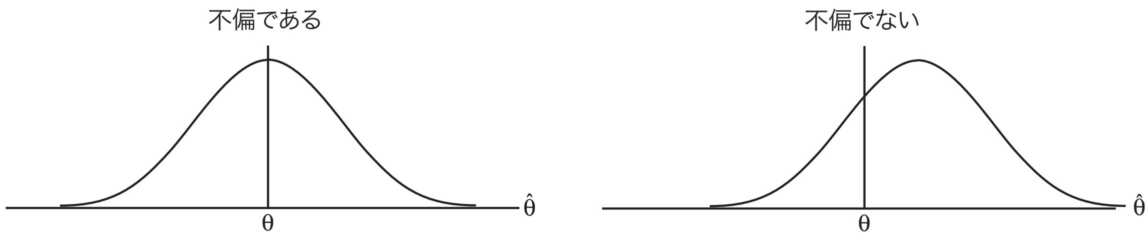

図 2: 推定量の確率分布と不偏性

不偏分散

ここまでで何度も述べているように，「標本平均 $\bar{X}$ の期待値」は母平均 μ に等しくなります。すなわち，標本平均は母平均の不偏推定量です。では，標本の分散を考えてみましょう。標本を $X_1, X_2, \ldots, X_n$ として，標本の分散を次のように求めてみます。

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n} \tag{1}$$

この S^2 は，母分散 σ^2 の不偏推定量になっているでしょうか？ それを調べるために，期待値 $E(S^2)$ を求めてみましょう。ここで $Y_i = X_i - \mu$ とおくと，

$$\begin{aligned} S^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}[(X_i - \mu) - (\bar{X} - \mu)]^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \end{aligned} \tag{2}$$

となりますから，

$$\begin{aligned} S^2 &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}((Y_i)^2 - 2\bar{Y}Y_i + \bar{Y}^2) \\ &= \frac{1}{n}\sum_{i=1}^{n}(Y_i)^2 - 2\bar{Y}\frac{1}{n}\sum_{i=1}^{n}Y_i + \bar{Y}^2\frac{1}{n}\sum_{i=1}^{n}1 \\ &= \frac{1}{n}\sum_{i=1}^{n}(Y_i)^2 - 2\bar{Y}\bar{Y} + \bar{Y}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i)^2 - \bar{Y}^2 \end{aligned} \tag{3}$$

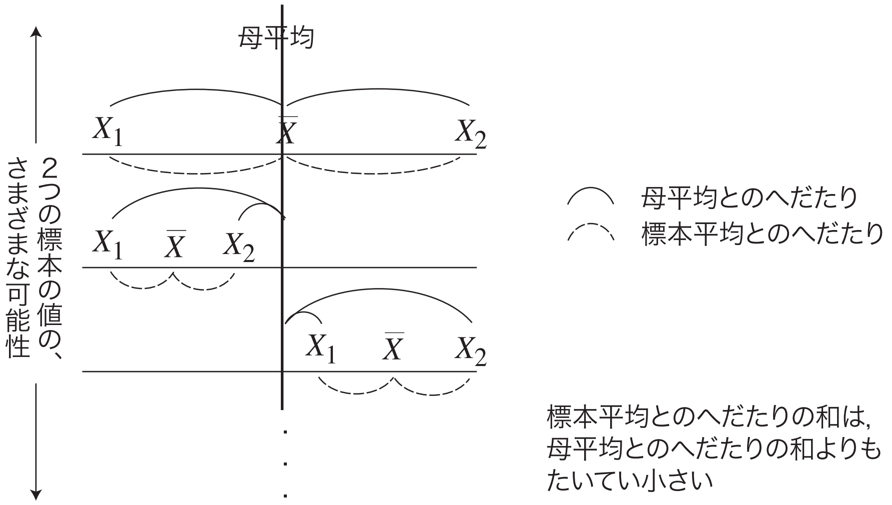

図 3: 不偏分散を求めるのに，なぜ $n-1$ で割るのか

となります。そこで，

$$\begin{aligned} E(S^2) &= E\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i)^2 - \bar{Y}^2\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}E[(Y_i)^2] - E[\bar{Y}^2] = \frac{1}{n}\sum_{i=1}^{n}E[(X_i-\mu)^2] - E[(\bar{X}-\mu)^2] \end{aligned} \tag{4}$$

となります。ここで，$E[(X_i-\mu)^2]$ は「標本サイズが十分大きいときの，標本と母平均との隔たりの２乗平均」ですから，母分散 σ^2 と等しくなります。また，$E[(\bar{X}-\mu)^2]$ は「標本平均の分散」で，これは第４回の講義で説明したように σ^2/n ですから，

$$E(S^2) = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 - \frac{\sigma^2}{n} = \frac{n\sigma^2}{n} - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \tag{5}$$

となります。つまり，S^2 の期待値は母分散 σ^2 とは等しくなく，S^2 は母分散の不偏推定量ではありません。母分散の不偏推定量は，上の式から，S^2 の $n/(n-1)$ 倍であることがわかります。この値，すなわち母分散の不偏推定量である

$$s^2 = \frac{(X_1-\bar{X})^2 + (X_2-\bar{X})^2 + \cdots + (X_n-\bar{X})^2}{n-1} \tag{6}$$

を**不偏分散**とよび，$n-1$ を自由度とよびます。

なぜ，標本分散 S^2 は不偏ではなく，母分散を小さく評価してしまうのでしょうか？ それを直観的に理解するために，図３をみてみましょう。図３では標本サイズを２とします。母分散は，各データと母平均とのへだたりの２乗の平均です。これに対して，n（この場合は 2）で割った標本分散は，標本と標本平均とのへだたりの２乗の平均になっています。２つの標本が，どちらも母平均から一方に偏ってへだたっていても，標本平均はつねに２つの標本の中間にあります。ですから，「標本と標本平均とのへだたり」は「標本と母平均とのへだたり」よりは小さくなります。この違いを調整するために，n ではなく $n-1$ で割っているのです。

最尤推定

ここまでは平均や分散を例にとって説明してきましたが，では一般的にはどうやって推定量を求めればよいでしょうか。それにはいろいろな方法がありますが，そのひとつとしてよく使われるのが**最尤原理**(さいゆう)です。これを理解するために，つぎの問題を考えてみましょう。

> あるくじびきは，当たる確率が p，はずれる確率が $1-p$ のベルヌーイ試行になっている。次の各段階で，母数 p をどのように推定したらよいか。
>
> (1) 1 回くじをひくと，当たりであった。
> (2) もう 1 回くじをひくと，今度ははずれであった。
> (3) さらに 3 回くじをひくと，当たり・当たり・はずれとなった。

上の問題で，まず，(1) の段階では「1 回当たった」という事実しかわかっていません。そこで，「現在わかっている現実の事象が起きる確率が最大になるような，母数の値」を，母数の推定値とします。このくじが 1 回当たる確率は p で表されます。p は 0 から 1 の間の何らかの値ですが，それらのうち「くじを 1 回ひいて，1 回当たる確率」がいちばん大きいのは，当然ながら $p=1$ のときです。そこで，この段階では p の推定値を $p=1$ とします。

(2) の段階では「1 回当たって，1 回はずれた」という事象が起きています。このような事象が起きる確率を $L(p)$ とすると，$L(p)=p(1-p)$ で表されます。(1) と同様に，0 から 1 の間で「1 回ひいて，1 回はずれる」確率が最大になるような p を $\hat{p}$ とすると，$\hat{p}=0.5$ となり，これをこの段階での p の推定値とします。

(3) の段階になると，現実に起きている事象は「3 回当たって，2 回はずれた」となっています。このような事象が起きる確率は $L(p)=p^3(1-p)^2$ で表されます。(1)(2) と同様に，0 から 1 の間でこの確率が最大になるような p を $\hat{p}$ とします。$L(p)$ が最大になるとき，$L(p)$ の微分 $L'(p)$ は 0 になります。すなわち $L'(\hat{p})=0$ で，$\hat{p}$ が 0, 1 でないことより

$$L'(\hat{p})=3\hat{p}^2(1-\hat{p})^2-2\hat{p}^3(1-\hat{p})=0 \tag{7}$$

となり，$\hat{p}=0.6$ となります。

ここまでの結論は当たり前のようですが，最尤原理とは，この例のように「考えられる母数の値のうち，今現在起きている事象が起きる確率がもっとも大きくなる値が，現在の時点では，その母数の推定値として一番尤もらしい」という考え方です。このとき，「今現在起きている事象が起きる確率」を**尤度**といい，尤度が最大になる母数の値を**最尤推定値**といいます。また，そのような値を与える推定量を**最尤推定量**といいます。標本抽出の場合でも考え方は同じで，「最も尤もらしい推定値とは，考えうる母数の値のうち『今現実に得られている標本が得られる確率』が最大になるもの」とします。

一般に，未知の母数を θ とし，その母集団から取り出した標本がしたがう確率分布（すなわち母集団分布）を，頻度関数または確率密度関数で表して $f(x;\theta)$ とします。いま，n 個のデータ $X_1=x_1, X_2=x_2, \ldots, X_n=x_n$ を標本として得たとき，尤度関数 $L(\theta)$ を

$$L(\theta)=f(x_1;\theta)\cdot f(x_2;\theta)\cdot\cdots\cdot f(x_n;\theta)=\prod_{i=1}^{n} f(x_i;\theta) \tag{8}$$

で定義し，これを最大にする $\theta=\hat{\theta}$ を最尤推定値とします。

なお，$L(\theta)$ は積の形になっていて扱いにくいので，これを和の形にかえるため，$L(\theta)$ のかわりに $\log L(\theta)$ を考えて

$$\log L(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta) \tag{9}$$

とするのが普通です。これを**対数尤度**といいます。

最尤推定の例：正規分布の場合

最尤推定量を求める例として，母集団分布が正規分布の時，母平均の最尤推定量が標本平均になることを確かめてみましょう。正規分布の密度関数は

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{10}$$

で表されます。この関数には 2 つの母数 μ と σ^2 を含むので，尤度関数は $L(\mu,\sigma^2)$ と表されます。標本としてとられた n 個のデータの値が $x_1, x_2, \ldots, x_n$ であるとき，対数尤度は

$$\begin{aligned}
\log L(\mu,\sigma^2) &= \sum_{i=1}^{n} \log\left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}\right] \\
&= \sum_{i=1}^{n} \left[\log \frac{1}{\sqrt{2\pi}\sigma} + \log\exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}\right] \\
&= n\log\frac{1}{\sqrt{2\pi}\sigma} + \sum_{i=1}^{n}\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= -n\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2
\end{aligned} \tag{11}$$

となります。ここで，μ の最尤推定値を $\hat{\mu}$ とすると，$\mu = \hat{\mu}$ のとき対数尤度が最大になるわけですから，対数尤度を μ で偏微分して $\mu = \hat{\mu}$ とおくと 0 となります。つまり，

$$\begin{aligned}
\frac{\partial}{\partial\mu}\log L(\mu,\sigma^2)\bigg|_{\mu=\hat{\mu}} &= \frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i-\hat{\mu}) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\hat{\mu}) = \frac{1}{\sigma^2}\left\{\sum_{i=1}^{n} x_i - n\hat{\mu}\right\} = 0
\end{aligned} \tag{12}$$

となります。したがって，このときの μ の最尤推定値 $\hat{\mu}$ は

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{13}$$

となります。これは標本の値 x_i が何であってもなりたちますから，一般に最尤推定量 $\hat{\mu}$ は標本平均に等しくなります。

問題

確率 p $(p \neq 0, p \neq 1)$ で成功するベルヌーイ試行を n 回行い，その i 回目の結果を

$$x_i = \begin{cases} 1 & i\text{回目の試行が成功} \\ 0 & i\text{回目の試行が失敗} \end{cases}$$

で表します。

(1) i 回目に結果 x_i が起きる確率を，p と x_i で表してください。
(2) 1 回め，2 回め，...，n 回めの試行の結果が，それぞれ $x_1, x_2, \ldots, x_n$ となる確率を求めてください。
(3) (2) の確率を最大にする p（p の最尤推定値）を求めてください。