Introduction to mathematical statistics ゼミ (第？回)

担当: 伊藤真道

未定

7　Chap.7 Sufficiency

7.1　Measures of Quality of Estimators

本節では，ある状況下における，いくつかの最適な推定値と検定を紹介する．記法は，これまでの章と同じ (だから，わからなくなったら，前の資料を参照してください．)．5.2 節では，点推定量が満たすべき性質として，一致性 (consistency) を紹介した．ある未知パラメータ θ の推定量 Y_n が一致推定量であるとは，Y_n が θ に確率収束する，つまり，

$$\lim_{n\to\infty} P[|Y_n-\theta|>\epsilon]=0,\ for\ \epsilon>0$$

が成立することであった．これは，サンプルサイズが大きくなると，推定量が真値に収束するということを表しており，非常に好ましい性質である．定理 6.1.3(テキストの番号に合わせています．) は，適切な状況下では，最尤推定量は一致推定量であることを主張する定理であり，大変重要である．他にも，推定量が満たすべき性質として，不偏性 (unbiasedness) がある．これは，推定量の期待値が，真値となること，つまり，

$$E[Y]=\theta$$

となることであり，少なくとも満たしてほしいものである．

もし，二つの不偏推定量があったとき，どちらの推定量を採用すべきだろうか．従来，二つの推定量の分散を比較し，より小さい分散を持つものが採用されてきた．これは，二つの推定量が漸近的に正規分布に従うとき，漸近分散の小さい推定量の方が，その漸近的な信頼区間が狭くなることから，妥当な方法であると判断できる．次に紹介する最小分散不偏推定量は，このことを基としている．

定義 7.1.1(最小分散不偏推定量 (Minimum Variance Unbiased Estimator, MVUE))

ある正定数 n について，$Y = u(X_1, X_2, \dots, X_n)$ が θ の最小分散不偏推定量 (Minimum Variance Unbiased Estimator(MVUE)) であるとは，Y が不偏 ($E(Y) = \theta$) であり，さらに，Y の分散が，他の全ての不偏推定量 Y^* と同じか，それ以下，つまり，

$$V(Y) \leq V(Y^*)$$

となることである．

例 7.1. $N(\theta, \sigma^2)$ の標本平均 $X_1, \dots, X_9 \overset{iid}{\sim} N(\theta, \sigma^2)$ とする．このとき，標本平均 $\bar{X} := (X_1 + \cdots + X_9)/9$ の分布は $N(\theta, \sigma^2/9)$ となる．ほにゃらら〜

¡note¿

次に，わずかに異なる複数の観測点から点推定を行うことを考える．$X_1, \dots, X_n$ を密度関数が $f(x, \theta)$ であるような分布からのサイズ n の無作為標本とし，$Y := u(X_1, X_2, \dots, X_n)$ を未知パラメータ θ を点推定する際の統計量とする．$\delta(y)$ を未知パラ θ の統計量である Y の観測値の関数とするなら，これは，統計量 Y の価値を決定する (decide) ため，decision function，あるいは，decision rule と呼ばれ，dicision function の値は，decision と呼ばれる．decision は正しいか，間

違っているかはわからないため，これを判断するために，真値と点推定値 $\delta(y)$ とのギャップを測る指標があると便利である．$[\theta, \delta(y)]$ のそれぞれの組み合わせについて，ギャップの深刻さを非負値 $\mathcal{L}[\theta, \delta(y)]$ として与えることを考えよう．この $\mathcal{L}[\theta, \delta(y)]$ は損失関数 $lossfunction$ と呼ばれる．さらに，損失関数の期待値は，リスク関数 (risk function) と呼ばれる．Y の pdf を $f_Y(y;\theta)$ と表すならば，リスク関数 $R(\theta, \delta)$ は，

$$R(\theta, \delta) = E\{\mathcal{L}[\theta, \delta(y)]\} = \int_{-\infty}^{\infty} \mathcal{L}[\theta, \delta(y)] f_Y(y;\theta) dy$$

で与えられる．$\forall\theta \in \Omega$ について，リスク $R(\theta, \delta)$ を最小化する decision function を選ぶことが望ましいが，一般的には不可能である．そのため，

1. decision function を特定のクラスに限定する
2. リスク関数の並び替えを行う方法を利用する

などが必要である．これに関して，テキストの p.376 の例がわかりやすい (ので，参照して)．

例 7.2. *minimax criterion* ミニマックス基準 *(minimax criterion)* とは，リスク関数の最大値を最小化するような決定関数を選んじゃおう！！という考え方に依る基準．*(p.376* の例を参照*)*

以上の例から，

1. decision function に対して，制約を課さなければ，リスク関数を一様に最小化する decision function を見つけることは困難.
2. 最良の decision function の選ぶ基準の一つに，minimax 基準と呼ばれるものがある．これは，以下のように表すことができる．任意の $\theta \in \Omega, \delta(y)$ に対して，

   $$\underset{\theta}{max} R[\theta, \delta_0(y)] \leq \underset{\theta}{max} R[\theta, \delta(y)]$$

   となるような $\delta_0(y)$ を選択する方法である．このような $\delta_0(y)$ を minimax decision function と呼ぶ.

もし，$E[\delta(y)] = \theta, \mathcal{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$ とするならば，リスク関数を最小化する decision function は，最小分散の不偏推定量を与える．しかし，$E[\delta(y)] = \theta$ なる条件が，別の条件に置き換えられ，θ に関して $E\{[\theta - \delta(y)]^2\}$ を一様に最小化するような decision function$\delta(y)$ がもし存在するならば，これは，最小平均二乗誤差推定量 (minimum mean squared-error estimator) と呼ばれる．
¡note¿

Y は統計量であるため，decision rule$\delta(Y)$ も統計量であり，これを $\delta_1(X_1,\dots,X_n)$ と表すとする．このとき，リスク関数は，

$$
\begin{aligned}
R(\theta,\delta_1) &= E\{\mathcal{L}[\theta,\delta_1(X_1,\dots,X_n)]\} \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\mathcal{L}[\theta,\delta_1(x_1,\dots,x_n)]f(x_1;\theta)\cdots f(x_n;\theta)dx_1\cdots dx_n
\end{aligned}
$$

と表現される．

これまで，二乗誤差損失関数 (squared-error loss function, $\mathcal{L}(\theta,\delta)=(\theta-\delta)^2$) と呼ばれる損失関数のみを取り扱ってきた．他にも絶対誤差損失関数 (absolute-error loss function, $\mathcal{L}(\theta,\delta)=|\theta-\delta|$) であったり，

$$
\mathcal{L}(\theta,\delta)=\begin{cases} 0, & |\theta-\delta|\le a \\ b, & |\theta-\delta|>a \end{cases}
$$

で定義される goalpost loss function も利用されている．これは，もし，ギャップが a 以下なら，損失はないとし，そうでないならば b の損失があるとする関数である．加えて，損失関数は，非対称的になりうることにも注意が必要である．例えば，自分が，コンビニの店員だとして，お弁当の受注量を過小に見積もることと，過大に見積もることの損失は一般的に等しくない (過大に見積もって商品が余った場合等)．

最後に，*likelihood principle* を紹介して (そのまま載せて)，本節を終える．

Suppose two different sets of data from possibly two different random experiments lead to respective likelihood ratios, $L_1(\theta)$ and $L_2(\theta)$, that are proportional to each other. These two data sets provide the same information about the parameter θ and **a statistician should obtain the same estimate of θ from either.**

¡note¿

7.2 A sufficient Statistic for a Parameter

統計量 (例えば $Y = u(X_1, \dots, X_n)$ とする) は，データを要約したものであると考えることができる．全てのデータを列挙するより，平均や分散といった統計量でデータの大部分を説明できるなら，効率的である．そのため，統計学者は，観測集合の全てに関連する情報や性質を損なわず，かつ，容易に把握できるようデータを要約する方法を探求している．

定義 7.2.1(十分統計量 (Sufficient Statistic))

$X_1, X_2, \dots, X_n$ を，pdf(もしくは，pmf)$f(x;\theta)$ を持つ分布からのサイズ n の無作為標本とする．$Y_1 = u_1(X_1, X_2, \dots, X_n)$ を，pdf(もしくは，pmf) が $f_{Y_1}(y_1;\theta)$ であるような統計量とする．このとき，Y_1 が θ の十分統計量 (sufficient statistic) であるとは，

$$\frac{f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)}{f_{Y_1}[u_1(x_1, x_2, \dots, x_n);\theta]} = H(x_1, x_2, \dots, x_n)$$

となること，つまり，$Y_1 = y_1$ と条件付けた同時分布が，$\theta \in \Omega$ に依存しないということである．ただし，ここで，$H(x_1, x_2, \dots, x_n)$ は θ に依存しない関数を表す．

より，一般的な状況において，確率変数が独立であることは必要なく，結果として，同一分布であることも必要ではない．ゆえに，より一般的な統計量 $Y_1 = u_1(X_1, X_2, \dots, X_n)$ の十分性の定義は，

$$\frac{f(x_1, x_2, \dots, x_n;\theta)}{f_{Y_1}[u_1(x_1, x_2, \dots, x_n);\theta]} = H(x_1, x_2, \dots, x_n)$$

と表すことができる．ただし，ここで，$f(x_1, x_2, \dots, x_n;\theta)$ は，$X_1, \dots, X_n$ の jpdf(or jpmf) である．

例 7.3. $X_1, \dots, X_n$ を $Gamma(2, \theta)$ からの無作為標本とする．この分布の mgf は $M(t) = (1-\theta t)^{-2}$, $t < 1/\theta$ で表されるため，$Y_1 := \sum_{i=1}^{n} X_i$ の mgf は，

$$\begin{aligned} E[e^{t(X_1+X_2+\cdots+X_n)}] &= E(e^{tX_1})E(e^{tX_2})\cdots E(e^{tX_n}) \\ &= [(1-\theta t)^{-2}]^n = (1-\theta t)^{-2n} \end{aligned}$$

と表すことができる．これは，$Gamma(2n, \theta)$ の mgf であり，この分布の pdf は，

$$f_{Y_1}(y_1;\theta) = \begin{cases} \frac{1}{\Gamma(2n)\theta^{2n}} y_1^{2n-1} e^{-y_1/\theta} & 0 < y_1 < \infty \\ 0 & elsewhere \end{cases}$$

で与えられる．このとき，$Y_1 = y_1$ が与えられたときの条件付き同時分布は，

$$\begin{aligned}\frac{f(x_1, x_2, \ldots, x_n)}{f(y_1)} &= \frac{\left[\frac{x_1^{2-1}e^{-x_1/\theta}}{\Gamma(2)\theta^2}\right]\left[\frac{x_2^{2-1}e^{-x_2/\theta}}{\Gamma(2)\theta^2}\right]\cdots\left[\frac{x_n^{2-1}e^{-x_n/\theta}}{\Gamma(2)\theta^2}\right]}{\frac{(x_1+x_2+\cdots+x_n)^{2n-1}e^{-(x_1+x_2+\cdots+x_n)/\theta}}{\Gamma(2n)\theta^{2n}}} \\ &= \frac{\Gamma(2n)}{\Gamma(2)^n}\frac{x_1x_2\cdots x_n}{(x_1+x_2+\cdots+x_n)^{2n-1}}\end{aligned}$$

となり，これは，θ に依存しないため，$Y_1 = \sum_{i=1}^{n} X_i$ は θ の十分統計量である．

¡note¿

もし，定義に従って，Y_1 が十分統計量であるか否かを判断するなら，Y_1 の pdf を求めることが必要である．しかし，多くの状況でこれはかなり難しい．以下の Neyman の因子分解定理は，Y_1 の pdf を求めることなく，Y_1 が十分統計量であるかどうかを判断できるものである．

定理 7.2.1(Neyman の定理，因子分解定理)

$X_1, X_2, \ldots, X_n$ を $f(x;\theta)$, $\theta \in \Omega$ を pdf，もしくは，pmf として持つ分布からの無作為標本とする．統計量 $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ が十分統計量であるための必要十分条件は，

$$f(x_1, \ldots, x_n; \theta) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta) = k_1[u_1(x_1, \ldots, x_n);\theta]k_2(x_1, \ldots, x_n) \tag{7.2.1}$$

のように，$X_1, \ldots, X_n$ の同時分布が，非負関数 k_1, k_2 を用いて分解できることである．ただし，ここで，$k_2(x_1, \ldots, x_n)$ は θ に依存しない関数である．

証明 7.1. *r.v.* が連続の場合の証明を行う．
(十分性):定理のような分解を仮定し，逆変換 $x_1 = w_1(y_1, y_2, \ldots, y_n), x_2 = w_2(y_1, y_2, \ldots, y_n), \ldots, x_n = w_n(y_1, y_2, \ldots, y_n)$ とヤコビアン J を持つような *1* 対 *1* 対応の変換 $y_1 = u_1(x_1, x_2, \ldots, x_n), y_2 = u_2(x_1, x_2, \ldots, x_n), \ldots, y_n = u_n(x_1, x_2, \ldots, x_n)$ を構成する．このとき，統計量 $Y_1, \ldots, Y_n$ の *pdf* は，

$$g(y_1, y_2, \ldots, y_n; \theta) = k_1(y_1;\theta)k_2(w_1, \ldots, w_n)|J|$$

で与えられる．Y_1 の *pdf* $f_{Y_1}(y_1;\theta)$ は，

$$
\begin{aligned}
f_{Y_1}(y_1;\theta) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(y_1,y_2,\ldots,y_n)dy_2\cdots dy_n \\
&= k_1(y_1;\theta)\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} |J|k_2(w_1,w_2,\ldots,w_n)dy_2\cdots dy_n
\end{aligned}
$$

のように，$y_2,\ldots,y_n$ を積分消去した形で得られる．ここで，k_2 は θ に依存せず，θ はヤコビアンや，積分の極限にも含まれていない．よって，上式の右辺の $(n-1)$ 重積分は y_1 のみの関数である．これを，例えば $m(y_1)$ として表すならば，

$$
f_{Y_1}(y_1;\theta) = k_1(y_1;\theta)m(y_1)
$$

のように表せる．もし，$m(y_1) > 0$ ならば，

$$
k_1[u_1(x_1,x_2,\ldots,x_n);\theta] = \frac{f_{Y_1}[u_1(x_1,\ldots,x_n);\theta]}{m[u_1(x_1,\ldots,x_n)]}
$$

と表現でき，最初に仮定した分解は，

$$
f(x_1;\theta)\cdots f(x_n;\theta) = f_{Y_1}[u_1(x_1,\ldots,x_n);\theta]\frac{k_2(x_1,\ldots,x_n)}{m[u_1(x_1,\ldots,x_n)]}
$$

のように書き換えられる．k_2, m のいずれも θ に依存しないことから，結果として，Y_1 は θ の十分統計量である．

(必要性) : 逆に，もし，Y_1 が θ の十分統計量なら，Y_1 の *pdf(f_{Y_1})* を k_1 として取ることで，定理のような分解を行うことができる．以上から，定理の結果を得る． □

この証明の中の一対一対応変換は必要ではない．詳しくは，Lehmann(1986) を参照．

例 7.4. $X_1,\ldots,X_n \overset{iid}{\sim} N(\theta,\sigma^2)$, $-\infty<\theta<\infty$, $\sigma^2>0$, $\bar{x} := \sum_{i=1}^n x_i/n$ とする．このとき，

$$
\sum_{i=1}^n (x_i-\theta)^2 = \sum_{i=1}^n [(x_i-\bar{x})+(\bar{x}-\theta)]^2 = \sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\theta)^2
$$

と分割できる．ゆえに，$X_1,\ldots,X_n$ の *jpdf* は，

$$
\begin{aligned}
&\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\sum_{i=1}^n (x_i-\theta)^2/2\sigma^2\right] \\
&= \exp[-n(\bar{x}-\theta)^2/2\sigma^2]\left\{\frac{\exp\left[-\sum_{i=1}^n (x_i-\bar{x})^2/2\sigma^2\right]}{(\sigma\sqrt{2\pi})^n}\right\}
\end{aligned}
$$

とできる．右辺の第一因子は $\bar{x}$ によって $x_1,\ldots,x_n$ に依存し，第二因子は θ に依存しないことから，因子分解定理によって，標本平均は θ の十分統計量であると判断できる．

¡note¿

7.3 Properties of a Sufficient Statistic

この節では，MVUE を決定するために，十分性がどのように利用されるかを議論する．$Y_1 = u_1(X_1, X_2, \ldots, X_n)$ が十分統計量であるとし，$Y_2 = g(Y_1)$ も統計量であるとする．ただし，g は一対一対応の関数である．このとき，

$$\begin{aligned} f(x_1;\theta)\cdots f(x_n;\theta) &= k_1[y_1;\theta]k_2(x_1, x_2, \ldots, x_n) \\ &= k_1[g^{-1}(y_2);\theta]k_2(x_1, x_2, \ldots, x_n) \end{aligned}$$

となり，このとき，因子分解定理から，Y_2 も十分統計量である．後に示す定理から十分性は，最良の点推定を導く．

まずは，Theorem 2.3.1 を思い出そう．確率変数 X_1, X_2 について，もし，X_2 の分散が存在するならば，

$$E[X_2] = E[E(X_2|X_1)]$$

であり，

$$Var(X_2) \geq Var[E(X_2|X_1)]$$

が成り立つ．十分統計量の文脈に戻るために，以上の条件にて，X_1 を十分統計量 Y_1，X_2 を θ の不偏推定量 Y_2 に置き換える．このとき，$E(Y_2|y_1) = \phi(y_1)$ を用いて，

$$\theta = E(Y_2) = E(\phi(Y_1))$$

および，

$$Var(Y_2) \geq Var[\phi(Y_1)]$$

を得る．つまり，十分統計量 Y_1 の関数 $\phi(Y_1)$ は θ の不偏推定量であり，同じく θ の不偏推定量 Y_2 よりも小さい分散を持つ．以上の議論を以下に定理の形でまとめる．

定理 7.3.1(Rao-Blackwell の定理)

n を固定された正定数とし，$X_1, \ldots, X_n$ を，pdf(or pmf) として $f(x;\theta), \theta \in \Omega$ を持つ分布からの無作為標本とする．さらに，$Y_1 = u_1(X_1, \ldots, X_n)$ を θ の十分統計量，$Y_2 = u_2(X_1, \ldots, X_n)$($Y_1$ だけの関数に限定しない) を θ の不偏推定量とする．このとき，$E(Y_2|y_1) =: \phi(y_1)$ として，統計量 $\phi(Y_1)$ を定義すると，$\phi(Y_1)$ は θ の十分統計量の関数であり，その上，それ自体 θ の不偏推定量であり．その分散は Y_2 の分散と同じかそれ以下である．

この定理は，パラメータの MVUE の探索は，もし，パラメータの十分統計量が存在するならば，その探索を十分統計量の関数に限定しても良いということを述べている．

以上の定理は，一見すると $\phi(Y_1)$ の探索のためには，不偏推定量 Y_2 を求めることが必要のようだが，必ずしもそうではない．定理 7.3.1(Rao-Blackwell の定理) は，**最良の推定量の探索範囲を十分統計量 Y_1 の関数クラスに絞って良い**ということを示している．その上，以下で示す定理のように，十分統計量と最尤推定量の間には関係がある．

定理 7.3.2

上の定理と同じセッティングに加えて，θ の最尤推定量 $\hat{\theta}$ が一意に存在するとする．このとき，$\hat{\theta}$ は $Y_1 = u_1(X_1, \ldots, X_n)$ の関数である．

証明 7.2. $f_{Y_1}(y_1;\theta)$ を Y_1 の *pdf(or pmf)* とする．このとき，十分性の定義から，尤度関数は

$$\begin{aligned} L(\theta; x_1, \ldots, x_n) &= f(x_1;\theta) \cdots f(x_n;\theta) \\ &= f_{Y_1}(u_1(x_1, \ldots, x_n);\theta) H(x_1, x_2, \ldots, x_n) \\ &(\because Y_1 = u(x_1, \ldots, x_n) \ is\ a\ sufficient\ statistic) \end{aligned}$$

のように，θ に関連する非負関数 f_{Y_1} と，θ に依存しない非負関数 $H(x_1, \ldots, x_n)$ の積に分解できる．L と f_{Y_1} は θ の関数であり，これらを最大化する θ は同一である．$L, f_{Y_1}[u_1(x_1, \ldots, x_n)]$ を最大化する θ は一意に決まり，θ は $u_1(x_1, \ldots, x_n)$ の関数となる．ゆえに，最尤推定量 $\hat{\theta}$ は十分統計量 Y_1 の関数となる． □

4 章と 6 章の議論から，最尤推定量は漸近的に θ の不偏推定量である．よって，まず，十分統計量を見つけ，その後に mle を見つける方法が推奨される．この考えに基づくと，多くの場合に十分統計量の関数である不偏推定量を見つけることができる．この過程はテキストの Example 7.3.1 にて例示されている．

Remark 7.1. 別の統計量，ただし十分統計量ではない Y_3 を用いて，$\Upsilon(y_3) = E[\phi(Y_1)|Y_3 = y_3]$ が与えられる．*Rao-Blackwell* の定理より，$E[\Upsilon(Y_3)] = \theta$ かつ $\Upsilon(Y_3)$ は $\phi(Y_1)$ よりも小さい分散を持つ．結果として，$\Upsilon(Y_3)$ は $\phi(Y_1)$ よりも良い θ の不偏推定量であると考えられる．しかし，これは正しくない．なぜならば，Y_3 は十分統計量ではなく，それゆえに，θ は $Y_1|Y_3 = y_3$ の条件付き分布と**条件付き平均 $\Upsilon(Y_3)$ の中に存在する**からである．$E[\Upsilon(Y_3)] = \theta$ であるが，未知パラメータ θ を含むため，$\Upsilon(Y_3)$ は統計量ではなく，それゆえに，推定値として利用できない．

¡note¿