訂正: $X_1, \ldots, X_n$ の確率密度関数が以下の形を持つときに $X_1, \ldots, X_n$ を n 次元正規確率変数と呼ぶ。

$$\begin{aligned} & f_{X_1 \cdots X_n}(x_1, \ldots, x_n) \\ = \quad & \frac{1}{\sqrt{2^n \pi^n |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{m})^T \Sigma^{-1} (\vec{x} - \vec{m})\right) \end{aligned}$$

推定の例と一般論

まず、最初に例を説明する。画鋲を投げて上を向くか下を向くか観察することを考える。画鋲が上を向く確率 p を推定することを考える。上を向いたときに 1 をとり下を向いたときに 0 をとる確率変数 X を考えると、その分布 P_X は p から完全に決定される。画鋲を n 回投げて上を向いた割合を p の値であると推定する手続きを考えることができる。これは推定の例である。

上の例は推定の例である。次に推定の一般論を述べる。まず何らかの確率分布 (離散型または連続型) があり、その分布に従う確率変数の実現値を n 回観察できるとする。i 回目の観察の結果を表わす確率変数を X_i と名付ける。また、$X_1, \ldots, X_n$ は統計的に独立である、言い替えると

$$P_{X_1 \cdots X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} P_{X_i}(x_i) = \prod_{i=1}^{n} P_{X_1}(x_i)$$

が成立する。２番目の等号は X_i の周辺分布が X_1 と等しいことによる。何か推定したい量を確率変数の実現値 $x_1, \ldots, x_n$ の関数 $T(x_1, \ldots, x_n)$ として推定する。x_i を**標本**と呼び n を**標本の大きさ** と呼ぶ。画鋲の例では

$$T(x_1, \ldots, x_n) = \frac{x_1 + \cdots + x_n}{n}$$

である。$T(x_1, \ldots, x_n)$ は実数だが $T(X_1, \ldots, X_n)$ は確率変数である。
$T(X_1, \ldots, X_n)$ のような確率変数 $X_1, \ldots, X_n$ の関数を**統計量**とよび推定に用いる統計量を**推定量**とよぶ。推定量 (これは確率変数である) の実現値を**推定値**と呼ぶ。
画鋲の例のようにパラメータを１つの値として推定する手法を**点推定**とよび、パラメータが含まれる区間を推定する手法を**区間推定**と呼ぶ。

パラメトリック推定

複数の有限個の実パラメータで記述される確率分布をまず仮定する。
例: 画鋲の例, $N(m, \sigma^2)$, $[a, b]$ に値を取る連続一様確率分布等

これらの分布を記述するパラメータを**母数**と呼ぶ。このように分布の形をある程度仮定して、分布を記述する母数や母数の関数を推定する手法をパラメトリック推定と呼ぶ。当面分布が有限個の母数 θ で記述されるパラメトリックな場合を説明する。θ は複数の実数の組であり、θ に対応する確率分布を P_θ と書くことにする。θ の取り得る値の集合を Θ と書く。画鋲の例では $\Theta = [0, 1]$ であり、正規分布の族を考えれば $\Theta = \mathbf{R} \times [0, \infty)$ である ($\mathbf{R}$ は実数すべてからなる集合)。
今までは平均を取る操作 E がどのような確率分布について平均を取っているかは明らかであったが、推定を論じるときにはどの分布について平均を取っているか明示する必要がある。P_θ について平均を取る操作を E_θ と書くことにする。

不偏推定量 (教科書220ページ)

推定量にはいろいろなものが考えられるが、複数の推定量の善し悪しを判定して、どれを用いればよいか判断するための基準があると便利である。
推定量 $T(X_1, \ldots, X_n)$ を省略して T_n と書くことにする。すべての $\theta \in \Theta$ とすべての n について $E_\theta[T_n]$ が P_θ から定まる推定したい量 $\omega(\theta)$ (これも θ の関数) に等しくなるときに T を**不偏推定量**と呼ぶ。

例：例えば画鋲の例では、画鋲が上を出る確率が p であるとき

$$
\begin{aligned}
& E_\theta[T_n] \\
= \quad & E_\theta\left[\frac{X_1 + \cdots + X_n}{n}\right] \\
= \quad & \frac{E_\theta[X_1] + \cdots + E_\theta[X_n]}{n} \\
= \quad & p
\end{aligned}
$$

がすべての p と n について成り立つ。従ってこれは不偏推定量になっていることが確認できた。

一致推定量 (教科書221ページ)

また、標本の大きさ n が大きいときにすべての $\theta \in \Theta$ について T_n が $\omega(\theta)$ に近い値になる確率が大きくなって欲しい。これを数式で表現するとすべてのの $\epsilon > 0$ について

$$\lim_{n\to\infty} P_\theta(|T_n - \omega(\theta)| < \epsilon) = 1$$

が成立することで、言い替えるとすべての $\theta \in \Theta$ について T_n が $\omega(\theta)$ に確率収束することである。この条件を満たす T_n を一致推定量と呼ぶ。

例：例えば画鋲の例では、画鋲が上を出る確率が p であるとき $\omega(\theta) = p$ であり、$T_n = (X_1 + \cdots + X_n)/n$ という統計量は n を大きくすると p に確率収束することが、弱大数の法則からわかる。従って T_n は一致推定量である。

一様最小分散不偏推定量 (教科書222ページ)

一つの母数に対する二つの不偏推定量 S_n と T_n が有ったとする。どちらが良いか考えたい。S_n と T_n は確率変数であるが、分散が小さいほうがより真の値に推定値が近くなる確率が大きくなる。従って分散は小さいほうがよい。不偏推定量の中ですべての母数 θ と標本の大きさ n について最小の分散を持つものを**一様最小分散不偏推定量**と呼ぶ。(ここで推定量の分散は n と θ に依存していることに注意)

教科書では、上記の性質を持つ推定量を「有効推定量」と呼んでいるが、有効推定量の定義として一様最小分散不偏推定量で更にもう一つの条件を満足する定義もあり、有効推定量と一様最小分散不偏推定量は異なる概念であるとする文献と同じであるとする文献があるから注意すること。

また、ただ分散を小さくしたいだけならば、常に0になる統計量を考えれば分散は0になるから、分散だけを小さくすることには意味がないことに注意せよ。

最尤推定 (教科書217ページ)

分布 P_θ から標本 $x_1, \ldots, x_n$ が得られたときに

$$P_\theta(x_1, \ldots, x_n)$$

を最大にする θ を推定量とする方式を**最尤推定**と呼び、上記の推定量を**最尤推定量**と呼ぶ。上の式を $x_1, \ldots, x_n$ を固定して θ の関数と見たときに尤度関数とよびその自然対数を対数尤度関数と呼ぶ。ここで P_θ は確率質量関数または確率密度関数を表わすものとする。以下、例として、正規分布の最尤推定量を求め、その不偏性と一致性について考察する。
期待値 m 分散 σ^2 正規分布の確率密度関数は

$$f_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-(x-m)^2/2\sigma^2\}$$

である。ここで分布を決める母数 θ は2つの実数 m と σ^2 の組からなる。独立な確率変数 X_i が上記分布に従うときに実現値 $x_1, \ldots, x_n$ が得られる確率密度は

$$\begin{aligned} & \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\{-(x_i-m)^2/2\sigma^2\} \\ = \quad & \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-m)^2\right) \end{aligned}$$

になる。上記の式を最大にする m と σ の組を推定量とする方式が最尤推定である。上の式から対数尤度関数 $l(m,\sigma^2)$ は

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

である。式の形から対数尤度を最大にする m が一つあることがわかるから、m で偏微分すると

$$\frac{\partial l(m,\sigma^2)}{\partial m} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - m)$$

を得る。この偏導関数は $m = (x_1 + \cdots + x_n)/n$ のときに 0 になるから $\hat{m} = (x_1 + \cdots + x_n)/n$ が尤度を最大にする m である。次に l を σ^2 を変数だと思って偏微分すると

$$\frac{\partial l(m,\sigma^2)}{\partial \sigma^2} = \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - m)^2 - \frac{n}{2\sigma^2}$$

を得る。これを 0 にする σ^2 は

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

である。m を含んでいてこのままでは、標本の関数にならないので m に $\hat{m}$ を代入して

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{m})^2$$

を分散 σ^2 の推定量とする。

正規分布の最尤推定の不偏性・一致性 (教科書221ページ)

前ページで求めた正規分布の母数の推定量 $\hat{m}$ と $\widehat{\sigma^2}$ の不偏性と一致性について考察する。まず、$\hat{m}$ は不偏性と一致性を持つ。その理由は画鋲の例とまったく同じである。

$\widehat{\sigma^2}$ は不偏性を持たない。このことを見るために $n=2$ として $\widehat{\sigma^2}$ を計算すると

$$\begin{aligned}
\widehat{\sigma^2} &= \frac{(x_1 - \frac{x_1+x_2}{2})^2 + (x_2 - \frac{x_1+x_2}{2})^2}{2} \\
&= \frac{(\frac{x_1-x_2}{2})^2 + (\frac{x_2-x_2}{2})^2}{2} \\
&= \frac{(x_1-x_2)^2}{4} \\
&= \frac{x_1^2 - 2x_1x_2 + x_2^2}{4}
\end{aligned}$$

これの平均を取ると

$$\begin{aligned}
& E_\theta\left[\frac{x_1^2 - 2x_1x_2 + x_2^2}{4}\right] \\
= & \frac{E_\theta[x_1^2] - 2E_\theta[x_1x_2] + E_\theta[x_2^2]}{4} \\
= & \frac{2E_\theta[x_1^2] - 2E_\theta[x_1]E_\theta[x_2]}{4}
\end{aligned}$$

$$= \frac{2E_\theta[x_1^2] - 2E_\theta[x_1]^2}{4}$$

最後の式は X_1 の分散の半分 $(= (n-1)/n)$ に等しい。一般に $\widehat{\sigma^2}$ の期待値は真の分散の $(n-1)/n$ 倍になることが、同様の計算で確かめられる。従って $\widehat{\sigma^2}$ は不偏推定量ではない。$\frac{n}{n-1}\widehat{\sigma^2}$ が不偏推定量になっている。

しかし、$\widehat{\sigma^2}$ と $\frac{n}{n-1}\widehat{\sigma^2}$ の両方が一致推定量になっていることが知られている。

ノンパラメトリック推定 (教科書224ページ)

今までは標本が有限個の母数で完全に決定される確率分布から得られていることを仮定して推定を行ったが、確率分布に何の仮定も置かずに行う推定をノンパラメトリック推定と呼ぶ。このときに推定の対象は分布の期待値や分散であることが多い。
正規分布の期待値と分散の不偏推定量であった $\hat{m}$ と $\frac{n}{n-1}\widehat{\sigma^2}$ は実はノンパラメトリックな場合にも、不偏推定量になっている。これは不偏性を証明するときに分布が正規分布である仮定をまったく用いていないため、証明をそのまま流用できることからわかる。

ポアソン分布

0, 1, 2, ... に値をとる離散型確率分布で確率質量関数が

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

で与えられる分布を**ポアソン**分布と呼ぶ。λ は分布を記述する**正の**パラメータ (母数) である。

ポアソン分布は非常に発生確率が低い現象を、数多く観察したときに、その現象が起きた合計が従う分布である。例えば、ポアソン分布の最初の当てはめ例は「プロシア陸軍において１年間に馬に蹴られて死んだ兵士の数」がある。ある兵士を考えると彼が１年間に馬に蹴られて死ぬ確率はとても少ないが、プロシア陸軍全体では兵士の数は多い。

母数 λ を持つポアソン分布に従う確率変数の期待値は λ であることは容易にわかる (教科書 115 ページ参照)。一生懸命計算すると分散も λ であることがわかる。

演習問題

原則として授業終了までに答案を提出して下さい。終らない場合は授業が行われた日の夜までに南３号館３１１号室前の箱に演習答案を提出して下さい。

λ を母数に持つポアソン分布の λ の最尤推定を考える。

71. この分布に従う３つの標本 x_1, x_2, x_3 を得たとする。x_1, x_2, x_3 を得る同時確率を λ, x_1, x_2, x_3 を用いて表わせ。x_1, x_2, x_3 は非負整数であると仮定してよい。

72. x_1, x_2, x_3 を定数と見て λ を変数と見たときに直前の問題で答えた式を最大化する λ を x_1, x_2, x_3 を用いて表わせ。これを T_3 とする。

73. T_3 は λ の推定量である。T_3 が不偏推定量であるか否か述べよ。

74. 直前の問に対する答えの大雑把な根拠または厳密な証明を記述せよ。厳密な証明のほうが点数は高くなる。

75. 標本が n 個あるときの λ の最尤推定量を $x_1, \ldots,$ x_n の式として表現し、最尤推定量が一致推定量であるか否か述べよ。

76. 直前の問で最尤推定量が一致推定量であるか否かの答えの大雑把な根拠または厳密な証明を記述せよ。

厳密な証明のほうが点数は高くなる。

77. 授業のわかりにくい点や要望・感想を書いて下さい。また前回の答案で×になったところで、理由がわからないところは申し出て下さい。