正則条件下における
最尤推定量の漸近正規性

渡辺澄夫
東京工業大学

問題

実数に値をとる確率変数 X が確率密度関数 $p(x|w_0)$ を持つとする。ここで w_0 はある実数。
確率変数 $X_1, X_2, \cdots, X_n$ は独立で X と同じ確率密度を持つとする。実数 w の関数(対数尤度関数を (−n) で割ったもの)

$$L(w) = -(1/n) \sum_{i=1}^{n} \log p(X_i|w)$$

を最小にする w を w^* と書いて**最尤推定量**という。
どのような条件のもとで、ある A>0 が存在して

「 $n^{1/2}(w^*-w_0) \rightarrow N(0,1/A)$ (分布収束) 」 が成り立つか。

この命題が成り立つとき「**最尤推定量は漸近正規性を持つ**」という。ここで N(0,1/A)は平均0分散1/Aの正規分布を表す。

漸近正規性 ⇒ 一致性

「 $n^{1/2}(w^* - w_0)$ → $N(0,A)$（分布収束）」 が成り立つならば

「（$w^* - w_0$） → 0（分布収束）」が成り立つ。定数への分布収束が成り立つとき確率収束が成り立つから

「（$w^* - w_0$） → 0（確率収束）」が成り立つ。

したがって、最尤推定量の漸近正規性が成り立つならば

最尤推定量の一致性も成り立っているということがわかった。

しかし証明ではまず一致性を示し、その後に漸近正規性を示すことになる。ここでは一致性は仮定する。

主定理

$(\partial/\partial w)$ を ∂_w と表記する。

関数 $f(x,w) = -\log p(x|w)$ とする。条件(1)–(5)を仮定する。

(1) w^* は w_0 に確率収束する（最尤推定量の一致性）。

(2) $\int (\partial_w)^2 p(x|w)\, dx = 0.$

(3) $f(x,w)$ は w で3回微分可能.

(4) $A \equiv \int \{\partial_w f(x, w_0)\}^2 p(x|w_0)\, dx > 0.$（フィッシャー情報量）

(5) 次の条件を満たす $d>0$ が存在する。$k=0,1,2,3$ について

$T_k(x) \equiv \sup_{|w-w_0|<d} |(\partial_w)^k f(x,w)|$ とおくと $\int T_k(x) p(x|w_0) dx < \infty.$

定理. 「$n^{1/2}(w^*-w_0) \to N(0,1/A)$（分布収束）」が成立

表記と例

定理が成り立つことを $w^* = w_0 + n^{-1/2} N(0, 1/A) + o_p(n^{-1/2})$

のように表記することがある。

例。$a>0$ を固定する。

確率モデルを $p(x|w) = (2\pi a)^{-1/2} \exp(-(x-w)^2/(2a))$ とする。

データ n 個が $p(x|w_0)$ から独立に発生しているとする。

このとき $f(x,w_0) = (x-w_0)^2/(2a) + (1/2)\log(2\pi a)$ であるから

$$A = \int \{ (\partial_w)^2 f(x,w_0) \} p(x|w_0)\, dx = 1/a.$$

従って 「$n^{1/2}(w^*-w_0) \to N(0,a)$ (分布収束)」が成り立つ。

準備(1)

補題0. 一般に「 $Z_n \to Z$ (分布収束), $W_n \to$定数$\neq 0$ (確率収束) ならば $Z_n/W_n \to Z/$定数 (分布収束)」が成り立つ。

補題1. $E_x[\partial_w f(X,w_0)] = 0$

証明: $E_x[f(X,w)] = -E_x[\log p(X|w)]$

$$= D(p(x|w_0)||p(x|w)) - E_x[\log p(X|w_0)]$$

だから $E_x[f(X,w)]$ は $w=w_0$ で最小値をとる。従って

$\partial_w E_x[f(X,w)] = 0.$ 条件(5) より (∂_w) と $E_x[\ \ ]$ の順序は交換できるので補題1が証明できた。

準備(2)

補題2.　次の式が成り立つ。

$$\int \{ \partial_w f(x,w_0) \}^2 p(x|w_0)\, dx = \int \{ (\partial_w)^2 f(x,w_0) \} p(x|w_0)\, dx = A$$

証明: ∂_w を ’ であらわす。定義より $f(x,w) = -\log p(x|w)$.

$$(\partial_w)^2 f(x,w) = \{ -\log p \}'' = (-(p'/p))'$$

$$= -p''/p + (p'/p)^2 = -p''/p + ((-\log p)')^2$$

条件(2) より $\int (p''/p)\, p\, dx = \int p'' dx = 0.$

したがって補題2が証明できた。

準備(3)

定義より $L(w) = (1/n) \sum_{i=1}^{n} f(X_i,w)$ である。

補題3 $L''(w_0) \to A$ (確率収束)

証明: $L(w)$ の定義から

$$L''(w_0) = (1/n) \sum_{i=1}^{n} (\partial_w)^2 f(X_i,w_0).$$

条件(5)と大数の法則から確率収束

$$L''(w_0) \to \int \{ (\partial_w)^2 f(x,w_0) \} p(x|w_0)\, dx$$

が成り立つが、これは補題2より A である。補題3が

証明できた。

準備(4)

$L(w) = (1/n) \sum_{i=1}^{n} f(X_i,w)$ である。

補題4 $n^{1/2}L'(w_0) \to N(0,A)$（分布収束）

証明： $L(w)$の定義から $L'(w_0) = (1/n) \sum_{i=1}^{n} \partial_w f(X_i,w_0)$ であるが、補題1よりこの平均値は0である。従って中心極限定理から $n^{1/2}L'(w_0) \to N(0,\sigma^2)$（分布収束）が成り立つ。ここで分散$\sigma^2$は

$$\sigma^2 = \int (\partial_w f(X,w_0))^2 \, p(x|w_0)dx$$

であるが、これは A である。補題4が証明できた。

定理の証明

最尤推定量 w^* は $L(w)$ を最小にするから $L'(w^*)=0$

平均値の定理から、ある $w^\#$ が存在して $|w^\# - w_0| < |w^* - w_0|$

かつ　$L'(w_0) + (w^* - w_0) L''(w^\#) = 0.$

したがって $w^* - w_0 = -\{L'(w_0)\} / L''(w^\#).$

ゆえに $n^{1/2}(w^* - w_0) = -\{n^{1/2} L'(w_0)\} / L''(w^\#).$

右辺の分子は補題4より $N(0,A)$ に分布収束する。

また分母は次ページのようにして $(A>0)$ に確率収束する

ことを証明することができる。

定理の証明(つづき)

不等式を分けてから条件(5)を適用すると

$|L''(w^{\#})-A| \leqq |L''(w^{\#})-L''(w_0)| + |L''(w_0) - A|$

$\leqq |w^{\#}-w_0|\ (1/n)\sum_i \sup_{|w-w_0|<d} |(\partial_w)^3 f(X_i,w)| + |L''(w_0) - A|$

$\leqq |w^{*}-w_0|\ (1/n)\sum_i T_3(X_i) + |L''(w_0) - A|$

となる。条件(1)より$|w^{*}-w_0| \to 0$ (確率収束), 条件(5)と大数の法則から $(1/n)\sum_i T_3(X_i) \to E[T_3(X)]$ (確率収束), 補題3から $|L''(w_0) - A| \to 0$ (確率収束)。従って $L''(w^{\#}) \to A$(確率収束)。

補題0から $n^{1/2}(w^{*}-w_0)$ は分布 $N(0,A)$ を持つ確率変数を $(-1/A)$倍した確率変数の分布に収束するが、それは $N(0,1/A)$ である。以上より定理が証明できた。