

問題の概要

ラプラス分布に関するパラメータ推定の問題である。正規分布と違って、最尤推定量が中央値になり従って十分統計量にならない例になっている。(4) で漸近正規性を使って良いのかは微妙なところだが、まあ (3) もそこそこ難しいし別に使って楽して良いと思う。

(1)

X_1 の平均は、

$$\begin{aligned}\int_{-\infty}^{\infty} x f(x; \mu) dx &= \int_{-\infty}^{\infty} \frac{x}{2} \exp(-|x - \mu|) dx \\ &= \mu + \int_{-\infty}^{\infty} \frac{x - \mu}{2} \exp(-|x - \mu|) dx \\ &= \mu\end{aligned}$$

X_1 の分散は、

$$\begin{aligned}\int_{-\infty}^{\infty} (x - \mu)^2 f(x; \mu) dx &= \int_{-\infty}^{\infty} \frac{1}{2} (x - \mu)^2 \exp(-|x - \mu|) dx \\ &= \int_0^{\infty} x^2 \exp(-x) dx \\ &= [(-x^2 - 2x - 2) \exp(-x)]_0^{\infty} \\ &= 2\end{aligned}$$

X_1, \dots, X_{2N+1} が独立同一に $f(x; \mu)$ に従うことから、サンプル平均の平均は、

$$\frac{1}{2N+1} \sum_{i=1}^{2N+1} \mu = \mu$$

サンプル平均の分散は、

$$\frac{1}{(2N+1)^2} \sum_{i=1}^{2N+1} 2 = \frac{2}{2N+1}$$

(2)

対数尤度 $l(\mu)$ は、

$$\begin{aligned}l(\mu) &= \sum_{i=1}^{2N+1} \log(f(X_i; \mu)) \\ &= - \sum_{i=1}^{2N+1} |\mu - X_i| - (2N+1) \log(2)\end{aligned}$$

よって、 X_1, \dots, X_{2N+1} を小さい順に並べたものを $X_{(1)}, \dots, X_{(2N+1)}$ とすると、 $\mu < X_{(N+1)}$ で増加、 $\mu > X_{(N+1)}$ で減少するので、 $\hat{\mu} = X_{(N+1)}$ すなわち中央値のとき尤度は最大になる。

(3)

$\hat{\mu} = X_{(N+1)} \leq x \Leftrightarrow X_1, \dots, X_{2N+1}$ のうち $N+1$ コ以上が x 以下

であるので、

$$P(\hat{\mu} \leq x) = \sum_{k=N+1}^{2N+1} \binom{2N+1}{k} F(x; \mu)^k (1 - F(x; \mu))^{2N+1-k}$$

よって $\hat{\mu}$ の確率密度関数は、

$$\begin{aligned} h(x; \mu) &= \frac{d}{dx} P(\hat{\mu} \leq x) \\ &= f(x; \mu) \sum_{k=N+1}^{2N+1} \binom{2N+1}{k} (k(1 - F(x; \mu)) - (2N+1-k)F(x; \mu)) F(x; \mu)^{k-1} (1 - F(x; \mu))^{2N-k} \\ &= (2N+1)f(x; \mu) \sum_{k=N+1}^{2N+1} \left(\binom{2N}{k-1} F(x; \mu)^{k-1} (1 - F(x; \mu))^{2N+1-k} - \binom{2N}{k} F(x; \mu)^k (1 - F(x; \mu))^{2N-k} \right) \\ &= (2N+1) \binom{2N}{N} F(x; \mu)^N (1 - F(x; \mu))^N \end{aligned}$$

最後の等号は総和の中身が添字を跨いで打ち消し合うことによる。

(3) 別解

X_1, \dots, X_{2N+1} の確率密度関数を $\hat{\mu}$ 以外で周辺化することを考える。

$$\begin{aligned} \{\hat{\mu} = x\} &= \bigcup_{\sigma \in \mathfrak{S}_{2N+1}} \{X_{\sigma(1)} \leq \dots \leq X_{\sigma(N)} \leq X_{\sigma(N+1)} = x \leq X_{\sigma(N+2)} \leq \dots \leq X_{\sigma(2N+1)}, X_{\sigma(N)} = x\} \\ \{X_1, \dots, X_N \leq x\} &= \bigcup_{\sigma \in \mathfrak{S}_N} \{X_{\sigma(1)} \leq \dots \leq X_{\sigma(N)} \leq x\} \\ \{X_{N+2}, \dots, X_{2N+1} \leq x\} &= \bigcup_{\sigma \in \mathfrak{S}_N} \{X_{N+1+\sigma(1)} \leq \dots \leq X_{N+1+\sigma(N)} \leq x\} \end{aligned}$$

と、 X_1, \dots, X_{2N+1} の独立同一性に注意して、

$$\begin{aligned} h(x; \mu) &= \int_{\{\hat{\mu}=x\}} \prod_{i=1}^{2N+1} f(x_i; \mu) dx_i \\ &= (2N+1)! f(x; \mu) \left(\int_{\{x_1 \leq \dots \leq x_N \leq x\}} \prod_{i=1}^N f(x_i; \mu) dx_i \right) \left(\int_{\{x \leq x_{N+2} \leq \dots \leq x_{2N+1}\}} \prod_{i=N+2}^{2N+1} f(x_i; \mu) dx_i \right) \\ &= \frac{(2N+1)!}{(N!)^2} f(x; \mu) \left(\int_{-\infty}^x f(\xi; \mu) d\xi \right)^N \left(\int_x^{\infty} f(\xi; \mu) d\xi \right)^N \\ &= (2N+1) \binom{2N}{N} f(x; \mu) F(x; \mu)^N (1 - F(x; \mu))^N \end{aligned}$$

(4)

X_1 の Fisher 情報量は、

$$\begin{aligned}
\int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \mu} \log f(x; \mu) \right)^2 f(x; \mu) dx &= \int_{-\infty}^{\infty} (\operatorname{sgn}(x - \mu))^2 f(x; \mu) dx \\
&= \int_{-\infty}^{\infty} f(x; \mu) dx \\
&= 1
\end{aligned}$$

最尤推定量の漸近正規性により、漸近分散は 1 に収束する。

(4) 別解

最尤推定量の漸近正規性の成立には、 $f(x; \mu)$ の μ に関する良い微分可能性を要求することが多い。物の本では、“良い条件のもと”と誤魔化されていたり、それなりに強い条件が要求されていたりする。例えば、久保川達也先生の「現代数理統計学の基礎」¹では、確率密度関数 $f(x; \mu)$ の μ に関する 3 回微分可能性が条件の一つとして含まれている。

しかしながらこの問題の確率密度関数は $x = \mu$ において微分不可能である。² よって、この問いで漸近正規性を解答の根拠として用いるには、このような確率密度関数に従う確率変数の最尤推定量の漸近正規性の成立を知っている必要がある。前述の通りこれは物の本には書いてないことが多いので、実際にこの問題を例として追ってみる。

大まかな方針としては、積分のラプラス近似法を用いることによる。ラプラス近似法とは、パラメータ M に依存する積分³

$$I_M = \int_a^b g(x) e^{Mf(x)} dx \quad (1)$$

の $M \rightarrow \infty$ での漸近展開を求める方法のことである。ラプラス法では、大雑把には f が最大になる点近傍で f を上に凸な二次関数によって近似する。 $M \rightarrow \infty$ では f が最大になる点近傍の寄与が大きいのだろうという発想である。

今求めたい積分は、

$$\begin{aligned}
(2N+1)V_{2N+1} &= (2N+1)^2 \binom{2N}{N} \int_{-\infty}^{\infty} (x - \mu)^2 f(x; \mu) F(x; \mu)^N (1 - F(x; \mu))^N dx \\
&= \frac{(2N+1)^2}{2^{2N}} \binom{2N}{N} \int_0^1 (G(z))^2 e^{N(\log(z) + \log(1-z))} dz
\end{aligned}$$

ただし、 $z = F(x; \mu)$ と変数変換しており、 $G(z)$ は、

$$F(x; \mu) = \begin{cases} 1 - e^{-(x-\mu)} & (x \geq \mu) \\ \frac{e^{x-\mu}}{2} & (x < \mu) \end{cases}$$

の $x - \mu$ に関する逆関数で、

$$G(z) = \begin{cases} \log(2z) & (0 < z \leq \frac{1}{2}) \\ -\log(2(1-z)) & (\frac{1}{2} < z < 1) \end{cases}$$

¹2017, 共立出版

²一点でしか微分不可能でないの、超関数の微分とか導入すればもっと一般的に良いことが言えるかもしれない。

³ f は問題文中のものとは違う一般の積分できるような関数とする。

である。⁴ また、関数 $\log(z) + \log(1-z)$ は $z = \frac{1}{2}$ で最大値をとることに注意する。
 f は C^3 級、 g は C^2 級として、 f が $x = c$ ($a < c < b$) で最大値をとる場合、式 1 の漸近展開は、 $\alpha = -\frac{Mf''(c)}{2}$ において、

$$I_M = e^{Mf(c)} \left(g(c) \sqrt{\frac{\pi}{\alpha}} + \frac{g''(c)}{4} \sqrt{\frac{\pi}{\alpha^3}} + \frac{g'(c)f'''(c)M}{8} \sqrt{\frac{\pi}{\alpha^5}} + O(M^{-\frac{5}{2}}) \right)$$

である。証明は気持ちだけならテイラー展開でできる。

したがって、 $f(x) = \log(x) + \log(1-x)$ 、 $g(x) = G(x)^2$ として計算すれば良い。

$$f\left(\frac{1}{2}\right) = -2\log 2, f''\left(\frac{1}{2}\right) = -8, f'''\left(\frac{1}{2}\right) = 0$$

$$g\left(\frac{1}{2}\right) = 0, g'\left(\frac{1}{2}\right) = 0, \quad g''\left(\frac{1}{2}\right) = 8$$

であることを用いれば良い。スターリングの公式 $\lim_{M \rightarrow \infty} \frac{M!}{\sqrt{2\pi M} \left(\frac{M}{e}\right)^M} = 1$ にも注意して、

$$\begin{aligned} \lim_{N \rightarrow \infty} (2N+1)V_{2N+1} &= \lim_{N \rightarrow \infty} (2N+1)^2 \binom{2N}{N} 2^{-2N+1} \sqrt{\frac{\pi}{64N^3}} \\ &= \lim_{N \rightarrow \infty} \frac{(2N+1)^2}{(2N)^2} \\ &= 1 \end{aligned}$$

ここでは、変数変換によってラプラス法に出てくる f の特異性を g に押し付けることで楽に計算したが、変数変換せずに計算するには積分区間を分けて計算すれば良い。
 この例を通して一般化ができそうではあるがここではやらない。

(5)

サンプル平均、最尤推定量ともに一致性や不偏性、漸近正規性など良い性質を持つ。サンプル平均は十分統計量であるのに対し最尤推定量は十分統計量ではないため、サンプル平均の方がパラメータに関する情報を失っていないという点でサンプル平均の方が良い推定量であると考えられる。一方で、サンプル平均の漸近分散は 2 であり、Cramer・Rao の不等式の下限を達成していないという点で最尤推定量の方が良い推定量であるとも言える。

⁴変数変換しなくても漸近展開を求めることはできるが、こちらの方が簡単になるのでこうしている。