一般化経験尤度法とそれを用いた推定関数の選択規準

46216 西村 友昭
指導教員 駒木 文保 助教授

2006 年 2 月 7 日

概要

一般化経験尤度法は分布の汎関数として表されるパラメータを，推定関数を用いてノンパラメトリックに推定を行うための手法のクラスであり，既存のいくつかの手法を含むことが知られている．一方，経験尤度法や指数ティルティング法は，Kullback-Leibler(KL) ダイバージェンスを用いたダイバージェンス最小化法として導かれる．そこで，本研究では一般化経験尤度法とダイバージェンス最小化法の関係を調べることを一つ目の目的とする．また，応用上は推定関数として複数の候補が得られていることがある．そこで，次元が最大の不偏性を持つ推定関数を選ぶという一致性を持つ選択規準を考え，一般化経験尤度法の枠組みで具体的な選択規準を提案することを本研究の二つ目の目的とする．

1 一般化経験尤度法

x_i $(i=1,\dots,n)$ は独立にある分布 F_0 に従う確率変数 x の観測値であるとし，分布 F_0 の汎関数として書ける p 次元のパラメータ θ_0 に興味があるとする．パラメータ θ_0 と分布 F_0 の関係は q 次元推定関数 $g(x,\theta)$ を用いて，$\mathrm{E}[g(x,\theta_0)]=0$ と表される．一般化経験尤度法 (Smith [4]) は $q\geq p$ のときに推定関数を用いてノンパラメトリックに θ_0 を推定する手法で，その推定量は凹関数 $\rho(v)$ を用いて，

$$\tilde{\theta}=\arg\min_{\theta}\ \max_{\lambda}\ \frac{1}{n}\sum_{i=1}^{n}\rho(\lambda^{\mathrm{T}}g(x_i,\theta)) \tag{1}$$

で与えられる．ただし，$\rho(v)$ の定義は $v=0$ を内点として含む区間であり，適当な微分可能性のもとで $\rho_{(k)}(v):=\partial^k\rho(v)/\partial v^k$ として一般性を失わず，

$$\rho(0)=0,\ \rho_{(1)}(0)=\rho_{(2)}(0)=-1 \tag{2}$$

であることを仮定する．$\rho(v)=\log(1-v),\rho(v)=1-\mathrm{e}^v,\rho(v)=-v^2/2-v$ としたとき，それぞれ経験尤度法，指数ティルティング法，連続更新一般化積率法に一致する．適当な仮定のもとで，推定量の漸近的な性質として次の定理が成り立つ．

定理 1.

$$\sqrt{n}(\tilde{\theta}-\theta_0)\xrightarrow{\mathrm{d}}\mathrm{N}(0,V),\ \sqrt{n}(\tilde{\lambda}-0)\xrightarrow{\mathrm{d}}\mathrm{N}(0,W), \tag{3}$$

$$2\sum_{i=1}^{n}\rho(\tilde{\lambda}^{\mathrm{T}}g(x_i,\tilde{\theta}))\xrightarrow{\mathrm{d}}\chi^2_{q-p} \tag{4}$$

が成り立つ．ただし，

$$\frac{1}{n}\sum_{i=1}^{n}\rho(\tilde{\lambda}^{\mathrm{T}}g(x_i,\tilde{\theta}))=\min_{\theta}\max_{\lambda}\frac{1}{n}\sum_{i=1}^{n}\rho(\lambda^{\mathrm{T}}g(x_i,\theta)),$$
$$\Omega=\mathrm{E}\left[g(x,\theta_0)g^{\mathrm{T}}(x,\theta_0)\right],\ G=\mathrm{E}\left[\partial g(z,\theta_0)/\partial\theta^{\mathrm{T}}\right],$$
$$V=(G^{\mathrm{T}}\Omega^{-1}G)^{-1},\ W=\Omega^{-1}-\Omega^{-1}GVG^{\mathrm{T}}\Omega^{-1}$$

である．

2 ダイバージェンス最小化法との双対性

一般化経験尤度法をダイバージェンス最小化法の双対問題として捉える場合，分布の確率の和が 1 になるという制約の取り扱いが問題となる．しかし，実は一般化経験尤度法は確率でない測度まで考察の範囲を拡げたダイバージェンス最小化法のラグランジュ双対問題になっている．まず，測度 P,Q に対して拡張された f-ダイバージェンスは

$$\bar{D}(P,Q)=\int p(x)\bar{f}(q(x)/p(x))\mathrm{d}\nu(x)$$

で定義される．ただし，$\bar{f}(u)$ は $\bar{f}(1)=\bar{f}_{(1)}(1)=0$ を満たす凸関数である．拡張された規格化されたダイバージェンス最小化法は，経験分布と同じサポートを持つ測度を考え，推定関数の不偏性という制約の下で経験分布からこの測度への拡張された f-ダイバージェンスを最小化することで得られる．つまり，凸最適化問題

$$\min.\ \frac{1}{n}\sum_{i=1}^{n}\bar{f}(nw_i)\quad \mathrm{s.\,t.}\ \sum_{i=1}^{n}w_ig(x_i,\theta)=0$$

の最適値を $\bar{d}_f(\theta)$ として，$\tilde{\theta}=\arg\min_{\theta}\bar{d}_f(\theta)$ を推定量とする．上の最適化問題のラグランジュ双対問題を考え

ると

$$\max.\ -\frac{1}{n}\sum_{i=1}^{n}\bar{f}^{*}(\lambda^{\mathrm{T}}g(x_i,\theta))$$

となる．ただし，$\bar{f}^{*}(v)$ は $\bar{f}(u)$ の共役関数 $\bar{f}^{*}(v)=\sup_u\{uv-\bar{f}(u)\}$ である．よって，$\rho(v)=-\bar{f}^{*}(v)$ とすれば一般化経験尤度法は拡張されたダイバージェンス最小化法のラグランジュ双対問題になっていることがわかる．また，$\rho(v)$ についての規格化条件 (2) は $\bar{f}(u)$ では

$$\bar{f}(1)=\bar{f}_{(1)}(1)=0,\ \bar{f}_{(2)}(1)=1 \tag{5}$$

と書ける．前の二つの条件はダイバージェンスであるための条件であった．三つ目の条件は簡単な計算により，ダイバージェンスによって導かれる計量を Fisher 情報量に一致させるためのものであることを確認できる．ところで，f-ダイバージェンスの一種である α-ダイバージェンスについて，任意の確率分布 P,Q と任意の $c_p,c_q>0$ について，

$$\bar{D}_{(\alpha)}(c_pP,c_qQ)=\mathrm{const.}+c_p^{\frac{1-\alpha}{2}}c_q^{\frac{1+\alpha}{2}}D_{(\alpha)}(P,Q)$$

という関係が成り立つので，α-ダイバージェンスの最小化は確率として規格化されていてもいなくても等価であるとわかる．従って，α-ダイバージェンスを用いた場合は確率分布に対するダイバージェンス最小化法も一般化経験尤度法に含まれることがわかる．なお，この事実は一般化経験尤度法と確率分布に対するダイバージェンス最小化法の一次最適性条件を比べることにより，Newey and Smith [3] によって示されている．

3　推定関数の選択規準

今，推定関数の候補を $\{g^{\gamma}(x,\theta)\,|\,\gamma\in\mathit{\Gamma}\}$ とする．ただし，$\mathit{\Gamma}$ は有限集合とし，これらの推定関数に対して $\mathrm{E}[g^{\gamma}(x,\theta_0)]=0$ となる θ_0 が唯一つ存在するか，あるいは任意の θ に対して $\mathrm{E}[g^{\gamma}(x,\theta)]\neq 0$ であるかのどちらかが成り立つものとする．つまり，$\mathrm{E}[g^{\gamma}(x,\theta_0)]=0$ となる θ_0 の値が二つ以上存在する場合は考えないことにする．$g^{\gamma}(x,\theta)$, $(\gamma\in\mathit{\Gamma})$ のうち $\mathrm{E}[g^{\gamma}(x,\theta_0)]=0$ となる θ_0 が唯一つ存在する γ の集合を $\mathit{\Gamma}_0$ とし，$\mathit{\Gamma}_0$ の要素のうち推定関数の次元 $q_{\gamma}\,(\geq p)$ が最大となる γ の集合を $M\mathit{\Gamma}_0$ とおく．ここで，次のような選択規準を考える．

$$\mathrm{EFSC}(\gamma)=T_n(\gamma)+h(q_{\gamma})\kappa_n.$$

ただし，$T_n(\gamma)$, $h(q_{\gamma})$, κ_n は次の性質を満たすものとする．

仮定 2.　1. $T_n(\gamma)$ は $x_i\,(i=1,\ldots,n)$ の統計量であり，以下を満たす．(a) $\gamma\in\mathit{\Gamma}_0$ のとき $T_n(\gamma)=\mathrm{O_p}(1)$ である．(b) $\gamma\notin\mathit{\Gamma}_0$ のとき $T_n(\gamma)/n\xrightarrow{\mathrm{p}}c>0$ である．2. $h(.)$ は単調減少である．3. $\kappa_n=\mathrm{o}(n)$ かつ $n\to\infty$ のとき $\kappa_n\to\infty$ である．

この選択規準は Andrews [1] が一般化積率法の枠組みで提案した選択規準を一般化したものである．この選択規準に対して次の定理が成り立つ．

定理 3.　$\hat{\gamma}=\arg\min_{\gamma\in\mathit{\Gamma}}\mathrm{EFSC}(\gamma)$ とすると，仮定 2 のもとで，$\lim_{n\to\infty}\Pr(\hat{\gamma}\in M\mathit{\Gamma}_0)=1$ となる．

ここで，$h(q_{\gamma})\kappa_n$ としては BIC タイプとして $h(q_{\gamma})\kappa_n=(p-q_{\gamma})\log n$ HQIC タイプとして $h(q_{\gamma})\kappa_n=(p-q_{\gamma})\log\log n$ などが考えられる．また，$T_n(\gamma)$ としては一般化経験尤度法の枠組みでは $T_n(\gamma)=2\sum_i\rho(\tilde{\lambda}^{\mathrm{T}}g(x_i,\tilde{\theta}))$ や $T_n(\gamma)=n||\tilde{\lambda}||^2$ などが適当な条件のもとで上の仮定を満たすことが確かめられる．なお，$T_n(\gamma)=2\sum_i\rho(\tilde{\lambda}^{\mathrm{T}}g(x_i,\tilde{\theta}))$ とした選択規準は Hong et al. [2] によってすでに提案されていた．

4　結論と今後の課題

一般化経験尤度法は一般の測度にまで考察を拡げたダイバージェンス最小化法のラグランジュ双対問題として表されることを示した．その結果，一般化経験尤度法の直観的な理解が可能となった．さらに，推定関数として複数の候補があるとき，どのような推定関数を用いるべきかという問題に対して，次元が最大でかつ不偏性を持つ推定関数を選ぶという一致性を持つ選択規準を与え，一般化経験尤度法の枠組みで具体的な選択規準を提案した．今後の課題としては，有限標本におけるパラメータの推定の良さという観点から今回提案した選択規準とは別の選択規準を提案する必要があると考えられる．

参考文献

[1] Andrews, D. W. K. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica*, **67**, 543-564.

[2] Hong, H., Preston, G. and Shum, M. (2003). Generalized empirical likelihood-based model selection criteria for moment condition models. *Econometric Theory*, **19**, 923-943.

[3] Newey, W. K. and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood. *Econometrica*, **72**, 219-255.

[4] Smith, R. J. (1997). Alternative semiparametric likelihood approaches to generalized method of moments estimation. *Economics Journal*, **22**, 300-325.