水域における効率的な水質調査のための最適採水時点の決定手法

Optimization of water-sampling times for efficient water quality investigation

○ 前田 滋哉 · 河地 利彦

○ Shigeya Maeda and Toshihiko Kawachi

1. はじめに

河川や湖沼のような水域の特定地点において，水中での窒素，リン等の濃度を調べるため，現地観測や採水後の室内分析がなされている．この観測・採水頻度は水質データの収集目的により，例えば1月ごと，1週間ごと，1日ごと，3時間ごとのように異なる．水質データの取得には通常多大な労力や費用がかかるため，目的に応じた情報量が得られる程度に観測・採水回数を抑えるよう，その時間的タイミング（以下,「採水時点」と呼ぶ）を計画することが合理的と考えられる．しかし，このような効率を重視した最適採水時点の決定法に関する研究はあまり行われていないようである．Harmancioglu (1984)[1], Harmancioglu *et al.* (1992)[2] は，それぞれ河川のアンモニア態窒素濃度と浮遊物質濃度のデータを用い，情報エントロピーに基づく最適な採水頻度について論じている．一方，疫学の分野で Cook *et al.* (2008)[3] は個体感染を表す確率モデルを考え，モデルパラメータに関する獲得情報量の最大化を目的とし，少数回（例えば4回以下）の個体数検査を行う時点のベイズ決定を示している．本報では Cook *et al.* (2008)[3] を参照し，水質モデル中のパラメータの効率的な同定のため，採水時点を最適化するベイズ決定について考察する．

2. 採水時点の決定問題

水域のある地点において，特定の期間中に m 回採水し，水質分析するだけの予算があるとする．このとき，採水する時間的タイミングを最適に決定する問題，すなわち最適スケジューリング問題を考える．

得られる水質データを $\mathbf{y} = (y_1, y_2, \cdots, y_m)$, 水質モデル中の未知パラメータを θ，θ の事前確率分布を $\pi(\theta)$ とする．ここで，添字 $1, 2, \cdots, m$ は採水時点番号である．決定変数は，採水時点の時刻 T_i を成分に持つベクトル $\mathbf{T} = (T_1, T_2, \cdots, T_m)$ とする．このときベイズ決定は，水質モデルとそれに含まれるパラメータの事前情報に基づき，効用関数

$$u(\mathbf{T}, \mathbf{y}, \theta^*) \tag{1}$$

の期待値

$$E_{\theta^*,\mathbf{y}}[u(\mathbf{T}, \mathbf{y}, \theta^*)] \tag{2}$$

を最大にするような採水時刻ベクトル $\mathbf{T}$ を決定することと表現できる．ここで θ^* = パラメータ θ の真の値である．得られた最適採水時点のすべては実際の採水を行う前に決められるという意味で，これは静的な最適計画問題である．

3. 水質モデルと事後確率分布の推測

採水地点の時刻 t におけるある水質項目の濃度を $C(t)$ とする．濃度の実現値の集合 $\mathcal{C} = \{C(t) : t > 0\}$ の要素の内，観測されるのが $\mathbf{y} = \{C(T_i) : i = 1, 2, \cdots, m\}$ である．水質モデルとして，時刻 t での濃度 $C(t)$ が未知のパラメータ θ を持つある確率分布に従うと仮定する．これに対し，適切な未知パラメータの事前確率分布 $\pi(\theta)$ を想定する．このとき，m 回の採水後の θ の事後確率分布について，ベイズの定理より

$$\pi(\theta|\mathbf{y}, \mathbf{T}) \propto L(\theta)\pi(\theta) \tag{3}$$

が成り立つ．ここで $\pi(\theta|\mathbf{y}, \mathbf{T})$ = 事後確率分布; $L(\theta)$ = 尤度関数であり，θ, $\mathbf{T}$ が与えられたときの $\mathbf{y}$ の確率分布 $f(\mathbf{y}|\theta, \mathbf{T})$ に等しい．

4. 効用関数の設定

効用関数として，採水により得られる情報量から採水・水質分析にかかると予想される総費用を引いたものが考えられる．しかし，総費用の明示は困難であるため，ここでは最大採水回数が m という制約の下での獲得情報量のみを考える．獲得情報量は m 回の採水により消滅した θ についての不確実性と考えることができるため，採水前後の θ の確率分布の情報エントロピーの差で表現できる．したがって，事後確

京都大学大学院農学研究科　Graduate School of Agricultural Science, Kyoto University

キーワード：水質，最適スケジューリング，ベイズ決定，情報量，効用関数

率分布から事前確率分布への Kullback-Leibler divergence を用いることにより，効用関数[3]は次のように表せる．

$$u(\mathbf{T},\mathbf{y}) = E_{\theta|\mathbf{y}}[\log\pi(\theta|\mathbf{y},\mathbf{T}) - \log\pi(\theta)] \quad (4)$$

$$= \int \pi(\theta|\mathbf{y},\mathbf{T}) \log\frac{\pi(\theta|\mathbf{y},\mathbf{T})}{\pi(\theta)} d\theta \quad (5)$$

この効用関数は，不確実に変動する水質 $\mathcal{C}$ と未知パラメータの真値 θ^* の関数であるため，確率変数である．期待効用関数は

$$E_{\mathcal{C},\theta^*}[u(\mathbf{T},\mathbf{y})] = \int u(\mathbf{T},\mathbf{y}) f(\mathcal{C}|\theta^*)\pi(\theta^*) d\mathcal{C} d\theta^* \quad (6)$$

$$= U(\mathbf{T}) \quad (7)$$

となる．これを最大にする $\mathbf{T}$ が求めたい最適な採水時刻ベクトルである．

5. シミュレーションに基づく最適計画

期待効用（式（7））を求めるには $\mathcal{C}$, θ^* について積分する必要があるが，これを直接行うのは困難である．これに対し，モンテカルロ法を用いて最適な決定を導くことがしばしばなされている．[3] モンテカルロ法を適用する場合，期待効用関数を

$$U(\mathbf{T}) \approx \frac{1}{M}\sum_{i=1}^{M} u(\mathbf{T},\mathbf{y}_i) \quad (8)$$

と近似する．ここで，M は十分大きな自然数であり，i はサンプル番号である．これは $u(\mathbf{T},\mathbf{y})$ の期待値を標本平均として近似できることを意味する．ここで式（8）右辺の評価に際しては，サンプル $(\theta_i^*, \mathcal{C}_i)$, $i = 1,2,\cdots M$ を，θ_i^* と $\mathcal{C}_i$ がそれぞれ確率分布 $\pi(\theta^*)$, $f(\mathcal{C}|\theta^*)$ に従うように生成する．サンプルが互いに独立ならば，大数の弱法則によりサンプル数 M が大きくなるほどこの近似は精度を増す．一般にサンプルは互いに独立ではないが，これに対応したサンプリング法のひとつとしてマルコフ連鎖モンテカルロ (MCMC) 法を用いる．[4]

さらに実際的な計算法として，定常分布を用いた MCMC 法[5]の適用を考えることができる．定常確率分布 $h(\cdot)$ を，$\mathbf{T}$ の周辺分布が期待効用関数 $U(\mathbf{T})$ に比例するように定義する．このとき

$$h(\mathbf{T},\mathcal{C},\theta^*) \propto u(\mathbf{T},\mathbf{y}) f(\mathcal{C}|\theta^*)\pi(\theta^*) \quad (9)$$

であり，

$$h(\mathbf{T}) \propto \int u(\mathbf{T},\mathbf{y}) f(\mathcal{C}|\theta^*)\pi(\theta^*) d\mathcal{C} d\theta^* = U(\mathbf{T}) \quad (10)$$

となる．すなわち，$U(\mathbf{T})$ の代わりにサンプリングが容易な関数 $h(\cdot)$ を導入し，これを最大化する $\mathbf{T}$ を MCMC 法で導くことで，$U(\mathbf{T})$ を最大にする採水時刻ベクトル $\mathbf{T}$ が得られると期待できる．

6. おわりに

水質調査において費用や労力の制約がある場合に，獲得情報量を最大にするような採水時点を決定する問題を，ベイズ決定の枠組みを用いて示した．滋賀県高島市今津町の河川，水路，内湖の3ケ所において，2006年～2008年の灌漑期に1日1回以上の頻度で採取された水質データが利用可能である．これを用いて，今後は具体的に水質モデルの決定，効用関数の設定，最適計画を導く効率的計算法の検討を考えていく予定である．

引用文献

[1] Harmancioglu, N.B. (1984): Entropy concept as used in determination of optimal sampling intervals, Proceedings of Hydrosoft '84, International Conference on Hydraulic Engineering Software, Portoroz, Yugoslavia, pp.6-99-6-110. [2] Harmancioglu, N.B., Alpaslan, N. and Singh, V.P. (1992): Application of the entropy concept in design of water quality monitoring networks, in Singh and Fiorentino (Eds.), Entropy and Energy Dissipation in Water Resources, Kluwer Academic Publishers, pp. 283-302. [3] Cook, A.R., Gibson, G.J. and Gilligan, C.A. (2008): Optimal observation times in experimental epidemic processes, Biometrics, 64, pp.860-868. [4] Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996): Introducing Markov chain Monte Carlo, in Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, pp.1-19. [5] Müller, P. (1999): Simulation-based optimal design, in Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M (Eds.), Bayesian Statistics, 6, pp.459-474.