確率過程における最適時点サンプリング

多　賀　保　志

(1966年9月受付)

Optimum Time Sampling in Stochastic Processes

Yasushi TAGA

In approximating a random variable Y by the linear combination of observations (X_{t_1}, ……, X_{t_n}) from a stochastic process $\{X_t,\ t\in T\}$, it is natural to consider how to select time points $(t_1, \cdots\cdots, t_n)$ optimully for a fixed n so that the expected mean square error $E\{|Y-a_0-\sum_{i=1}^{n} a_i X_{t_i}|^2\}$ is to be minimized. Since the expected mean square error is represented as a function of covariances between $(Y,\ X_{t_1}, \cdots\cdots, X_{t_n})$, the optimum selection of time points $(t_1,\cdots\cdots,t_n)$ can be obtained based on the informations about them. When T is discrete and finite, the problem is regarded as a sort of integer programming, and the solution can be obtained by investigating all possible combinations of $(t_1,\cdots\cdots, t_n)$. When T is continuous, informations of covariance functions are needed to solve the problem under the assumption that the process is weakly stationary. In the special case when the process is weakly stationary and has Markov property in the wide sense, covariance function $C(\tau)=E\{X(t+\tau)X(t)\}$ is represented as $\sigma^2 e^{-\lambda\tau}$(see [I]) where $E\{X(t)\}=0$ and $E\{[X(t)]^2\{=\sigma^2$. Then the optimum time sampling can be made analyticaly, and the solution is obtained in the case when $Y=\int_0^1 X(t)dt$.

Institute of Statistical Mathematics

§1 序

$\{X_t,\ t\in T\}$ はある確率過程，Y はある時刻 t までには観測しえない確率変数とするとき，予め定められた n ケの時点 $t_1<t_2<\cdots\cdots t_n(<t)$ における観測値 $(X_{t_1}\cdots\cdots X_{t_n})$ の1次式で Y の最適近似をえようという問題は線形予測とよばれ，実用的にもしばしば現われるものである．しかし，$\{X_t\}$ が定常でない場合やパラメーター空間が離散的な場合には，いままでの線形予測の方法は適用できない．また，$\{X_t\}$ が弱定常の場合でも，共分散関数

$$C(\tau)=E\{X(t+\tau)X(t)\},\ (E\{X(t)\}=0 \text{ とする})$$

が簡単な関数形で表わされるときには，それを用いて平均２乗誤差 $E\{|Y-\sum_{i=1}^{n} a_i X_{t_i}|^2\}$ を最小にするような最適時点（$t_1,\cdots\cdots,t_n$）を求めることができる．とくに，$\{X_t\}$ が広義マルコフ過程かつ $Y=\int_0^1 X(t)dt$ なる場合について，最適時点を実際に求めてみた．

§2 *T* が離散的な場合

確率過程を $\{X_i,\ i=1,\cdots\cdots,N\}$，共分散を $C_{00}=E\{Y^2\}, C_{0i}=E\{YX_i\}$，$C_{ij}=E\{X_iX_j\}$ と表わす．（ただし $E\{Y\}=0,\ E\{X_i\}=0, i=1,\ \cdots\cdots,N$）とするすると，

（１） $S=E\{|Y-\sum_{j=1}^{n} a_j X_{i_j}|^2\}=C_{00}-2\sum_{i=1}^{n} a_j C_{0i_j}+\sum_{j,k=1}^{n} a_j a_k C_{0i_j} C_{0i_k}$

であるから，これを最小ならしめる（$a_1^*,\cdots\cdots,a_n^*$）は

（２） $\sum_{j=1}^{n} a_j C_{i_j i_k}=C_{0i_k},\ i=1,\cdots\cdots,n$

の根として求められる．（$a_1^*,\cdots\cdots,a_n^*$）を (1) に入れた式を S^* とすれば，

（３） $S^*_{(i_1,\cdots\cdots,i_n)}=C_{00}-\sum_{j=1}^{n} a_j^* C_{0i_j}=C_{00}-\sum_{j,k=1}^{n} a_j^* a_k^* C_{i_j i_k}$

となるが，さらに（$i_1,\cdots\cdots,i_n$）を許される範囲で動かすことによって，

（４） $S^{**}=\min_{(i_1,\cdots\cdots,i_n)} S^*_{(i_1,\cdots\cdots,i_n)}$

を attain する（$i_1^*\cdots\cdots,i_n^*$）を求めることができる．これは整数計画法（integer programming）の一種で，これ以上簡単に（$i_1^*,\cdots\cdots,i_n^*$）を求める方法は今までのところ見当らない．この問題は，Y も $\{X_i\}$ もともに多次元の場合に拡張することができる．ただし (1) の代りに $S=\sum_h W_h S_h$ を最小にするようにする）

実際問題への応用としては，$\{X_i\}$ を荷主単位とみたときの日出荷量，Y を年間出荷量（または年末などの繁忙期における月出荷量）としたとき，何日かのデータにもとづいて Y を最良近似したい，というような場合に適用できる．

§3 *T* が連続的な場合

$$C_{00}=E\{Y^2\},\ C_{0t}=E\{YX_t\},\ C_{st}=E\{X_tX_s\}$$

とおくと，前節と同様に (1)～(3) が成り立つ．（ただし $i_1,\cdots\cdots,i_n$ を $t_1,,\cdots\cdots t_n$ でおきかえる） しかし，こんどは T が連続的であるから，

$$S^{**}=\min_{(t_1,\cdots\cdots,t_n)} S^*_{(t_1,\cdots\cdots,t_n)}$$

を求めるに際しては

（５） $\dfrac{\partial S^*_{(t_1,\cdots\cdots,t_n)}}{\partial t_j}=0,\ j=1,\cdots\cdots,n$

の根として（$t_1^*,\cdots\cdots,t_n^*$）を求めることになる．したがって，共分散 $\{C_{st}\}$ の関数形が既知で t について偏微分可能でないと，解を求めることはむつかしい．とくに $\{X_t\}$ が弱定常であれば，$C_{st}=f(t-s)$（$t-s$ の関数）と表わされ，さらに $\{X_t\}$ が広義マルコフ過程であれば，

（６） $C_{st}=f(t-s)=\sigma^2 e^{-\lambda|t-s|},\ \lambda\geqq 0$

$$\sigma^2=E\{X^2{}_t\},\ E\{X_t\}=0$$

となるから，解を求めることが容易になる（[1] をみよ）．

［例］ $\{X_t,\ -\infty<t<\infty\}$ は弱定常な広義マルコフ過程で，$E\{X_t\}=0,\ E\{X_t^2\}=1$，とする．このとき，$Y=\int_0^1 X(t)dt$ を $\sum_{j=1}^{n} a_j X_{t_j}$ で最良近似したい．

$$S^*_{(t_1,\cdots\cdots,t_n)}=b-\sum_{j=1}^{n} a_j^* u_j=b-\boldsymbol{u}'\boldsymbol{C}^{-1}\boldsymbol{u},$$

$$ここで \quad b=\int_0^1\int_0^1 e^{-\lambda|t-s|}dsdt$$

$$u_j=\int_0^1 e^{-\lambda|t-t_j|}dt,$$

$$\boldsymbol{u}'=(u_1,\cdots\cdots,u_n),$$

$$\boldsymbol{C}=(C_{t_it_j}),\ C_{t_it_j}=e^{-\lambda|t_i-t_j|}$$

となるから，一般性を失うことなく，$\lambda=1$ とおいて，$S^{**}=\min\limits_{(t_1,\cdots\cdots,t_n)} S^*(t_1,\cdots\cdots,t_n)$ を attain する $(t_1^*,\cdots\cdots,t_n^*)$ および $(a_1^*,\cdots\cdots,a_n^*)$ を求めてみた．

1) $n=1$ のとき

$$t_1^*=\frac{1}{2},\quad a_1^*=2(1-e^{-\frac{1}{2}})=0.787$$

$$S^{**}=2e^{-1}-4(1-e^{-\frac{1}{2}})^2=0.117$$

2) $n=2$ のとき

対称性より $t_1+t_2=1$ となる．t_1^* は

$$y^3+e^{-1}y-2e^{-1}=0,\quad y=e^{-t_1}$$

の根として求められる．

$$t_1^*=0.264,\quad t_2^*=0.736$$

$$b_1^*=b_2^*=0.464,\quad S^{**}=0.037$$

n が3以上になると4次以上の代数方程式をとくことが必要になるが，数値計算によって根を求めることは可能である．

統計数理研究所

参考文献

[1] Dood, J.L., Stochastic Processes, John Wiley, (1953).