Innovation Theory の歴史

Si Si
愛知県立大学　情報科学部

概 要

確率過程の詳しい確率論的な構造をしらべるには、それと同じ情報量を持つ独立確率変数の列で、かつ時間についての因果性を考慮したもの、すなわち Innovation を用いるのが有効な手段である。確率場についても同様である。

Innovation の研究は 1950 年頃から種々の方法で行われてきており、現在も重要な研究課題である。その古いアイデイアを見つけて将来の研究に役立たせたい。

1　はじめに

確率解析の systematic な理論の出発点を明らかにして、これまでの解析の方法によって random complex system の研究を進めようというのは、orthodox な方法であろう。その出発点として決めるのに一番自然なものが **innovation** である。Innovation という言葉は 1950 年代から使われてきたようであるが、この言葉は使わなくても実質的なものはそれよりずっと古くから確率論の中に現われていた。現在の確率論やランダムな複雑系の理論に直接つながるようなはっきりした形で出てきたのは、P. Lévy や N. Wiener による二つの流れと考えてよいであろう。

1) Lévy によるものは、今一般的な立場で言えば、**要素還元主義** (Reductionism) 的な考えによるものであり、

2) Wiener によるよる方法は**非線型予測** (nonlinear prediction) のための最も基本的な量として出てきたものである。

Innovation は、大まかに言えば、時系列 X_n の場合でいうと、**独立**な確率変数の列 Y_n で、各 n について $\{X_k, k \le n\}$ の持つ情報（事象) と $\{Y_k, k \le n\}$ の持つ情報とが等しくなるような時系列である。

確率過程 $X(t)$ のときは、各時点で独立な確率過程は、実は**超過程**になってしまう。その確率分布が時間の shift で不変、すなわち定常であるとしよう。時間的一様性は自然な仮定である。その超過程の各要素（すなわち確率素子）を集めていけば（時間について積分する)、加法過程が得られる。それが確率連続であることを仮定すれば、よく知られた Lévy 過程になる。これを $Z(t)$ と書こう。この過程 $Z(t)$ の Lévy 分解は

$$Z(t) = mt + \sigma B(t) + X(t) \ ,$$

となる。ここで m は定数 で以後無視する。 $B(t)$ はブラウン運動、$X(t)$ は複合ポアソン過程であり、ジャンプの高さ u がいろいろな場合のポアソン過程 $X_u(t)$ の複雑な組み合わせになっている。こうして elemental な加法過程は $B(t)$ と　系 $\{X_u(t), u \in U \subset R\}$ ということができる。innovation ならば、それらの時間微分をとればよい。こうして、ガウス型のホワイトノイズ（単にホワイトノイズという）またはポアソンノイズの関数（実は汎関数）を考えることになる。

この定義は確率場 $X(C)$ のときにも一般化できる。

厳密な innovation の定義は文献 [2], [6] に詳しい。

問題は時系列、あるいは確率過程とか確率場などが与えられたときに、どのような方法で innovation をもとめるか、またそれがどのようにして与えられたランダムな現象の研究に役立つかを考えることである。おもな方向として、1) と 2) の二つがあるが、それらについて、理論の歴史と、それがどんな形で現代数学（現代科学といってもよい）の中で発展してきたかを述べてみよう。

2　P. Lévy の innovation

文献 [11] の第 8 章は時系列について、 innovation （この言葉は使っていないが）の説明から始まっている。時系列　X_n、すなわち

離散パラメターの場合から初めているが、それは次のような方法で innovation を構成しようとしている。すなわち、独立な列 $\{Y_n\}$ で、各 n について

1) Y_{n+1} は $\mathbf{B}_n(X)$ と独立、

2) $\mathbf{B}_n(X) = \mathbf{B}(Y_n, X_k, k \leq n-1)$

となるようなものを求めることである。

文献 [11] には innovation が構成できるための必要条件も示されているが、かなり強い制限であるように思われる。パラメターが離散のときは簡単そうに見えるが、パラメターについての連続性がないので、かえって難しくしている。

しかし M. Rosenblatt などの研究 [17] があり、その他、特別な場合にはいくらかの成果がえられている。

P. Lévy は [13] において、パラメターが連続のときに一つの symbolical formula を出している。確率過程 $X(t)$ について、時刻 t と $t+dt$ との微小区間での変分 $\delta X(t)$ は

$$X(t+dt) - X(t) = \delta X(t) o(dt).$$

をみたすものとして

$$\delta X(t) = \Phi(X(s), s \leq t, Y(t), t, dt).$$

と表されたとしよう。ここで $Y(t)$ は $X(s), s \leq t,$ と独立で、$X(t)$ が $[t, t+dt)$ において得た新しい情報を表す infinitesimal な確率変数であり，Φ はランダムでない関数とする。このとき $Y(t)$ が innovation になる。この方程式は **stochastic infinitesimal equation** と呼ばれる。formal な式ではあるが、suggestion をあたえてくれる。数学的に厳密な定義は [6] で述べている。　Innovation を求めるための困難さは時系列のときとほぼ同様であるが、$X(t)$ のときは t についての連続性を仮定することが多いので，制限が多く少し事情が違ってくる。

Innovation についてこのような setup が出てくるのには、当時の確率過程の研究の動きが影響しているように思われる。すなわち、次節

で述べるガウス過程の研究が始まっていて、自然にinnovationの考え方が出てきていたことがあり、その発展と考えられる。また定常過程の研究研究に対して1950年に Karhunen の standard な論文 [10] が出て linear prediction の一つの研究体系が完成した。そこで次の step として、多くの人がこれからは nonlinear prediction の時代だと叫んでいたのは1950年代である。

単にnonlinear prediction というのは容易であるが,多くの実例をふまえて研究内容の具体的な formulation に向かう地道な努力が必要である。その時期に Lévy の infinitesimal equation があらわれて、研究の guideline を示したのはよい timing であった。

3 ガウス過程の場合

ガウス過程 $X(t), t \geq 0$, が与えられたとき, Reductionism の立場からは、各 t 毎に独立な確率変数を $X(t)$ から構成して、それを変数として元の $X(t)$ を書き表すことが問題 となる。独立な確率変数系をブラウン運動 $B(t)$ の時間微分であるホワイトノイズ $\dot{B}(t)$ にとり

$$X(t) = \int_0^t F(t,u)\dot{B}(u)du \tag{3.1}$$

とあらわすことができれば、よく知られた $\dot{B}(t)$ と non-random な関数 $F(t,u)$ によって $X(t)$ の詳しい性質をしることができる。これが**ガウス過程の表現**である。

一般のガウス過程 $X(t)$ について、表現の話を厳密に進めるため、仮定を明らかにしておく(詳しくは論文 [4] 参照)。まず、平均は 0 とする:

$$E(X(t)) = 0.$$

また $X(t)$'s の張るヒルベルト空間 $M(X)$ は separable であるとする。また initial data はないとしよう。 このとき、innovation が作れて、それは有限または可算無限個の独立なホワイトノイズ $\{\dot{B}_i(t)\}$ で

あり、$X(t)$ は

$$X(t) = \sum_i \int_0^t F_i(t,u)\dot{B}_i(u)du \tag{3.2}$$

とあらわされる。当然 sigma-field について

$$\mathbf{B}_t(X) = \bigvee_i \mathbf{B}(\dot{B}_i)$$

がなりたつ。その意味で **generalized canonical representation** である。

ただ一つの $\dot{B}$ で表現できる場合が好都合なときであり、詳しい性質が知られる。このときの $X(t)$ の表現が **標準表現** (**canonical representation**) である。

標準表現をもつ $X(t)$ について、$F(t,u)$ がなめらかで、$F(t,t)=0$ になることがなければ、その変分が存在して、変分 $\delta X(t)$ は

$$\delta X(t) = dt \int^t \frac{\partial}{\partial t} F(t,u)\dot{B}(u)du + F(t,t)\dot{B}(t)dt,$$

で、第 2 項は $\sqrt{dt}$ の order である。

標準表現の存在を知って、それが具体的に求められるよい例として、多重マルコフ ガウス過程がある。$X(t)$ は N 重マルコフ ガウス過程とし、その標準表現を

$$X(t) = \int_0^t F(t,u)\dot{B}(u)du$$

とする。ここで $F(t,u)$ はグルサー核 (Goursat kernel) で

$$F(t,u) = \sum_1^N f_i(t)g_i(u)$$

となる N 個の t_j にたいして

$$det(f_i(t_j)) \neq 0$$

であり、$\{g_i\}$ は一次独立である。

今共分散 $\Gamma(t,s), t \geq s$, が t の関数として C^N-class と仮定しよう。このとき、N 階の常微分作用素 L_t があって、$\{f_i\}$ は

$$L_t f = 0$$

の基本解系となる。L_t を Frobenius formula で書けば（T. Hida,[4]）

$$L_t = \prod_0^{N-1} \frac{d}{dm_j} \cdot \frac{1}{v_N(t)},$$

ここで $dm_j(t) = v_j(t)dt$ である．いま

$$L_t = \frac{d}{dm_0} L_t^1$$

によって L_t^1 を定義する。$\{v_j\}$ を適当に選んでおけば、$t > s$ のとき

$$L_t^1 E(X(t)|\mathbf{B}_s(X)) = \int_0^s g(u)\dot{B}(u)du,$$

となる g が存在する。右辺は加法過程で、これから容易に $\dot{B}(s)$ が求められる。それは innovation である。

また純非決定的 (purely nondeterministic) な強連続定常ガウス過程の場合は，一般の（ガウス型と限らない）弱定常過程の理論を利用することができる。

純非決定的, 強連続弱定常過程については 前節であげた K. Karhunen の論文 [10] は orthogonal increment process による表現といえる。それは linear theory であり、これをガウス過程の場合にあてはめれば、上の話の特別な場合になる。

Linear process への一般化 (Win Win Htay [23]).

$X(t)$ が linear process とは、次の条件をみたすものをいう。すなわち、任意の時刻 t にたいして、t 以前の値が知られたとき、以後の値は知られた値の linear function とそれらの値とは独立な確率変数との和としてあらわされる。（一般的な定義は P. Lévy [15] 参照。）そのような linear process の典型として次のような $X(t)$ を考えよう。

$$X(t)=\int_0^t F(t,u)\dot{B}(u)du+\int_0^t G(t,u)\dot{P}(u)du.$$

ここで、$\dot{P}$ はポアソン ノイズである。（複合ポアソン ノイズとしてもよい。）ただし、$F(t,u)$ は canonical kernel で， なめらか、かつ $G(t,u)$ は連続で $G(t,t)$ は 0 にならないとする。このとき sample function (path) を見ることにより $\dot{P}(t)$ が、ついで $\dot{B}(t)$ がもとまる。そして, 二つをたしたもの

$$\dot{B}(t)+\dot{P}(t)$$

を innovation としてよい。またベクトル $\{\dot{B}(t),\dot{P}(t)\}$ をとってもよい。

これは linear といっても見本関数 (sample function) について見た場合で，linear な L^2-theory ではない。次節の考え方と同じである。

4 Linear prediction と innovation

N. Wiener は前々から、定常過程の予測の問題に、今の言葉でいえば innovation あるいは標準表現の問題を提起し、いくらかの結果を得ている。その後、Wiener は、著書 [20] の中で、（特に Coding, Decoding の部分を参照）この課題を扱っている。そこまでに到る経過を見よう。

まず linear prediction である。Wiener は 1934 年の著書で Harmonic analysis の立場から確率関数 (random function) を扱っているが、これは定常過程の解析ということで、結果として Linear prediction theory の準備となっている。

Wiener は 1949 年にいたり，定常過程の研究と通信工学の理論との融合をめざした本を書いた ([20]). 私達はその結果だけでなく、その思想に注目したい。基本的な考え方は，定常確率過程 $X(t,\omega), t\in R, \omega\in\Omega$, について、ある ω を決めたときの見本関数 (sample function, path) をとりあげる。$X(t,\omega)=f(t)$ とおく。すなわち、$f(t)$ は考えている偶然現象の一つの観測結果を表す。ある時刻 t まで、この偶然

現象を観測したとき、それ以後の時点における値をよりよく予測しようというのが目標である。これは数式により、次のように表される：有界変分関数 $K(t)$ により時刻 $t+h$ $(h>0)$, における予測値 $\hat{f}(t,h)$ を

$$\hat{f}(t,h) = \int_0^\infty f(t-s)dK(s)$$

で与えようとする。この時、予測の誤差を出来るだけ小さくしたい。

ここで、誤差の測り方が問題になるが、それには２乗平均誤差をとるのが普通である。したがって

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |f(t+h) - \hat{f}(t,h)|^2 dt$$

を最小にするような K を選べばよいことになる。すなわち K を変数とした変分法の計算が必要になる。解析としてのポイントは、考えている定常過程の共分散関数のあつかい、スペクトル密度の factorization の議論である。いわゆる調和解析の話になる。これにより最良予測値が explicit にもとまることになる。

この議論の仮定になることがらできわめて重要な注意がある。予測したものが良いか悪いかを見るために２乗平均誤差を使うといったが、それなら、見本関数よりも $X(t)$ のままで、誤差を直接

$$E\{|X(t+h) - \int_0^\infty X(t-s)dK(s)|^2\}$$

で測ったらよかろう、と思うかもしれない。ここに, E は平均、すなわち $dP(\omega)$ による積分をあらわす（念のため）。たしかに、この方法は線形予測理論として、ヒルベルト空間の方法を使って、大きな成功をおさめた。仮定といったのは、この方法では、平均も共分散関数 $\Gamma(h)$ もわかっているとして計算ができることである。具体的な問題では、それは仮定できない。そこで前に使った式

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t+h)f(t)dt$$

を共分散関数とすることになる。この極限値が存在して ω に無関係であることが必要である。もっと一般に

$$dP\text{ による平均 (相平均) }E = \text{時間平均}$$

となること、すなわち**エルゴード性**がなりたつことが基礎にある。innovation が存在し、remote past が 0 であるとき、この式はなりたつ。

いずれにしても、共分散関数をきめるのは、この方法によることになる。それが前から与えられるような場合は少ない.

なお、ここでは触れないが、線形予測理論については

A.N. Kolmogorov, Interpolation and extrapolation. Bull. Academy of Sci. USSR, ser. Math. 5 (1941), 3-14.
のすぐれた論文 がある。

5 Wiener の非線型予測と innovation

定常過程の線形予測の理論は、残された問題はたくさんあるけれども、一応 setup がすんだので、次は nonlinear prediction theory だという声が高くなって innovation の方法がとりあげられる。 一つには Cramér volume に出た Masani-Wiener の論文 Nonlinear prediction (1959 年) が suggestion になった。

Nonlinear theory といっても、ただ与えられた $X(t)$ の nonlinear function を作るのでは、case by case でしか成功しない。**一般的な setup** のもとに **systematic に議論**することがもとめられた。それには、$X(t)$ を evolutional random complex system とみて、その innovation をもとめ、$X(t)$ を t と innovation (それは独立な確率変数のシステムである) の関数としてあらわせば解析しやすくなると考えられる。予測という応用の問題から始まったことが、Lévy の考えのように、確率過程の innovation の問題としてとりあげられることになったのである。

Wiener は 1958 年の本 [21] で nonlinear system, coding, decoding などの章でこのことを説明している。一方 P. Lévy は 1953 年の California lecture note [13] で **stochastic infinitesimal equation** をあげて、このことを説明して入ることは 2 章で述べたとうりで、数学的には同じような idea から来ていることに注意したい 。

[註] N. Wiener の nonlinear prediction と innovation との自然な関係については Kallianpur の解説 [9] が参考になる。

Innovation $\{Y(t)\}$ が求まると、与えられた $X(t)$ は, remote past がなければ

$$X(t) = \Psi(Y(s), s \leq t, t)$$

と表されよう。

$(\Psi, \{Y(t)\})$ は完全に $X(t)$ の確率論的な構造を決定している。こうして、$\{Y(t)\}$ を変数のシステムとする汎関数解析が始まる。

Innovation が直接求められる例をあげよう。

(1) ガウス過程。これについてはすでに詳しくのべた。

(2) 確率微分方程式できまる場合。$X(t)$ は

$$\delta X(t) = A(X(t))dt + \sigma(X(t))Y(t)dt \tag{5.1}$$

と適当な初期条件できまるとしよう。ここで $Y(t)$ はガウス型または複合ポアソンノイズとする。それらを加えたものでもよい。

(2.a) 方程式

$$\delta X(t) = (-aX(t) + f(t))dt + g(t)Y(t)dt \ ,$$

があたえられたとする. 適当な initial condition のもとで、解が存在して $Y(t)$ は innovation になり、$X(t)$ は

$$X(t) = e^{-at} \int_0^t \{e^{au} f(u) + g(u)Y(u)\} du + C \ .$$

とあらわされる。

(2.b) 上の g を $X(t)$ の 1 次関数として

$$\delta X(t) = -aX(t)dt + (b + X(t))Y_t dt, \quad Y = \dot{B} \ .$$

とすれば $Y(t)$ は innovation となり

$$X(t) = \sum_1^{\infty} \int_{(-\infty,t]^n} \frac{b^n e^{-at}}{n!} e^{a min(u_1, \ldots u_n)} : \dot{B}_{u_1} \ldots \dot{B}_{u_n} : du^n + C \ ,$$

となる。: : は Wick product である.

(3) ガウス過程とポアソン型の linear process の積になっている場合。

このような場合であるかどうかは、特性汎関数によってきめることができる。特性汎関数は、与えられるかまたは、やはりエルゴード性を使って見本関数から求めることになる (cf.[2])。

(4) Martingale に reduce されるとき。[2] による。

[註] 興味あるガウス過程のときでも、stochastic infinitesimal equation がいつでも得られるわけではない。例を挙げよう。

例　fractional order Brownian motion.

P. Lévy は1953年の Berkeley の講義録 [13] で次のような fractional order のブラウン運動を定義した。もちろんガウス過程である。

$$X_\alpha(t) = \frac{1}{\Gamma(\alpha+1)} \int_0^t (t-u)^\alpha \dot{B}(u) du.$$

ただし $\alpha > -1/2$. n を $\alpha + 1/2$ の整数部分とすれば、$X_\alpha(t)$ は n 回微分できて、k 次導関数を $X_\alpha^{(k)}(t)$ と書けば

$$\delta X_\alpha(t) = \sum_1^n X_\alpha^{(k)} \frac{(dt)^k}{k!} + c\dot{B}(t)(dt)^{\alpha+1} + o(dt)^{\alpha+1}.$$

そして、やはり $\dot{B}(t)$ が innovation となり、$\mathbf{B}_t(X) = \mathbf{B}_t(\dot{B}),\ \forall t$ である。

kernel が (t-u) の関数であるため定常過程となることが期待される。そのためには確率積分の下限を $-\infty$ にしなければならないが、積分を収束させるために Mandelbrot 達は次の項を加えた：

$$\frac{1}{\Gamma(\alpha+1)} \int_{-\infty}^0 [(t-u)^\alpha - (-u)^\alpha] \dot{B}(u) du.$$

このときも innovation は $\dot{B}(t)$ である。

もし、定常性だけにこだわるならば fractional order Ornstein Uhlenbeck Brownian motion を定義すればよい。すなわち

$$U_\alpha(t) = \frac{1}{\Gamma(\alpha+1)} \int_{-\infty}^t (t-u)^\alpha e^{-(t-u)} \dot{B}(u) du.$$

6　確率場の場合

確率場については、報告者が確率過程に対する結果を一般化して確率場の innovation を定義し、議論している（[6] 参照）。考え方は stochastic infinitesimal equation を一般化した stochastic variational equation

$$\delta X(C) = \Phi(X(C'), C', (C') \subset (C), Y(C), C, \delta C).$$

によるが、歴史とするにはまだ早いようである。

References

[1] L. Accardi et al, Selected papers of Takeyuki Hida. World Scientific Pub. Co.Ltd. 2001

[2] L. Accardi, T. Hida and Si Si, Innovation approach to some stochastic processes and random fields. Volterra Center Notes. 2002.

[3] I.M. Gel'fand and N. Ya. Vilenkin, Generalizede functions. vol.4.

[4] T. Hida, Canonical representations of Gaussian processes and their applications. Memoires College of Sci. Univ. Kyoto, A. 33 (1960), 109-155.

[5] 飛田武幸，櫃田倍之、ガウス過程、紀伊国屋書店、1976.

[6] T. Hida and Si Si, An innovation approach to random fields. Application of white noise theory. World Scientific Pub. Co. Ltd. 2003.

[7] T. Hida and Si Si, Lévy field as a generator of elemental noises. Proc. Levico Conference on classical and quantum Lévy process. 2003, to appear.

[8] T. Kailath, An innovation approach to least-square estimation Part

I, II, IEEE Transactions on Automatic Control Vol. AC-13, 1968, 646-654, 665-660.

[9] G. Kallianpur, Some ramifications of Wiener's ideas on nonlinear prediction. Norbert Wiener: Collected Works, vol.III, The MIT Press, 1981, 402-424.

[10] K. Karhunen, Uber die Struktur station arer zuf allige Funktionen. Arkiv f. Mat. 1 (1950), 141-160.

[11] P. Lévy, Théorie de l'addition des variables aléatoires. Gauthier-Villars. 1937. 2ème éd.1954.

[12] P. Lévy, Processus stochastiques et mouvement brownien. Gauthier-Villars, 1948; 2ème ed. 1965.

[13] P. Lévy, Random functions: general theory with special reference to Laplacian random functions.I, (1953), 331-388. Univ. of Calif. Pub.

[14] P. Lévy, A Special problem of Brownian motion, and a general theory of Gaussian random functions. Proc. of the 3rd Berkeley Symp. on Math. Statist. and Probability, vol. II, 1955. 133-174.

[15] P. Lévy, Fonction aléatoire à corrélation linéaire. Illinois J. Math. 1 (1957), 217-258.

[16] P. Masani and N. Wiener, Non-linear prediction. Probability and Satatistics, The Harald Cramér volume. ed. U.Grenander, 1959, 190-212. Almqvist & Wiksell, John Wiley & Sons.

[17] M. Rosenblatt, Stationary processes as shifts of functions of independent random variables, J. Math. Mech. 8 (1959), 665-681.

[18] A. N. Shirjaev, Nonlinear filtering of Markov diffusion processes, Trudy Mat. Inst. Steklov, 104(1968), 135-180.

[19] N. Wiener, (with R.E.A.C. Paley), Fourier transforms in the com-

plex domain. Amer. Math. Soc. 1934.

[20] N. Wiener, Extrapolation, interporation, and smoothing of stationary time series. The MIT Press, 1949

[21] N. Wiener, Nonlinear problems in random theory. The MIT Press, 1958.

[22] Nobert Wiener, Special issue of Bull. American Math. Soc. 72 (1966).

[23] Win Win Htay, Note on linear processes, Proc. Meijo Winter School, Jan 2003.

おわり