清水 泰隆 研究室（確率統計解析研究）

確率過程モデリングと統計的推測論
— 統計学で明日は見えますか？—

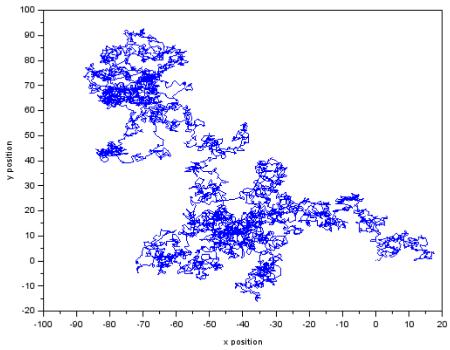
研究室 HP

確率過程による将来予測

時刻 t における株価を S_t と書けば，S_t の値は t が進むにつれて “ランダム” に変わります．このように時間を変数としてランダムに変化する系列 $S = (S_t)_{t\geq 0}$ を**確率過程**といいます．最も基本的な確率過程の一つである**ブラウン運動**は，ある自然現象を数学的にモデル化したもので，これを基に多くのランダムな現象が確率モデルとして記述できるようになります．

代表的な確率過程：Brown 運動

植物学者 R.Brown 博士は，花粉の中から出てくる微粒子が水面上で不規則に動くのを発見しました．始めはなにかの生物が現れたと思ったらしい Brown 博士ですが，後に，これが水分子と微粒子との衝突によって起こる物理現象であることがわかりました．



A.Einstein による定式化 (1905 年・奇跡の年)」：

時刻 t において，位置 x に微粒子が存在する密度 $p(t,x)$ は，

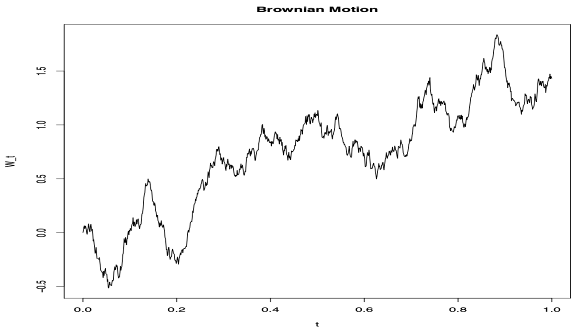
$$\frac{\partial p}{\partial t} = \frac{\kappa}{2}\frac{\partial^2 p}{\partial x^2}, \quad p(0,x) = \xi(x)\ (\text{初期分布})$$
$$\Rightarrow p(t,x) = \int_{\mathbb{R}} \xi(y)\phi(\kappa t, x, y)\, dy$$

ここに，有名な**正規（Gauss）分布** $\phi(t,x,y)$ が現れます：

$$\phi(t,x,y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}}, \quad \Rightarrow \quad \mathcal{N}(x,t) \qquad (1)$$

つまり，最初に密度 $\xi(x)$ のように分布していた粒子の拡散密度は，正規分布に従うように広がっていきます．

この 2 次元の Brown 運動を一方向から射影し，時系列に沿って軌跡を描いてみると，**1 次元 Brown 運動**のパスが描かれます（次図）.



$$\text{確率表現 } B_t \sim \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k, \quad Z_k \sim^d \mathcal{N}(0,t)$$

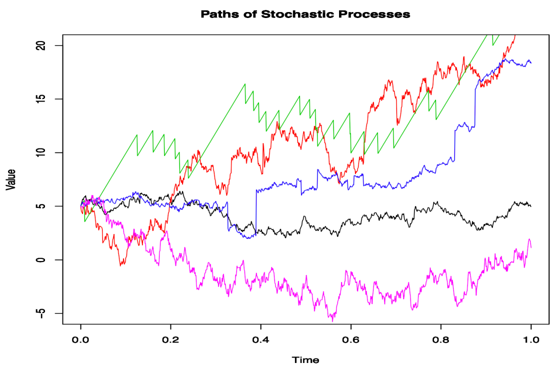
資産過程モデル：確率微分方程式

Brown 運動を用いて表現される代表的な確率過程が確率微分方程式によって記述される拡散過程です．

$$S_t = x + \int_0^t \mu(S_u, \theta_1)\, du + \int_0^t \sigma(S_u, \theta_2)\, dB_u \qquad (2)$$

ここに，$\theta = (\theta_1, \theta_2)$ はパラメータです．

関数 μ, σ やパラメータ θ の値によって，様々なパス（グラフ）がランダムに現れ，これらのグラフの形が株価や企業価値などの資産モデルとして用いられます．



統計的推測論：データによるパラメータ推定

株価のデータ $S_0, S_{t_1}, \ldots, S_{t_n}$ から**疑似尤度解析**によって θ を推定し，モデル (2) を特定することで，将来の株価の予測に利用します．

$$\widehat{\theta}_n := \arg\min_{\theta} \sum_{k=1}^{n} \log\phi\left(\Delta\sigma^2_{k-1}(\theta_2), \Delta^n_k S, \mu_{k-1}(\theta_1)\Delta\right)$$
$$\xrightarrow{p} \theta_0 \quad (\text{真のパラメータ}) \quad n \to \infty.$$

数学が分野を繋ぐ：人口・保険統計への応用

外見上は全く異なる現象であっても，数学的に表現すれば同一のモデルになるところは，応用数学の面白いところです．先程は株価のような金融資産を確率微分方程式でモデリングしましたが，今度は，これを人間の死亡率予測に応用することを考えてみます．

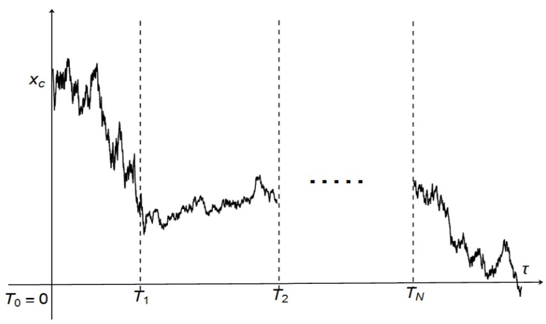
人はなぜ死ぬのか?: 生命エネルギー仮説と SEM

人はなぜ死ぬのか？ということを考えるとき，仮想的に次のような仮説を考えてみましょう．「人は生命エネルギー (Survival Energy) を持って生まれてきて，エネルギーが消滅すると死亡する」このような生命エネルギーの確率モデルを **SEM** (Survival Energy Model) といいます．

ある c 年生まれの人が，初期エネルギー x_c を持って生まれてきたとき，生命エネルギーが次の非斉時拡散過程に従うとします（**ID-SEM**）：

$$X_t^c = x_c + \int_0^t U_c(s;\theta)\,ds + \int_0^t V_c(s;\theta)\,dB_s$$

ただし，B は Brown 運動，$U_c(s,\theta)$ は SE の平均的な増加率，$V_c(s,\theta)$ は SE のぶれの大きさ（分散）を表します．



人の死亡時刻は $\tau_c := \inf\{t > 0 \,|\, X_t^c < 0\}$ と書けて

$$q_c(t) := \mathbb{P}(\tau^c \le t), \quad t > 0$$

を**死亡率関数 (mortality function)** と呼びます．

死亡率関数のパラメトリックモデル

$$U_c(t;\theta_c) = \alpha_c + \beta_c \exp\left(\gamma_c(t-T)\right)\mathbf{1}_{\{t>T\}},$$
$$\kappa_c S(t;\theta_c) = M(t;\theta_c)$$

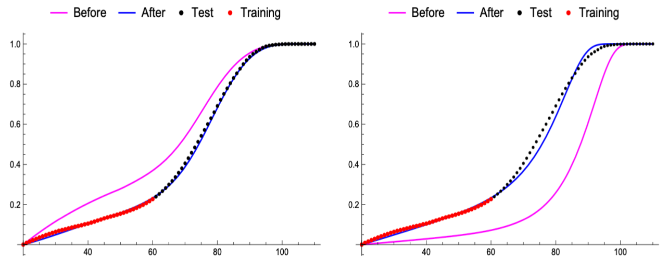
ただし，

$$S(t;\theta) = \int_0^t V_c^2(s;\theta)\,ds, \quad M(t;\theta) = \int_0^t U_c(s;\theta)\,ds$$

とおくと，死亡率関数が以下のように陽な形で得られます：

$$q_c(t;\theta_c) = 1 - \int_0^\infty \left(\frac{1}{\sqrt{\pi S(t;\theta_c)}}\exp\left(-\frac{(z-x_c-M(t;\theta_c))^2}{4S(t;\theta_c)}\right)\right.$$
$$\left. e^{-\kappa_c x_c}\frac{1}{\sqrt{\pi S(t;\theta_c)}}\exp\left(-\frac{(z+x_c-M(t;\theta_c))^2}{4S(t;\theta_c)}\right)\right) dz$$

これを，実際に得られている経験死亡率と**最小２乗法**で合わせることで，パラメータ θ_c を推定し，将来のコホート c' 対して $\theta_{c'}$ を予測します．



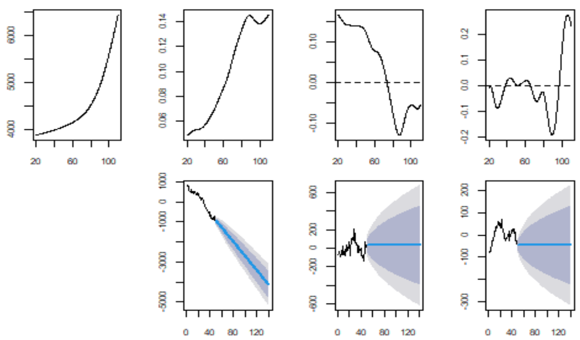
Norway-1900, Male: ID-SEM (left) vs. IG-SEM (right)

ノンパラメトリック法：関数データ解析

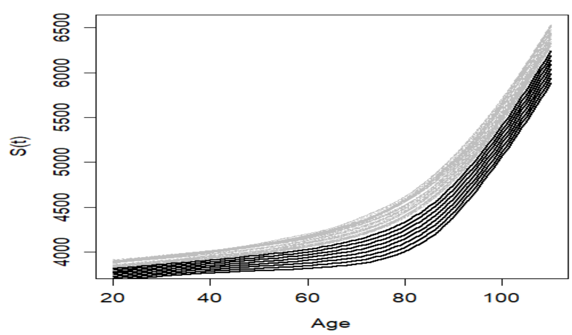
パラメトリックモデルでの特定が難しい場合には，ノンパラメトリックな手法もあります．**関数データ解析**は関数データに対する主成分分析です．時間発展するデータを何本も観測したとき，その関数の将来の形状をダイレクトに予測する統計手法です：

$$S(t) = \widehat{\mu}(t) + \sum_{k=1}^K \widehat{\xi}_{c,k}\widehat{\nu}_k(t), \quad (c_1 \le c \le c_m,\ T_0 \le t \le \omega)$$

ここで，$\widehat{\mu}(t) = \frac{1}{m}\sum_{c=c_1}^{c_m}$（過去データの平均），$\widehat{\nu}_k$ は $(S(T_0),\ldots,S(\omega))$ の標本分散共分散行列の第 k 固有値に対する固有ベクトル．



上段は平均 $\widehat{\mu}(t)$ と固有ベクトル，下段が ξ_k の (時系列的) 予測値とその 95%予測区間（$K=$ 第３主成分まで）．



グレーの曲線は過去データ，黒の曲線が予測値で，これを用いて ID-SEM により将来の死亡率関数を予測できます．

— 早稲田大学理工学術院 基幹理工学部/研究科 応用数理学科/数学応用数理専攻 （清水研究室）—