時系列入門

マーケティング実務
への応用をめざして

基礎知識

二つのモデル

- 物理モデル
  - システムの挙動を記述する状態変数の時間的変動が（微分、差分方程式などの）物理法則から導かれる
  - 関数形は固定、パラメータは微分方程式またはシミュレーションから推定
- 時系列モデル
  - 観測データだけが与えられている場合に、その背後にあるメカニズムを排除した統計的な分析を行う

時系列モデル

- 統計モデル
  - 基本的な変数は観測データ
  - 例: $x_n = f(x_{n-1}, x_{n-2}, x_{n-3}....)$
- 状態空間モデル
  - 観測データから適当な変換を経て得られるベクトル変数Ynに関するモデル
  - 例: $\mathbf{Y}_n = \mathbf{F}(\mathbf{Y}_{n1}, \mathbf{Y}_{n2}, \mathbf{Y}_{n3}....)$
  - ここで、n1、n2、n3は時間通りの順番とは限らない

確率過程の定義

- 時間の経過に伴って変動する確率変数の無限の集合を確率過程と呼ぶ
- 大きさNの時系列 $\{x_n\} \equiv \{x_1, x_2, \ldots x_N\}$ は確率過程 $\{X_n\} \equiv \{X_1, X_2, \ldots X_N\}$ の実現値として与えられる
- 現象に関する結合確率密度関数Pが与えられればそのモーメントからすべての統計量がきまる
- 確率密度関数Pは以下の規格化条件を満たさなければならない

$$\sum_{x_1, x_2, \ldots, x_N} P(x_1, x_2, \ldots, x_N) = 1 \quad \text{離散値}$$

$$\int_{x_1, x_2, \ldots, x_N} P(x_1, x_2, \ldots, x_N)\, dx_1, dx_2, \ldots, dx_N = 1 \quad \text{連続 値}$$

- 次のとき、各確率変数は独立であるが、時系列ではほとんどありえない

$$P(x_1, x_2, \ldots, x_N) = P(x_1)P(x_2)\ldots P(x_N)$$

基本的な統計量

- 1次モーメント（時刻nにおける平均値）

$$\mu_n = E[x_n] \quad (1 \le n \le N)$$

- 2次モーメント（時刻nにおける分散・自己共分散）

$$\mathrm{var}(x_n) = E\left[(x_n - \mu_n)^2\right] \quad (1 \le n \le N)$$
$$\mathrm{cov}(x_n, x_{n+k}) = E\left[(x_n - \mu_n)(x_{n+k} - \mu_{n+k})\right] \quad (1 \le n \le N-k)$$

- kはラグとよばれ、k=0の時、自己共分散は分散に一致する
- 期待値の定義

$$E[q] = \sum_{x_1, x_2, \dots, x_N} qP(x_1, x_2, \dots, x_N) = 1$$

定常性条件

- 時系列では観測値は1時刻に1つだけ与えられるために、時間の関数である前頁の統計量は正確には得られない
- そこで以下の（弱*）定常性条件を課して、少ない観測値から統計量を推定する

*1次と2次のモーメントの条件だけを取り出しているため

$$
\begin{aligned}
&\mu_n = \mu_n = const \\
&\mathrm{var}(x_n) = \gamma(0) = const \\
&\mathrm{cov}(x_n, x_{n+k}) = \gamma(k)
\end{aligned}
$$

平均値・分散は時刻に寄らず一定
自己共分散はラグ*k*のみに依存

$$
\begin{aligned}
&\sigma_x^2 \equiv \mathrm{var}(x_n) \equiv \gamma(0) \\
&\gamma(-k) = \gamma(k) \\
&|\gamma(k)| \le \sigma_x^2
\end{aligned}
$$

k=0の時自己共分散は分散に一致する。自己共分散は分散を超えない。

代表的な定常過程:ガウス過程

- 以下のガウス(正規)分布に従う。

$$P(\xi)=\frac{1}{\sqrt{2\pi\sigma_\xi^2}}\exp\left(-\frac{(\xi-\mu_\xi)^2}{2\sigma_\xi^2}\right)$$

ガウス分布

μ_ξ : 平均、　σ_ξ^2 : 分散

$$\int_{-\infty}^{\infty} P(\xi)d\xi = 1$$

規格化条件

$$\mu_\xi = \int_{-\infty}^{\infty} \xi P(\xi)d\xi$$

$$\sigma_\xi^2 = \int_{-\infty}^{\infty} (\xi-\mu_\xi)^2 P(\xi)d\xi$$

$$\xi_n \sim N\left(\mu_\xi, \sigma_\xi^2\right)$$

「正規過程に従う」

ホワイトノイズ

- N(0,1)に従う互いに独立な系列を正規白色過程あるいは正規ホワイトノイズという。（互いに無相関なデータを並べた時系列のこと）

$$
\mu_{\xi} = 0
$$

$$
\gamma_{\xi}(k) = \delta_k
$$

$$
\delta_k = \begin{cases} 1 & (k = 0) \\ 0 & (k \neq 0) \end{cases}
$$

デルタ関数

自己相関関数

- 自己共分散を分散で正規化したものを自己相関関数（コレログラム）と呼ぶ。

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)(=\sigma_x^2)}$$

$\mu_x = E[x_n]$とすると、

$$\mathrm{cov}(x_n, x_{n+k}) = E[(x_n - \mu_n)(x_{n+k} - \mu_{n+k})] \quad (1 \le n \le N) \quad \text{より}$$

$$\rho(k) = \frac{E[(x_n - \mu_n)(x_{n+k} - \mu_{n+k})]}{\sigma_x^2}$$

$$\rho(0) = 1$$

$$\rho(k) = \rho(-k)$$

$$|\rho(k)| \le 1$$

異なる2種類の時系列の相関

- 異なる2種類の時系列に対する相互共分散および相互相関関数は、

$$\gamma_{xy}(k) = E\left[(x_n - \mu_x)(y_{n+k} - \mu_y)\right]$$

$$\rho_{xy}(k) = \frac{E\left[(x_n - \mu_x)(y_{n+k} - \mu_y)\right]}{\left\{\gamma_x(0)\gamma_y(0)\right\}^{1/2}}$$

$$\mu_x = E[x_n]$$

$$\mu_y = E[y_n]$$

$$\gamma_x(0) = \sigma_x^2$$

$$\gamma_y(0) = \sigma_y^2$$

（中略）

- で、いろいろ中略して…

自己回帰モデル

- Autoregressive model (AR model)

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t, t = 1, \ldots, T$$

$$E(\varepsilon_t) = 0$$

$$E(\varepsilon_t^2) = \sigma^2$$

$$E(\varepsilon_t \varepsilon_{t-\tau}) = 0, \text{for } \tau \neq 0$$

$$y_t \sim AR(p)$$

プロセスの平均は一般 性を失うことなくゼロ と仮定

$$E(y_t) = 0, \forall t$$

ラグパラメータ

$Ly_t = y_{t-1}, L^2 y_t = y_{t-2} \cdots$

$AR(1)$をラグオペレータを用 いて書き直すと、

$y_t = \phi y_{t-1} + \varepsilon_t$

$\leftrightarrow y_t - \phi L y_t = \varepsilon_t$

$\leftrightarrow (1 - \phi L) y_t = \varepsilon_t$

$|\phi| < 1 \leftrightarrow 1 - \phi L = 0$の解が1より大きい

特性多項式と定常性

- AR(p)における定常性を考える

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

$$\leftrightarrow \left(1 - \phi_1 L - \ldots - \phi_p L^p\right) y_t = \varepsilon_t$$

$AR(1)$から類推すると、

$$1 - \phi_1 L - \ldots - \phi_p L^p = 0$$

の解がすべて1より大きいことが定常性の条件

移動平均表現(1)

- AR(1)に後退代入を繰り返す
- 観測地はたまたまt＝1から得られているが、時系列自体は非常に遠い過去から出発していると考える

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, t = 1,\ldots,T$$

$$y_t = \phi(\phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

$$= \phi^2 y_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t$$

$$= \ldots$$

$$= \phi^J y_{t-J} + \sum_{j=0}^{J-1} \phi^j \varepsilon_{t-j}$$

すなわち初期値とホワイトノイズの加重和

$$E(y_t) = E\left(\sum_{j=0}^{J-1} \phi^j \varepsilon_{t-j}\right) + E(\phi^J y_{t-J}) = \phi^J y_{t-J}$$

$|\phi| \geq 1 \rightarrow$ 平均は初期値に依存

$|\phi| < 1 \rightarrow \phi^J y_{t-J} \rightarrow 0 \text{ as } J \rightarrow \infty$ 定常,この時

$$y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}, t = 1,\ldots,T$$

これを無次元の移動平均表現という

移動平均表現(2)

- 定常AR(1)をMA(∞)と表現すると、時系列の値とは、現在までのホワイトノイズの加重和になっていることがわかる。現在から遠い時点のノイズほど影響が小さい。

$$\gamma(0) = E\left(y_t^2\right) = E\left(\left(\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}\right)^2\right)$$

$$= \sum_{j=0}^{\infty} \phi^{2j} E\left(\varepsilon_{t-j}^2\right) = \sigma^2 \sum_{j=0}^{\infty} \phi^{2j}$$

$|\phi| < 1$ならば

$$\gamma(0) = \frac{\sigma^2}{1-\phi^2}$$

これは有限であり時間に依存せず一定

自己回帰移動平均、他

- とりあえず省略
- 他もとりあえず省略

情報量基準

- 赤池(1973)は、統計モデルのよさを、モデルが定める予測分布と真の分布との近さで測ることとし、その近さをカルバックライブラー情報量で評価することを提案。
- 真の分布が未知である場合にはこの情報量は直接計算できないが、対数尤度がその推定値とみなせることを示す。(つまり、最尤法は予測分布の最適化を目指した推定法)
- 推定されたモデルを評価するときに生じる対数尤度のバイアスを以下の情報量基準によって補正

$$AIC = -2(\text{対数尤度}) + 2(\text{パラメータ数})$$

時系列の予測と状態

- Yn-を時刻nまでの観測値、Yn+を時刻n以降の将来の観測値で張られる空間とする。ともに現在の値y_nを含み共通部分を持つが、この共通部分を経由して過去の情報は将来に伝達される。共通部分がy_nだけである場合は時系列はマルコフ性を持つが、一般にはより複雑な構造を持つ。赤池(1974)は時系列の予測においてはこの共通部分が時系列の状態に対応し、その最小次元の決定によって最適なモデルが決定されることを示した。
- 状態の推定の問題とは、観測値y_nに基づいて時刻nの状態x_nの推定を行うことである。時刻jまでの観測値Yjに基づいて状態x_nを推定する場合、以下の3つの場合に分類される

$j < n$:予測

$j = n$:フィルタ

$j > n$:平滑化

観測値$Y_j = \{y_{1,}, y_2, .. y_j\}$が与えられた時、状態 x_nを推定するためには、

x_nの条件付分布 $p(x_n | Y_j)$を求めればよい。

状態空間モデル（1）

- 過去の状態は何らかの形で現在に集約されていると考え、これをシステムの状態としてとらえたうえで、この状態からシステムに関する推定と予測をより効率的に行うことを考える。知りうる状態は時間の推移とともに得られる状態によって変化する。その様子を表すのがシステムの状態空間表現であり、状態空間表現によるモデルを状態空間モデルと呼ぶ。（これに対して、システムへの入力とそこからの出力に注目する統計モデルを入出力モデルと呼ぶ。）
- 一般に状態空間モデルは、時間推移とともに状態の変化する様子をモデル化したシステムモデルと、状態と観測値の関係を表現した観測モデルの二つのモデルからなる
- 特に、線形・ガウス型の状態空間モデルは、このような構造を線形モデルと正規ノイズで近似したものである。

状態空間モデル(2)

- すなわち時系列ynに対して以下のような2つのモデルを合わせて状態空間モデルという。

$$x_n = \mathbf{F_n} x_{n-1} + \mathbf{G_n} v_n \text{(システムモデル)}$$

$$y_n = \mathbf{H_n} x_n + w_n \text{(観測モデル)}$$

ただし、

x_n :状態ベクトル

$v_n \sim N(0, Q_n)$:ホワイトノイズ(システムノイズ)

$w_n \sim N(0, R_n)$:ホワイトノイズ(観測ノイズ)

$\mathbf{F_n}, \mathbf{G_n}, \mathbf{H_n}$:適当な次元のベクトル

- 時系列解析で用いられる多くのモデルは、状態空間モデルの形で表現し、統一的に扱うことができる。仮に観測モデルを時系列ynが観測される仕組みを扱う回帰モデルであると考えると、状態xnはその回帰係数であり、回帰係数が時間的に変化する場合を考えることができる。

カルマンフィルタ

- 線形・ガウス型状態空間モデルの場合には状態の分布 $p(x_n|Y_j)$ は正規分布となるので、状態推定のためには平均ベクトルと分散共分散行列だけを求めればよい。状態 x_n の条件付平均と分散共分散行列をそれぞれ $x_{n|j}, V_{n|j}$ とあらわすと、$p(x_n|Y_j) \sim N(x_{n|j}, V_{n|j})$ となる。
- 一般に、観測値 Y_j が与えられた下での状態 $\{x_1, x_2, \ldots, x_N\}$ の条件つき同時分布を求めるためには多大な計算量を要するが、状態空間モデルの場合、逐次的な計算アルゴリズムによって条件付の周辺分布 $p(x_n|Y_j)$ はデータ数Nに比例する計算量で効率的に計算できる。
- 特に線形・ガウス型状態空間モデルの場合にはカルマンフィルタと呼ばれるアルゴリズムが利用できる。

カルマンフィルタの計算

- 以下の2ステップからなる

[一期先予測]

$$x_{n|n-1} = F_n x_{n-1|n-1}$$

$$V_{n|n-1} = F_n V_{n-1|n-1} F_n^t + G_n Q_n G_n^t$$

ただしQ_nは確率項v_nの分散共分散行列であり、

$$E\left[v_n v_m^{'}\right] = \begin{cases} Q_n : n = m \\ 0 : n \neq m \end{cases}$$

[フィルタ]

$$K_n = V_{n|n-1} H_n^t \left(H_n V_{n|n-1} H_n^t + R_n\right)^{-1}$$

$$x_{n|n} = x_{n|n-1} + K_n \left(y_n - H_n x_{n|n-1}\right)$$

$$V_{n|n} = \left(I - K_n H_n\right) V_{n|n-1}$$

ただしK_nはカルマンゲイン、R_nは確率項w_nの分散共分散行列であり、

$$E\left[w_n w_m^{'}\right] = \begin{cases} R_n : n = m \\ 0 : n \neq m \end{cases}$$

カルマンフィルタ・数学補足(1)

$$V_{n|n-1} = E\left(x_n - x_{n|n-1}\right)^2$$
$$= E\left(F\left(x_{n-1} - x_{n-1|n-1}\right) + Gv_n\right)^2$$
$$= FE\left(x_{n-1} - x_{n-1|n-1}\right)^2 F^{'} + GE\left(v_n^2\right)G^{'}$$
$$= F_n V_{n-1|n-1} F_n^t + G_n Q_n G_n^t$$

$$K_n = V_{n|n-1} H_n^t \left(H_n V_{n|n-1} H_n^t + R_n\right)^{-1}$$

ここで$H_n V_{n|n-1} H_n^t + R_n$は予測誤差の分散

$$x_{n|n} = x_{n|n-1} + K_n\left(y_n - H_n x_{n|n-1}\right)$$

ここで$y_n - H_n x_{n|n-1}$は予測誤差（イノベーション）

カルマンゲインとは、予測誤差分散に占める状態の変動分を示したシグナルノイズ比

$V_{n|n} = (I - K_n H_n) V_{n|n-1}$はフィルタリングによって分散を小さくしようとしている

カルマンフィルタ・数学補足(2)

$$x_{n|n} = x_{n|n-1} + K_n\left(y_n - H_n x_{n|n-1}\right)$$
$$= K_n y_n + \left(I - K_n H_n\right)x_{n|n-1}$$

つまりフィルタ平均は 、予測分布の平均と観 測値の加重和になっている。

$$V_{n|n} = (I - K_n H_n)V_{n|n-1}$$
$$= V_{n|n-1} - K_t H_n V_{n|n-1}$$

つまり、観測値からの情報によって、状態推定の精度が改善されている(分散が小さくなっている。)

固定区間平滑化

- カルマンフィルタは時刻nまでの観測値だけを用いて状態x_nを推定するが、平滑化のアルゴリズムではより多くの観測値を用いて状態推定を行う。したがって、平滑化を行えばフィルタよりも精度の良い状態推定が実現できる。平滑化に関しては、以下のアルゴリズムが一般的。

$$
\begin{aligned}
A_n &= V_{n|n} F_{n+1}^t V_{n+1|n}^{-1} \\
x_{n|N} &= x_{n|n} + A_n \left(x_{n+1|N} - x_{n+1|n} \right) \\
V_{n|N} &= V_{n|n} + A_n \left(V_{n+1|N} - V_{n+1|n} \right) A_n^t
\end{aligned}
$$

- まずカルマンフィルタによって $\left\{ x_{n|n-1}, x_{n|n}, V_{n|n-1}, V_{n|n} \right\}_{n=1,\dots,N}$ を求め、次に上式によって時間的に逆方向に $\left\{ x_{n|N}, V_{n|N} \right\}_{n=N-1,\dots,1}$ を順次計算すればよい

トレンドモデル

- 1階のトレンドモデルは、以下で表される

$$t_n = t_{n-1} + v_n \left(v_n \sim N\left(0, \tau^2\right)\right)$$
$$y_n = t_n + w_n \left(w_n \sim N\left(0, \sigma^2\right)\right)$$

- 2階のトレンドモデルは以下で表される

$$t_n = 2t_{n-1} - t_{n-2} + v_n \left(v_n \sim N\left(0, \tau^2\right)\right)$$
$$y_n = t_n + w_n \left(w_n \sim N\left(0, \sigma^2\right)\right)$$

- 以下の確率差分方程式がシステムモデルの一般式となる

$$\Delta^k t_n = v_n \left(v_n \sim N\left(0, \tau^2\right)\right)$$

またまた数学補足

- 定義

$$t_n = t_{n-1} + v_n$$
$$(1-L)t_n = v_n \ (L\text{はラグオペレータ})$$
$$\Delta t_n = v_n \ (\Delta\text{は階差オペレータ})$$

- 2階のトレンドモデルを例に取る

$$\Delta^2 t_n = v_n$$
$$(1-L)^2 t_n = v_n$$
$$(1-2L+L^2)t_n = v_n$$
$$t_n - 2t_{n-1} + t_{n-2} = v_n$$
$$t_n = 2t_{n-1} - t_{n-2} + v_n$$

季節調整モデル(1)

- トレンド、季節変動、不規則変動は直接には観測できない要素(潜在変数)なので、システム方程式でそのダイナミクスを与えてやるためのモデル

$$y_n = t_n + s_n + w_n \sim N\left(0, \sigma^2\right)$$

周期pの季節変動は、

$$s_n \approx s_{n-p}$$

をみたし（例えば一年前と今の値はほぼ同じ）

$$s_n + \dots + s_{n-p+1} = v_{n2} \sim N\left(0, \tau_2^2\right) = \sum_{j=0}^{p-1} s_{n-j} = v_{n2} \sim N\left(0, \tau_2^2\right)$$

によって表現できる。なぜならば、

$$s_n = s_{n-p} + v_{n2}$$

$$s_n - s_{n-p} = v_{n2}$$

$$\left(1 - B^p\right) s_n = v_{n2}$$ （Bはバックオペレータ）

$$(1 - B)\left(1 + B + \dots + B^{p-1}\right) s_n = v_{n2}$$

*注：ここよくわからない

季節調整モデル(2)

- 4半期のデータに以下の仮定を与えてシステムモデルを構成する

$$\Delta^2 t_n = v_{n1} \sim N\left(0, \tau_1^2\right)$$

$$t_n = 2t_{n-1} - t_{n-2} + v_{n1} \sim N\left(0, \tau_1^2\right)$$

$$s_n = -s_{n-1} - s_{n-2} - s_{n-3} + v_{n2} \sim N\left(0, \tau_2^2\right)$$

$$x_{n-1} = \left(t_{n-1}, t_{n-2}, s_{n-1}, s_{n-2}, s_{n-3}\right)'$$

$$x_n = \left(t_n, t_{n-1}, s_n, s_{n-1}, s_{n-2}\right)'$$

$$F = \begin{pmatrix} 2 & -1 & & & \\ 1 & 0 & & & \\ & & -1 & -1 & -1 \\ & & 1 & 0 & 0 \\ & & 0 & 1 & 0 \end{pmatrix}, G = \begin{pmatrix} 1 & \\ & \\ & 1 \\ & \\ & \end{pmatrix}$$

$$v_n = \left(v_{n1}, v_{n2}\right)'$$

$$Var\left(v_n\right) = Q = diag\left(\tau_1^2, \tau_2^2\right)$$

$$x_n = Fx_{n-1} + Gv_n$$

季節調整モデル(3)

- 観測方程式(モデル)は以下で与える

$$x_n = \left(t_n, t_{n-1}, s_n, s_{n-1}, s_{n-2}\right)'$$

$$H' = (1,0,1,0,0)$$

$$y_n = H' x_n + w_n$$

すなわち

$$y_n = t_n + s_n + w_n$$

時変係数ARモデル(1)

- ここからは、時系列をあらわす添え字はtで表現
- 通常の自己回帰モデルではラグつき変数にかかる係数は時間不変

$$y_t = \sum_{j=1}^{m} a_{tj} y_{t-j} + w_t \sim N\left(0, \sigma^2\right)$$
$$a_{tj} \approx a_{t-1j}$$
$$a_{tj} - a_{t-1j} \approx a_{t-1j} - a_{t-2j}$$

- 平滑化制約

$$Bx_t = x_{t-1}$$
$$a_{tj} = a_{t-1j} + v_{tj}$$
$$a_{tj} = Ba_{tj} + v_{tj}$$
$$a_{tj} - Ba_{tj} = v_{tj}$$
$$(1-B)a_{tj} = v_{tj}$$
$$\Delta a_{tj} = v_{tj}$$
$$\Delta^2 a_{tj} = v_{tj} \leftrightarrow 2a_{t-1j} - a_{t-2j} + v_{tj}$$

時変係数ARモデル(2)

- 状態空間表現(2次AR、2階差の例)
  - システム

$$\begin{bmatrix} a_{1t} \\ a_{1t-1} \\ a_{2t} \\ a_{2t-1} \end{bmatrix} = \begin{bmatrix} 2 & -1 & & \\ 1 & & & \\ & & 2 & -1 \\ & & 1 & \end{bmatrix} \begin{bmatrix} a_{1t-1} \\ a_{1t-2} \\ a_{2t-1} \\ a_{2t-2} \end{bmatrix} + \begin{bmatrix} 1 & \\ & \\ & 1 \\ & \end{bmatrix} \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}$$

$$Var(v_t) = diag(\tau^2, \tau^2)$$

  - 観測

$$y_t = (y_{t-1}, 0, y_{t-2}, 0) \begin{bmatrix} a_{1t} \\ a_{1t-1} \\ a_{2t} \\ a_{2t-1} \end{bmatrix} + w_t$$

時変係数回帰モデル

- 時間変化する係数が状態（観測不能な変数）になっている
- 係数はランダムウォークとしてモデル化されることが多く、システム方程式は簡単な形
- 観測方程式は各時点での回帰モデル

$$\beta_t = \beta_{t-1} + v_t \sim N(0, Q)$$
$$y_t = H_t^{'} \beta_t + w_t \sim N(0, R_t)$$

一般化状態空間モデル

- 非線形非ガウス状態空間モデル
- ここではとりあつかわない