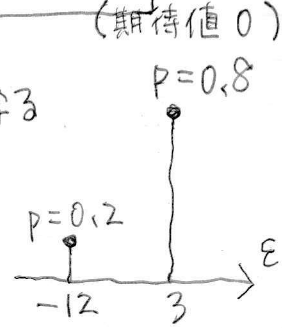
3. 予測

参考文献：経済・ファイナンスデータの計量時系列分析（沖本竜義 著 朝倉書店

3.1 予測の基礎



問. y_t が $y_t = 10 + 0.9 y_{t-1} + \varepsilon_t$ (3.1) なる定常AR(1)過程にしたがっているとする。
ただし、ε_t は右図の確率関数に i.i.d. にしたがう（離散値をとるホワイトノイズ）。
いま、時刻 t に $y_t = 110$ が観測されたとする。
このとき y_{t+1} をどのように予測すればよいだろうか？

答A. AR(1)過程 $y_t = c + \phi_1 y_{t-1} + \varepsilon_t$ の期待値は $E(y_t) = \frac{c}{1-\phi_1}$ なので、$\frac{10}{1-0.9} = 100$ を y_{t+1} の予測値と考える。

（時刻 t に $y_t = 110$ が観測されたことは完全に無視して、何も観測データをもっていない下での y_{t+1} の期待値 $E(y_{t+1})$ になるのでは、と考えるタイプ。）

答B. 次の時刻のノイズ ε_{t+1} は確率 0.8 で 3 をとるので（3 をとる確率が最も高いので）$y_{t+1} = 10 + 0.9 \cdot 110 + 3 = 112$ を y_{t+1} の予測値と考える。

（y_t を観測した下での y_{t+1} の最頻値になるのでは、と考えるタイプ。）

答C. 次の時刻のノイズ ε_{t+1} の期待値は 0 なので、$y_{t+1} = 10 + 0.9 \cdot 110 + 0 = 109$ を y_{t+1} の予測値と考える。

（y_{t+1} を観測した下での y_{t+1} の期待値を予測値とするべきでは、と考えるタイプ。）

→ 色々な予測値が出てしまったがどれが一番よいのか？

→ 一つの基準として、予測値と実現値の誤差（予測誤差）が小さいほどよい予測とする。

- といっても、実現値 y_t は確率変数である。→ 誤差も確率変数。
- 誤差にも、絶対誤差や2乗誤差など色々ある。

	予測値	実現値 $\varepsilon_{t+1}=3$ のとき ($p=0.8$)	実現値 $\varepsilon_{t+1}=-12$ のとき ($p=0.2$)	予測誤差 $\varepsilon_{t+1}=3$ のとき ($p=0.8$)	予測誤差 $\varepsilon_{t+1}=-12$ のとき ($p=0.2$)	予測誤差の絶対値の期待値	予測誤差の2乗の期待値
予測 A	100	112	97	12	−3	10.2	117
予測 B	112	112	97	0	−15	3	45
予測 C	109	112	97	3	−12	4.8	36
	①	②	③	②−①	③−①	④	⑤

$$④ = 0.8\cdot|②-①| + 0.2\cdot|③-①|$$

$$⑤ = 0.8\cdot(②-①)^2 + 0.2\cdot(③-①)^2$$

- $y_{t+1}=112$ が実現したときは予測 B が一番当たっている (②−①).
- $y_{t+1}=97$ が実現したときは予測 A が一番当たっている (③−①).

「どれがよい予測かは実現値によって違う」だと評価が定まらない.
実現値によって異なりうる誤差を1つの値にまとめたい.
誤差の期待値を取ることが考えられる.
例えば予測誤差の絶対値の期待値 (④) や予測誤差の2乗の期待値 (⑤) を予測のよさの指標とすることが考えられる. これらの指標では予測 C の誤差の期待値が一番小さい.

経済やファイナンスの分野では、この予測誤差の2乗の期待値 (⑤) を基準とすることが多い. この値を平均2乗誤差 (mean squared error: MSE) とよぶ.
MSE を最小にするような予測を最適予測という (定義 3.1).

問. 予測 C は MSE を基準とすると予測 A〜C の中で一番よい予測だった.
予測 C は最適予測なのか? (もっと MSE が小さい予測はないのか?)

答. 予測値を a としたときの MSE は

$$MSE_a = 0.8\cdot(112-a)^2 + 0.2\cdot(97-a)^2$$
$$= a^2 - 2(0.8\cdot112 + 0.2\cdot97)a + 0.8\cdot112^2 + 0.2\cdot97^2$$

$$\frac{dMSE_a}{da} = 2a - 2(0.8\cdot112 + 0.2\cdot97)$$

なので MSE_a は $a = 0.8\cdot112 + 0.2\cdot97 = 109$ のときに最小値をとる. よって 109 と予測するのが最適予測である.
よって予測 C は最適予測である.

(実は「現時点までの値を観測した下での将来の値の期待値を予測値とする」ことは「MSE が最小になる予測をする」ことと同じ.)

※ 注意

予測Aも予測Cもどちらも期待値である。
但し前提としている情報が異なる。

予測A：$E(y_{t+1})$ 何も知らない条件での y_{t+1} の期待値
予測C：$E(y_{t+1}|y_t)$ y_t を知っている条件での y_{t+1} の期待値
（条件付き期待値）

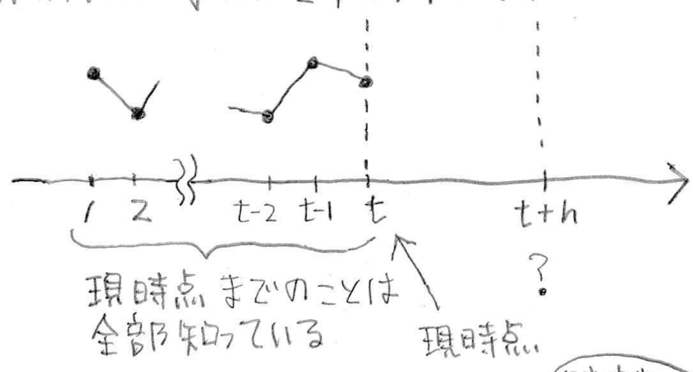
同様に、「y_t を知っている下での y_{t+1} の確率分布」のように
ある情報を前提とした確率を条件付き確率という。

$$P(y_{t+1}=112 \mid y_t=110)=0.8$$
$$P(y_{t+1}=97 \mid y_t=110)=0.2$$

3.1.2 表記と仮定

これ以降、以下を仮定する。

- 予測するときはいつも、現時点までに観測した情報集合 $\Omega_t=\{y_t, y_{t-1}, \cdots, y_1\}$ を知っているという条件の下で h 期先の y_{t+h} を予測する $(h=1,2,\cdots)$。



$$V(X)=E(X^2)-E(X)^2$$
$$\Leftrightarrow E(X^2)=V(X)+E(X)^2$$

- 予測量を $\hat{y}_{t+h|t}$ と表記する。（確率変数ではない）
- 予測誤差を $\hat{e}_{t+h|t}=y_{t+h}-\hat{y}_{t+h|t}$ と定義する。（確率変数）
- MSEの定義を $MSE(\hat{y}_{t+h|t})=E(\hat{e}_{t+h|t}^2 \mid \Omega_t)$ とする。

問. 条件付き期待値 $E(y_{t+h}|\Omega_t)$ を予測値とすると最適予測となる（MSEが最小になる）ことを示せ。

答.
$$MSE(\hat{y}_{t+h|t})=E(\hat{e}_{t+h|t}^2 \mid \Omega_t)=E\left((y_{t+h}-\hat{y}_{t+h|t})^2 \mid \Omega_t\right)$$
$$=V(y_{t+h}-\hat{y}_{t+h|t} \mid \Omega_t)+E(y_{t+h}-\hat{y}_{t+h|t} \mid \Omega_t)^2$$
$$=V(y_{t+h} \mid \Omega_t)+\left(E(y_{t+h} \mid \Omega_t)-\hat{y}_{t+h|t}\right)^2$$

なので、これが最小になるのは $\hat{y}_{t+h|t} = E(y_{t+h}|\Omega_t)$ のとき。□

ARMA過程の予測をするとき、ε_t が正規ホワイトノイズであると仮定することが多い(といっても、最尤法なりでARMA過程のパラメータを特定する時点で ε_t の過程は仮定していると思うのですが)。

- 以降の点予測の議論では、ε_t は期待値0のiid系列でさえあればいい(正規分布でなくてもいい)(独立ではあってほしい)。
- 区間予測の議論では、正規分布であることまで仮定する。
  - ※ 一般にホワイトノイズは自己無相関だが独立ではない。
  - ※ 60ページ(初版第6刷)の最後に「もし、ε_t がホワイトノイズであることだけを仮定するのであれば、以下の点予測の議論は最適線形予測を求めていると解釈することができる」とあるが、これは「ε_t の過程によってはもっとよい非線形予測が存在する」という示唆に読める。実際 ε_t が独立過程でなければ(3.6)が成り立たず、(3.7)が最適予測にはならないが、$\{y_t, \cdots, y_1\}$ の線形和でかける範囲では(3.7)が最適予測ということらしいがその証明はよくわかりません。

これ以降、以下も仮定する。

- y は定常かつ反転可能なARMA(p, q)過程であり、p, q は既にわかっており、モデルのパラメータ $\phi_1, \cdots \phi_p, \theta_1, \cdots, \theta_q, c, \sigma^2$ も既にわかっているとする。
- また、$t \geqq p$, $t \geqq q$ とする(例えばモデルが AR(5) なのに現在が $t=2$ だったら5時点分参照できないが、そういうことはないとする)。

3.2 AR過程の予測

※ ε_t がホワイトノイズなのは大前提。

一般に次が成り立つ。$E(y_\tau|\Omega_t) = y_\tau \quad (\tau \leqq t) \qquad (3.5)$

(既に観測し終わっている $y_1, \cdots, y_t$ は、もはや確率変数でないので、期待値をとっても観測した値そのものである。)

また、ε_t が独立過程なら次も成り立つ(以下、これも仮定)。

$E(\varepsilon_{t+k}|\Omega_t) = 0 \quad (k>0) \qquad (3.6)$

独立なら この情報があろうとなかろうと任意の時点のノイズの期待値はゼロ。

以下では $y_t = c + \phi_1 y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid\ N(0, \sigma^2)$

というAR(1)過程の最適予測を考える。

1期先予測は、$y_{t+1} = c + \phi_1 y_t + \varepsilon_{t+1}$ なのでこの最適予測は

$$\hat{y}_{t+1|t} = E(y_{t+1} | \Omega_t) \quad \text{さっきやった}$$
$$= E(c + \phi_1 y_t + \varepsilon_{t+1} | \Omega_t)$$
$$= c + \phi_1 y_t + E(\varepsilon_{t+1} | \Omega_t) \quad \varepsilon_t \text{が独立なら}$$
$$= c + \phi_1 y_t$$

また、予測誤差は $\hat{e}_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t} = \varepsilon_{t+1}$

MSEは $MSE(\hat{y}_{t+1|t}) = E(\hat{e}_{t+1|t}^2 | \Omega_t) = E(\varepsilon_{t+1}^2 | \Omega_t) = \sigma^2$

AR(1)過程の1期先予測の平均2乗誤差はノイズの分散に他ならない。不確かな要素がノイズの項だけなので当然である。

同様に、2期先予測も考える。

$$y_{t+2} = c + \phi_1 y_{t+1} + \varepsilon_{t+2}$$
$$= c + \phi_1 (c + \phi_1 y_t + \varepsilon_{t+1}) + \varepsilon_{t+2}$$
$$= (1 + \phi_1) c + \phi_1^2 y_t + \phi_1 \varepsilon_{t+1} + \varepsilon_{t+2}$$

$$\hat{y}_{t+2|t} = (1 + \phi_1) c + \phi_1^2 y_t$$

$$\hat{e}_{t+2|t} = \phi_1 \varepsilon_{t+1} + \varepsilon_{t+2}$$

$$MSE(\hat{y}_{t+2|t}) = E((\phi_1 \varepsilon_{t+1} + \varepsilon_{t+2})^2 | \Omega_t)$$
$$= E(\phi_1^2 \varepsilon_{t+1}^2 + 2\phi_1 \varepsilon_{t+1} \varepsilon_{t+2} + \varepsilon_{t+2}^2 | \Omega_t)$$
$$= \phi_1^2 \sigma^2 + \sigma^2$$

もっと一般にh期先予測を考える。

$$y_{t+h} = (1 + \phi_1 + \cdots + \phi_1^{h-1}) c + \phi_1^h y_t + \varepsilon_{t+h} + \phi_1 \varepsilon_{t+h-1} + \cdots + \phi_1^{h-1} \varepsilon_{t+1}$$
$$= c \sum_{k=1}^{h} \phi_1^{k-1} + \phi_1^h y_t + \sum_{k=0}^{h-1} \phi_1^k \varepsilon_{t+h-k}$$

$$\hat{y}_{t+h|t} = c \sum_{k=1}^{h} \phi_1^{k-1} + \phi_1^h y_t = \frac{1 - \phi_1^h}{1 - \phi_1} c + \phi_1^h y_t \quad (3.7)$$

$$\hat{e}_{t+h|t} = \sum_{k=0}^{h-1} \phi_1^k \varepsilon_{t+h-k}$$

$$MSE(\hat{y}_{t+h|t}) = \sigma^2 \sum_{k=0}^{h-1} \phi_1^{2k} = \frac{1 - \phi_1^{2h}}{1 - \phi_1^2} \sigma^2 \quad (3.8)$$

以上からわかること

- いま $\{y_1, \cdots, y_t\}$ が全てわかっているが最適予測につかうのは y_t のみである。AR(1)過程が1期の過去しか参照しないので y_t が手に入っていたら $y_{t-1}, y_{t-2}, \cdots$ の情報はもう y_t に入っているので $y_{t-1}, y_{t-2}, \cdots$ は要らない。
- 最適予測はAR(1)過程の条件なし期待値 $\frac{c}{1-\phi_1}$ に近づいていく。
  遠い未来を予測するほど $(h \to \infty)$
  - 遠い先の予測値ほど過程の条件なし期待値に近づいていくことを平均回帰的(mean reverting)という。
- 遠い先を予測するほど $(h \to \infty)$ MSEは $\frac{\sigma^2}{1-\phi_1^2}$ に近づく。
  - 1期先予測のMSE σ^2 からどんどん大きくなっていき過程の条件なし分散 $\frac{\sigma^2}{1-\phi_1^2}$ に収束する。

次に、一般のAR(p)過程の最適予測を考える。

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim iid\, N(0, \sigma^2) \qquad (3.9)$$

まず1期先予測は、

$$y_{t+1} = c + \phi_1 y_t + \phi_2 y_{t-1} + \cdots + \phi_p y_{t-p+1} + \varepsilon_{t+1} \qquad (3.10)$$

$$\hat{y}_{t+1|t} = c + \phi_1 y_t + \phi_2 y_{t-1} + \cdots + \phi_p y_{t-p+1} \qquad (3.11)$$

$$\hat{e}_{t+1|t} = \varepsilon_{t+1}$$

$$MSE(\hat{y}_{t+1|t} \mid \Omega_t) = \sigma^2 \qquad (3.12)$$

2期先予測は、

$$y_{t+2} = c + \phi_1 y_{t+1} + \phi_2 y_t + \cdots + \phi_p y_{t-p+2} + \varepsilon_{t+2} \qquad (*)$$

$$\begin{aligned} &= (1+\phi_1)c + (\phi_1^2 + \phi_2) y_t + (\phi_1\phi_2 + \phi_3) y_{t-1} + \cdots \\ &\quad + (\phi_1\phi_{p-1} + \phi_p) y_{t-p+2} + \phi_1\phi_p y_{t-p+1} \\ &\quad + \varepsilon_{t+2} + \phi_1 \varepsilon_{t+1} \end{aligned}$$

$$\begin{aligned} \hat{y}_{t+2|t} &= (1+\phi_1)c + (\phi_1^2 + \phi_2) y_t + (\phi_1\phi_2 + \phi_3) y_{t+1} + \cdots \\ &\quad + (\phi_1\phi_{p-1} + \phi_p) y_{t-p+2} + \phi_1\phi_p y_{t-p+1} \end{aligned} \qquad (3.13)$$

$$MSE(\hat{y}_{t+2|t}) = (1+\phi_1^2)\sigma^2$$

AR(p)過程の2,3,…期先予測は書くのが大変。
実はh期先予測はh−1期先予測をつかってかける。先のy_{t+2}の式(*)のy_{t+1}を(3.11)式の$\hat{y}_{t+1|t}$で置き換えて期待値$E(\cdot|\Omega_t)$をとればこれがy_{t+2}の予測値になる。

$$\hat{y}_{t+2|t} = c + \phi_1 \hat{y}_{t+1|t} + \phi_2 y_t + \cdots + \phi_p y_{t-p+2} \quad (3.14)$$

これに(3.11)を代入すれば(3.13)に一致する。
このように予測値を計算することを逐次予測という。
逐次予測を用いるとAR(p)過程のh期先予測は、

$$\hat{y}_{t+h|t} = c + \phi_1 \hat{y}_{t+h-1|t} + \phi_2 \hat{y}_{t+h-2} + \cdots + \phi_p \hat{y}_{t+h-p}$$

とかける(証明はy_{t+h}の式の両辺のΩ_tの下での期待値をとれば明らか)。
最後にAR(p)過程の最適予測の性質をまとめる(証明略)。

定理3.1

AR(p)過程の最適予測は以下の性質をもつ。

(1) 最適予測は直近p期の値$y_t, y_{t-1}, \cdots, y_{t-p+1}$のみに依存する。

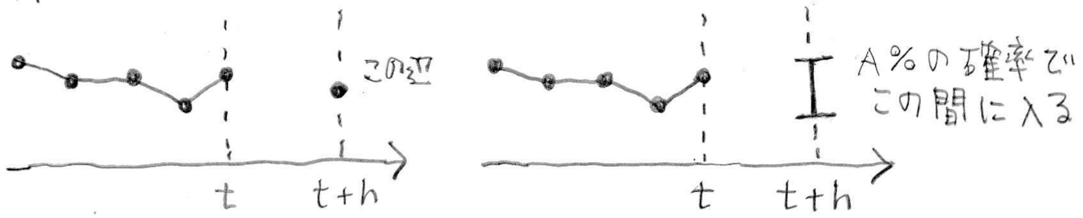
(2) 予測期間が長くなるにつれて、最適予測における過去の観測値の影響は指数的に減衰していき、最適予測は過程の期待値に収束する。

(3) 予測期間が長くなるにつれて、MSEは単調に増大し、過程の分散に収束する。

3.3 区間予測

これまで、t+h期先の値を1つの値で予測してきた。
このような予測を点予測という。
他方、t+h期先の値を含む区間を予測することを区間予測という。

点予測　　　　　　　　　　区間予測



区間予測のいいところ

- 「何%の確率でどうなる」という予測なので確率で評価できる。
- 点予測の精度はMSEで示されるがMSEがいくらというのはわかりにくい。区間予測のよさは区間の長さになるので「95%の確率で株価が100〜103円になる」のようになる。どれくらい精度がよい予測かわかりやすい。
- 平均的な場合どうなるか、だけでなく、最悪な場合どうなるか、最良な場合どうなるか、がわかる。

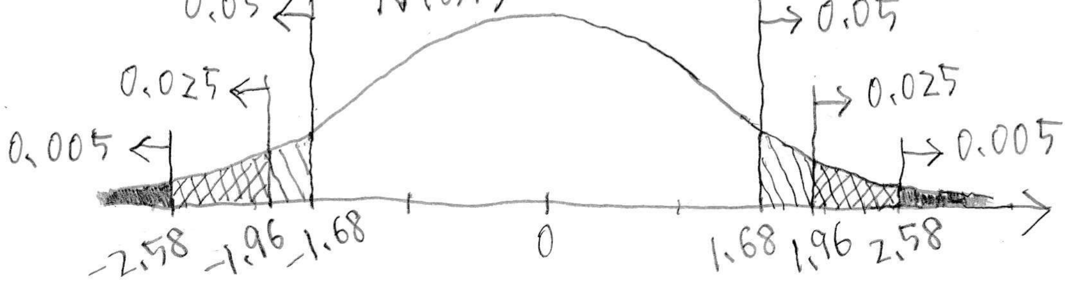
ただし、区間予測をするにはノイズの分布の形が特定できていなければならない（正規ノイズでないのに正規ノイズを仮定して区間予測してもそれは正しくない）。

(3.9)のAR(p)過程の区間予測を考える。

y_{t+1}は(3.10)とかけるので y_tまでの値がわかっている下で 不確実性があるのはε_{t+1}の項のみ。$\varepsilon_{t+1} \sim iid\,N(0, \sigma^2)$ なので Ω_t所与の下でy_{t+1}は分散σ^2の正規分布にしたがう。

期待値を知りたい → $E(y_{t+1}|\Omega_t)$は最適予測$\hat{y}_{t+1|t}$に他ならない。よって $y_{t+1|t} \sim N(\hat{y}_{t+1|t}, \sigma^2)$。これが1期先予測の分布なので95%区間予測は、

$$(\hat{y}_{t+1|t} - 1.96\sigma,\ \hat{y}_{t+1|t} + 1.96\sigma) \qquad (3.15)$$



一般に時点 t までの情報 Ω_t が手に入っている下で h 期先の値は $y_{t+h|t} \sim N(\hat{y}_{t+h|t}, MSE(\hat{y}_{t+h|t}))$ にしたがう。

$$
\begin{aligned}
\because\ E(y_{t+h}|\Omega_t) &= \hat{y}_{t+h|t} \\
V(y_{t+h}|\Omega_t) &= V(y_{t+h} - \hat{y}_{t+h|t}|\Omega_t) \\
&= V(\hat{e}_{t+h|t}|\Omega_t) \\
&= E(\hat{e}^2_{t+h|t}|\Omega_t) - E(\hat{e}_{t+h|t}|\Omega_t)^2 \\
&= MSE(\hat{y}_{t+h|t}) - (E(y_{t+h|t}|\Omega_t) - \hat{y}_{t+h|t})^2 \\
&= MSE(\hat{y}_{t+h|t})
\end{aligned}
$$

ので、h 期先 95% 区間予測は、

$$
\left(\hat{y}_{t+h|t} - 1.96\sqrt{MSE(\hat{y}_{t+h|t})},\ \hat{y}_{t+h|t} + 1.96\sqrt{MSE(\hat{y}_{t+h|t})}\right) \tag{3.16}
$$

しかし、AR(p) 過程の h 期先予測の MSE を求めることは難しい。
($\hat{y}_{t+h|t}$ は逐次的に求められるが $MSE(\hat{y}_{t+h|t})$ はそうはいかない。)
なので、実際には (3.16) で区間予測はできない。

解決策1. カルマンフィルタをつかって MSE を求める。
解決策2. シミュレーションで数値的に MSE を求める。

- まず逐次予測で $\hat{y}_{t+h|t}$ を求める。
- $k=1,2,\cdots,N$ について以下を繰り返す。
  - $\varepsilon^{(k)}_{t+1}, \varepsilon^{(k)}_{t+2}, \cdots, \varepsilon^{(k)}_{t+h}$ を生成する。
  - 生成したノイズを用いて h 期先までシミュレーションし $y^{(k)}_{t+h}$ を得る。
- 得られた N 個の標本から MSE を得る。

$$
\widehat{MSE}(\hat{y}_{t+h|t}) = \frac{1}{N}\sum_{k=1}^{N}(y^{(k)}_{t+h|t} - \hat{y}_{t+h|t})^2
$$

3.4 MA過程の予測

3.4.1 無限個の観測値がある場合の予測

AR過程との違い：参照する過去の値が直接手に入らない。

$y_t = c + \phi_1 y_{t-1} + \varepsilon_t$ 直接手に入る（さっき観測した値）

$y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t$ 直接手に入らない（さっきのノイズというものは直接わからない）

但し、MA過程は（反転可能条件を満たす表現を選択すれば）AR(∞)過程に書き直すことができる（2.2.2 AR過程の反転可能性）。

MA(1)の例．

$$y_t = \theta\varepsilon_{t-1} + \varepsilon_t \iff \varepsilon_t = -\theta\varepsilon_{t-1} + y_t$$

$$\begin{aligned}\varepsilon_t &= -\theta\varepsilon_{t-1} + y_t \\ &= -\theta(-\theta\varepsilon_{t-2} + y_{t-1}) + y_t \\ &= (-\theta)^m \varepsilon_{t-m} + \sum_{k=0}^{m-1}(-\theta)^k y_{t-k} \xrightarrow[m\to\infty]{} \sum_{k=0}^{\infty}(-\theta)^k y_{t-k}\end{aligned}$$

（時点 t に y_t を観測しても ε_t の値はわからない。
ただ、ε_t は y_t から $\theta\varepsilon_{t-1}$ を抜いたものではある。
といっても、ε_{t-1} の値もわからない（y_{t-1} はわかるけど）。
でも、ε_{t-1} は、y_{t-1} から $\theta\varepsilon_{t-2}$ を抜いたものではある。
…というようにどんどんさかのぼると、ε_t の値は
$y_t, y_{t-1}, \cdots, y_{t-m+1}, \varepsilon_{t-m}$ の式でかける。しかも、
m を $+\infty$ にとばせば、ε_{t-m} の係数が0に収束するので
結局 ε_t は $y_t, y_{t-1}, \cdots\cdots$（無限に続く）の式でかける。）

ε_{t-1} を $y_{t-1}, y_{t-2}, \cdots$ の式でかいたものをMA(1)に代入すると、

$$\begin{aligned}y_t &= \theta\varepsilon_{t-1} + \varepsilon_t \\ &= \theta\sum_{k=0}^{\infty}(-\theta)^k y_{t-1-k} + \varepsilon_t = -\sum_{k=1}^{\infty}(-\theta)^k y_{t-k} + \varepsilon_t\end{aligned}$$

一般の反転可能なMA(q)でも

とにかく無限の過去まで $y_t, y_{t-1}, \cdots$ が手に入っているなら
$\varepsilon_t, \varepsilon_{t-1}, \cdots$ も求まる。現実的ではないがまずこの仮定の下で
MA過程の予測を考える。

反転可能な MA(q) 過程は一般に $y_t = \sum_{k=1}^{\infty}\eta_k y_{t-k} + \varepsilon_t$
と AR(∞) 過程にかき直せる。これより

$$\varepsilon_t = y_t - \sum_{k=1}^{\infty}\eta_k y_{t-k} \quad (3.17)$$

として過去の ε_t の値を求められる。

MA(2)過程の予測を考えてみる.

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, \quad \varepsilon_t \sim iid\, N(0, \sigma^2) \quad (3.19)$$

まず1期先予測を考える.

$$y_{t+1} = \mu + \varepsilon_{t+1} + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1}$$

$$\hat{y}_{t+1|t} = \mu + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} \quad (3.20)$$

← $y_t, y_{t-1}, \cdots\cdots$ を無限の過去まで使用.
(AR(p)過程の予測が直近p期のみ使用していたのと対照的.)

$$\hat{e}_{t+1|t} = \varepsilon_{t+1}$$

$$MSE(\hat{y}_{t+1|t}) = \sigma^2 \quad (3.21)$$

次に2期先予測を考える.

$$y_{t+2} = \mu + \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1} + \theta_2 \varepsilon_t$$

$$\hat{y}_{t+2|t} = \mu + \theta_2 \varepsilon_t$$

$$\hat{e}_{t+1|t} = \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1}$$

$$MSE(\hat{y}_{t+2|t}) = (1+\theta_1^2)\sigma^2$$

3期先予測も考える.

$$y_{t+3} = \mu + \varepsilon_{t+3} + \theta_1 \varepsilon_{t+2} + \theta_2 \varepsilon_{t+1} \quad (3.22)$$

$\hat{y}_{t+3|t} = \mu$ ← 過程の期待値に等しい.

$$\hat{e}_{t+3|t} = \varepsilon_{t+3} + \theta_1 \varepsilon_{t+2} + \theta_2 \varepsilon_{t+1}$$

$MSE(\hat{y}_{t+2|t}) = (1+\theta_1^2+\theta_2^2)\sigma^2$ ← 過程の分散に等しい.

4期先以降も、過程の期待値になるのでは、という予測にしかならない.

定理3.2

MA(q)過程の最適予測は以下の性質をもつ.

(1) q期までの最適予測はすべての観測値 $y_t, y_{t-1}, \cdots$ に依存する.

(2) q+1期先以上の予測は単に予測の期待値に等しい.

(3) q期までの予測のMSEは過程の分散に等しくなる.

3.4.2 有限個の観測値しかない場合の予測

実際には有限個の観測値しかないのでその場合の予測を考える。
先のMA(2)過程の例だと、3期先以降に最適予測が過程の期待値になるのは変わらない。
1,2期先予測に出てくる $\varepsilon_t, \varepsilon_{t-1}$ は有限個の観測値しかないときは求まらない。そこで、近似値を求める。

(3.19)より
$$\varepsilon_t = y_t - \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$
$$\varepsilon_{t-1} = y_{t-1} - \mu - \theta_1 \varepsilon_{t-2} - \theta_2 \varepsilon_{t-3}$$
$$\cdots$$
$$\varepsilon_3 = y_3 - \mu - \theta_1 \varepsilon_2 - \theta_2 \varepsilon_1$$
$$\varepsilon_2 = y_2 - \mu - \theta_1 \varepsilon_1$$
$$\varepsilon_1 = y_1 - \mu$$

ε_t を知るには $\varepsilon_{t-1}, \varepsilon_{t-2}$ を知りたいがわからない。のでどんどん過去までさかのぼる。
手持ちのデータでさかのぼれるところまでさかのぼる。
標本期間以前の ε はゼロとする(標本期間以前の y を過程の期待値 μ とみなすことに等しい)。
無限個の観測値がある場合でも、過去の y の影響は急速に小さくなるので、大きな問題にはならない。

この要領で $\varepsilon_1, \varepsilon_2, \cdots$ と特定していけば(近似値ではあるが)現在の ε_t まで求まる。

MA過程でもAR過程と同様に区間予測を考えることができる。

3.5 ARMA過程の予測

AR過程とMA過程の予測を組み合わせればよい。

- MA過程部分の予測値は近似値になる。
- しかし、$q+1$ 期先以降はMA過程部分の影響は消える。
- 区間予測はシミュレーションなどで求められる。