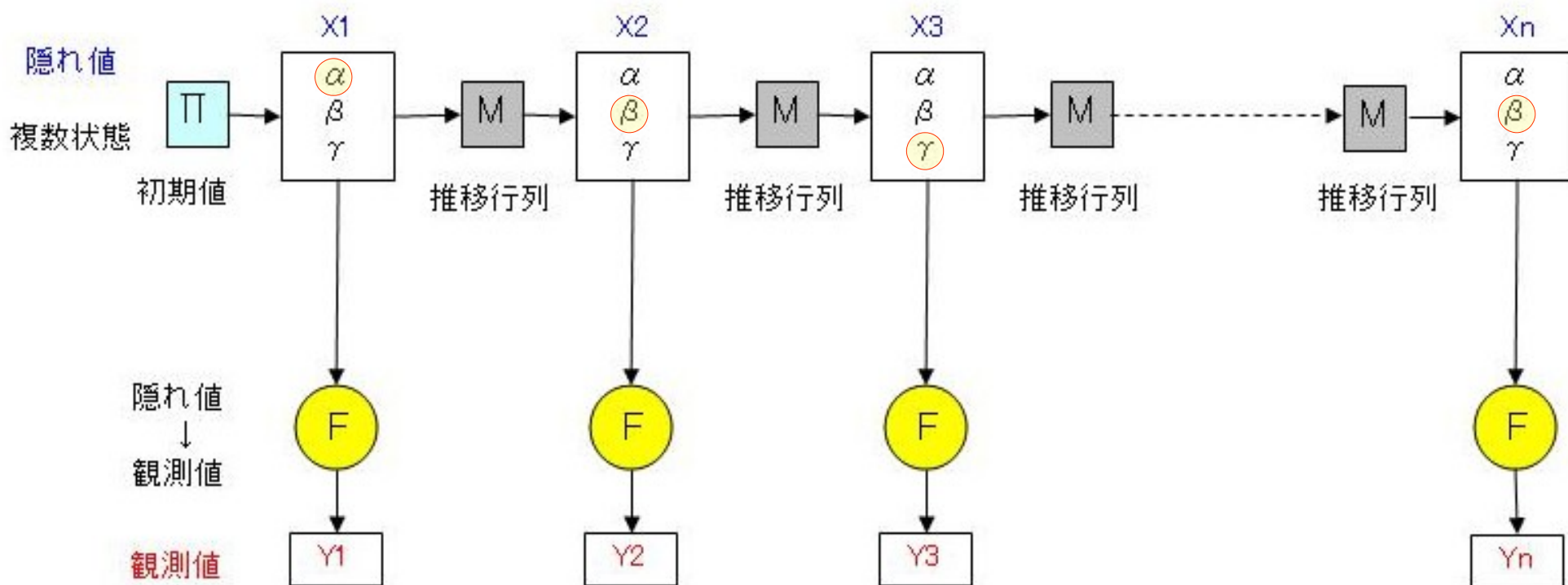


# 隠れマルコフモデル(HMM)例

- 自己紹介 @mabo0725  
20年以上データ分析
- 離散型HMMの計算例の紹介  
計算はEM版(MCMC版の方が実装は簡単)  
観測データは富士通(2000年ごろの)の株価

# 離散型HMMの考え方

観測値は隠れ値の時系列が起因して観測されているとします



# 離散型HMMのアルゴリズム

- 観測値の時系列は背後の要因で変動する
- 背後にある要因を隠れ状態とする  
状態の推定なのでカルマンフィルターと同類
- 以下を隠れ変数としてEMアルゴリズムで解く

$\pi[I]$  : 初期確率  
隠れ値 $X[1]$  が  $I$  の状態になる初期状態確率

$M[I][J]$  : 遷移確率  
隠れ値 $X[t]$ が  $I$  の状態 →  
次の隠れ値 $X[t+1]$ が  $J$  の状態になる遷移確率

$F[I][Y[t]]$  : 出現確率  
隠れ値 $X[t]$ が  $I$  の状態 → 観測値 $Y[t]$ になる出現確率

# 離散型HMMのアルゴリズム

①観測値 $Y[t]$ の読み込み

②隠れ変数を乱数で初期化

隠れ変数  $\pi$ (初期確率)、 $M$ (遷移確率)、 $F$ (出現確率)

③EMの繰返し

▶ a) Eステップでパラメータ推定

隠れ変数 $\pi, M, F$  → パラメータ $\alpha, \beta, \gamma, \xi$ 推定

※パラメータ  $\alpha, \beta$  (フォワード・バックワード法で計算)

b) Mステップで隠れ変数推定

パラメータ $\alpha, \beta, \gamma, \xi$  → 隠れ変数 $\pi, M, F$ 推定

c) 隠れ変数値が変わらなくなる(収束) → 終了

④隠れ変数の出力

# 理論は複雑だがプログラムは簡単

① 観測データ  $Y[t]$  を  $N$  個読み込む

② 隠れ変数を乱数で仮に設定

$\pi[i]$  : 隠れ値  $X[1]$  が  $i$  の状態になる確率

$M[i][j]$  : 隠れ値  $X[t]$  が  $i$  の状態から次の隠れ値  $X[t+1]$  が  $j$  の状態になる遷移確率

$F[i][Y[t]]$  : 観測値が  $Y[t]$  の場合、隠れ値  $X[t]$  が  $i$  の状態になる出現確率

③ EM過程 (以下を収束まで繰り返す)

a) Eステップ

$$\alpha[1][i] = Y[1] * F[i] * \pi[i]$$

$$\alpha[t][i] = F[i][Y[t]] + \sum \alpha[t-1][j] * M[i][j]$$

$\alpha$  の漸化式  $[1 \rightarrow N]$  (フォワード)

$$\beta[t][i] = \sum \beta[t+1][j] * F[j][Y[t+1]] * M[j][i]$$

$\beta$  の逆漸化式  $[N \rightarrow 1]$  (バックワード)

$$\xi[t][i][j] = \alpha[t][i] * F[i][Y[t]] * M[i][j] * \beta[t][j]$$

$$\gamma[t][i] = \alpha[t][i] * \beta[t][i]$$

b) Mステップ

$$\pi[i] = \gamma[1][i]$$

$$M[i][j] = \sum \xi[t][i][j] / \sum \gamma[t][i]$$

$$F[i][K] = \sum \gamma[t][i] * I(Y[t]=K) / \sum \gamma[t][i]$$

$I(Y[t]=K)$  が真の時 1 それ以外は 0

c)  $\pi$   $M$   $F$  が収束  $\rightarrow$  終了

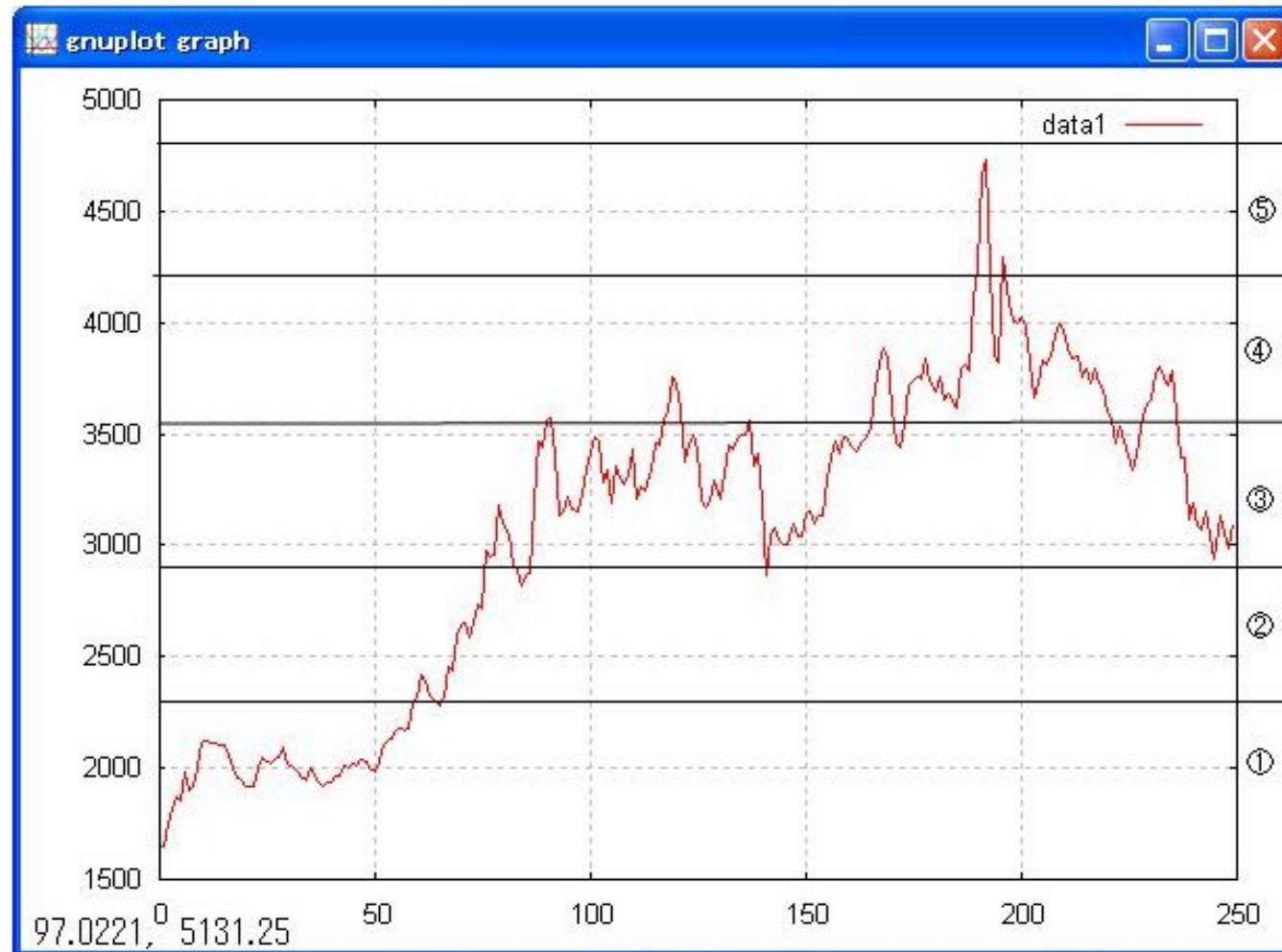
④  $\pi[i]$   $M[i][j]$   $F[i][K]$  を出力

# 観測値「250日の株価」(入力)

計算のため離散化する

観測値は 5個(低→高)の状態とする

隠れ値は 3個の状態とする(根拠なし)



# 離散型HMM 計算結果(出力)

ANALYSIS

隠れ状態数      観測状態数      繰返数

Time = 250 Dim of Hidden = 3 Dim of Observed = 5 iterations = 263

Final Hidden Initial probability  
0.00000 0.00000 1.00000      Π      初期確率

Final Hidden Transer Matirx  
0.91322 0.08678 0.00000  
0.05556 0.94444 0.00000      M      遷移確率  
0.00000 0.01695 0.98305

Final Hidden To Observed Matirx  
0 0.00000 0.00000 1.00000  
1 0.00000 0.18128 0.00000  
2 0.00000 0.80283 0.00000  
3 0.93258 0.01589 0.00000  
4 0.06742 0.00000 0.00000      F      出現確率

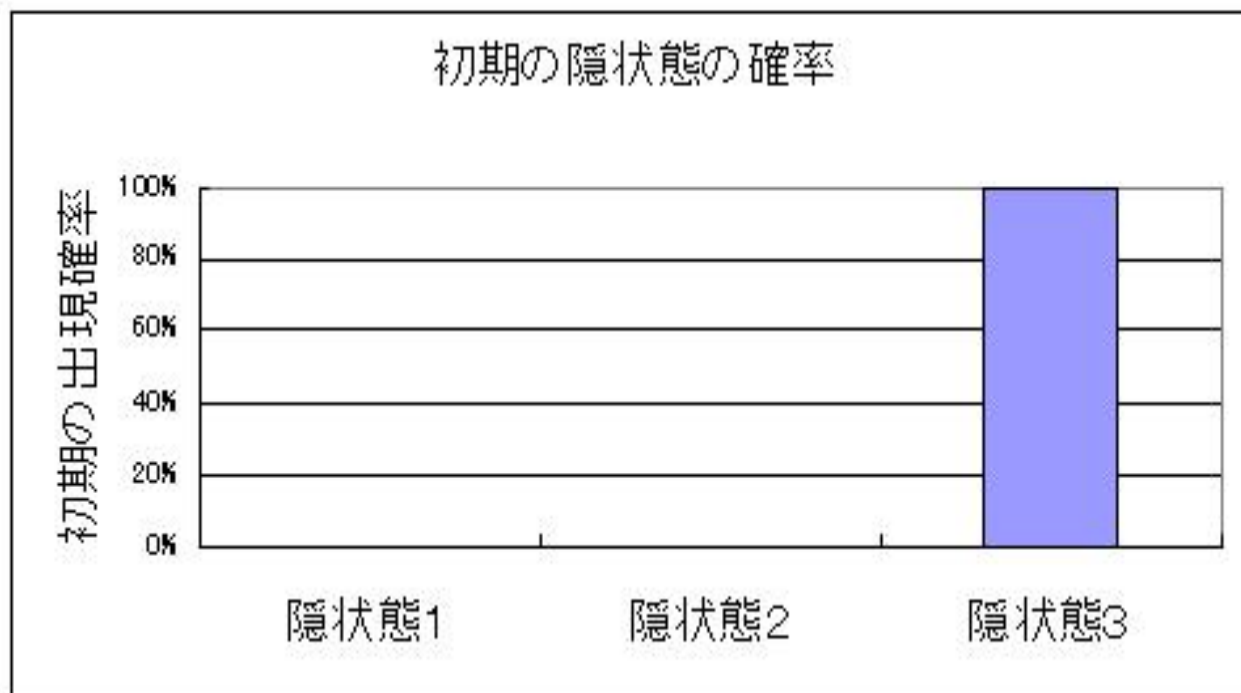
log [P(observations | lambda)] = -131.805482

キャンセル      OK

# 離散型HMM 計算結果(初期確率)

π 初期の隠状態の確率

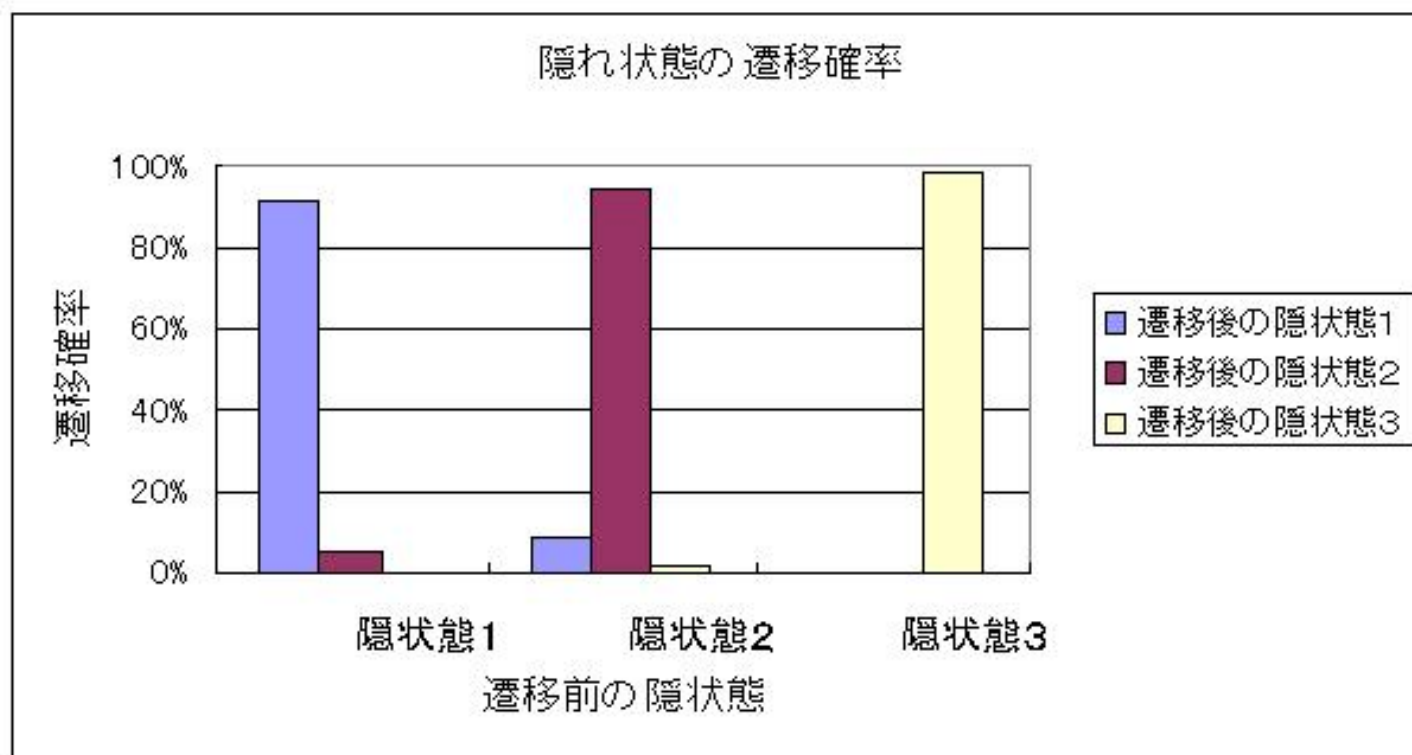
隠状態1	隠状態2	隠状態3
0.00000	0.00000	1.00000





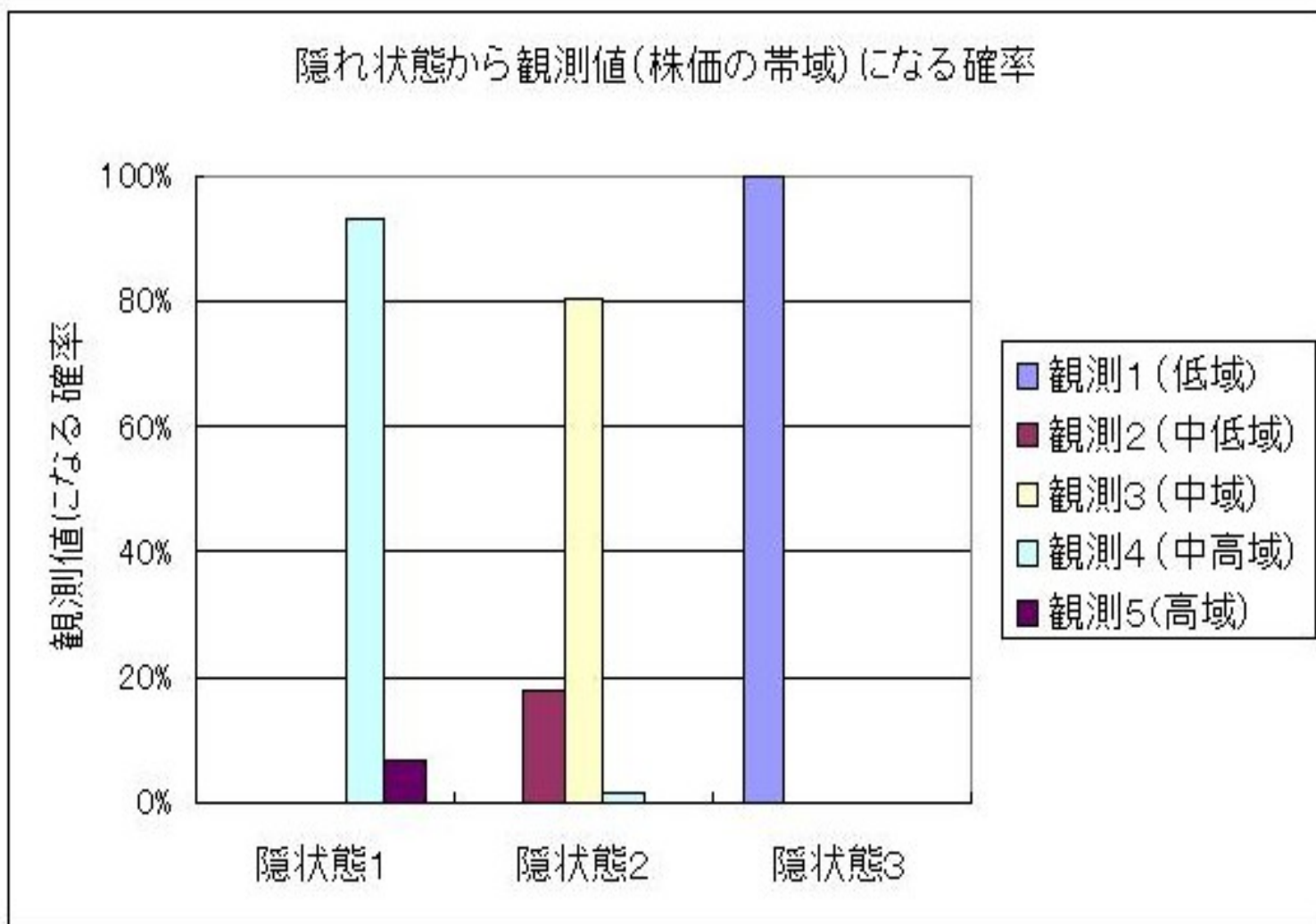
# 離散型HMM 計算結果(遷移確率)

	隠状態1	隠状態2	隠状態3
遷移後の隠状態1	0.91322	0.08678	0.00000
遷移後の隠状態2	0.05556	0.94444	0.00000
遷移後の隠状態3	0.00000	0.01695	0.98305



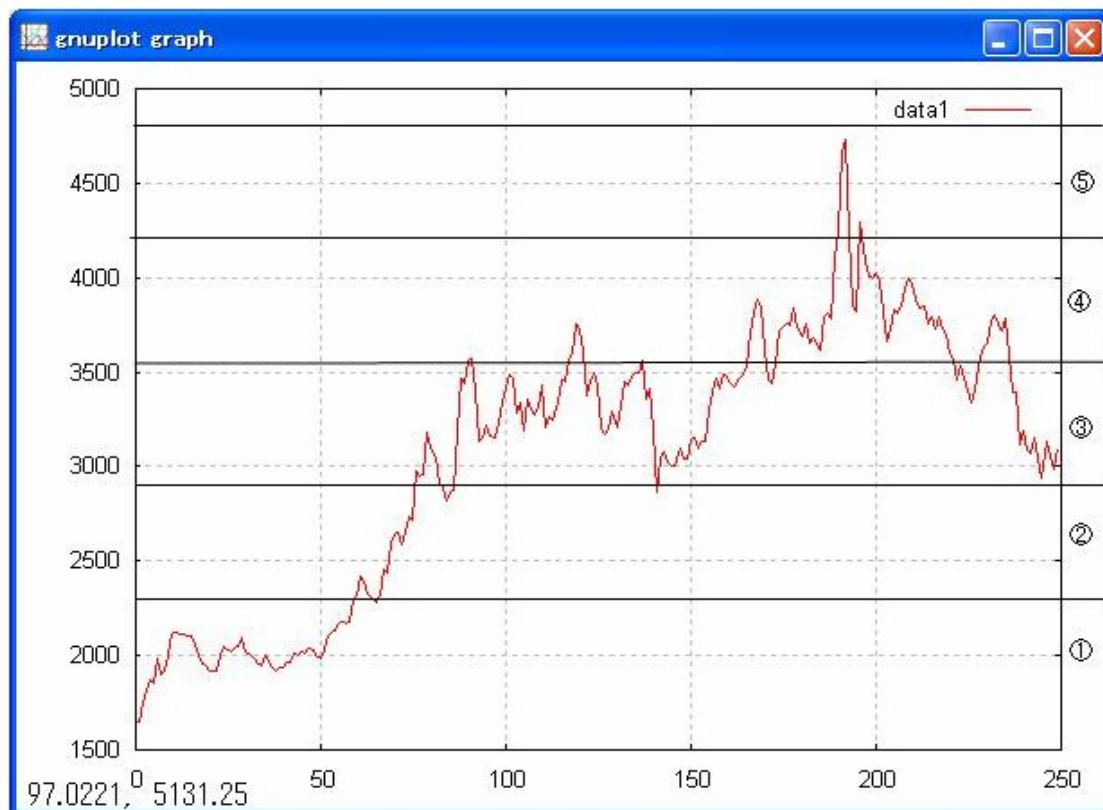
隠れ値の推移確率行列からは異なる状態になり難い事を示しています。

# 離散型HMM 計算結果(出現確率)



# 離散型HMM 計算例 まとめ

- 隠れ状態1 → 観測値は中高域④ですが、偶に高域⑤で現れます
- 隠れ状態2 → 観測値は中域③に留り、殆ど中高域④で現れない
- 隠れ状態3 → 観測値は低価格①に留まる事を示しています。



一旦隠れ状態が変わると  
その状態が続くことを示し  
ている  
(安定的な銘柄)