統数研共同研究集会「経済物理学とその周辺」
H24年度第一回研究会 キヤノングローバル戦略研究所　2012.8

乱数度計RMTテストの実データへの応用
～ハッシュ値とTick株価～

楊 欣　　田中 美栄子

鳥取大学大学院工学研究科
情報エレクトロニクス専攻

研究背景

- 乱数の性質を調べるためには，統計的な手法を使って，その性質を解析するという手段が一般的であり，様々な統計的検定法が提案されている 。
- しかし，乱数検定ツールや文献によって，採用している検定の種類や数はまったく異なっているのが現状である。

RMTテストの定性評価

相関行列

固有値

固有値分布

既知のM-P関数

比較

RMTテストの定性評価

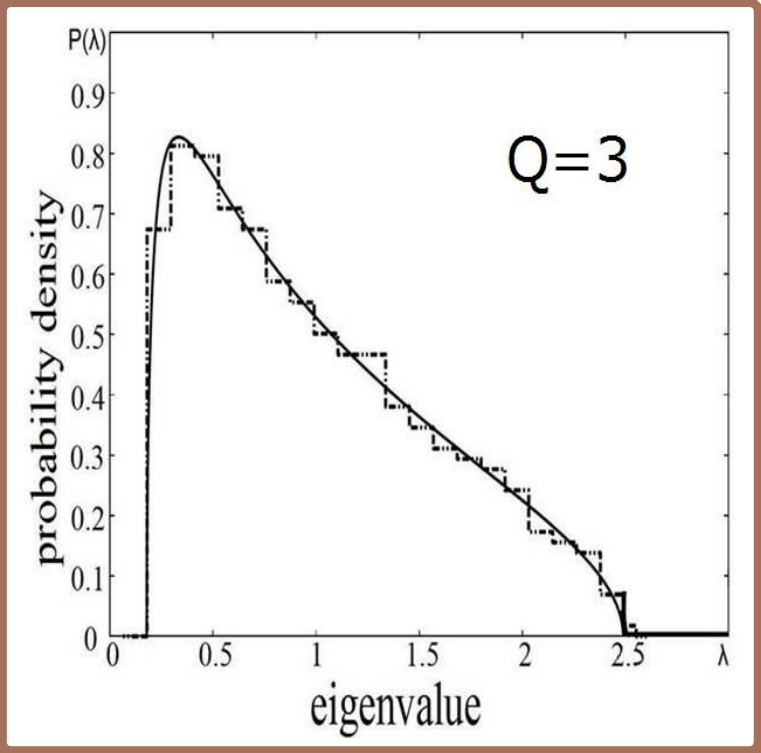
- 評価対象

✓擬似乱数:線型合同法(LCG)
メルセンヌ・ツイスター(MT)

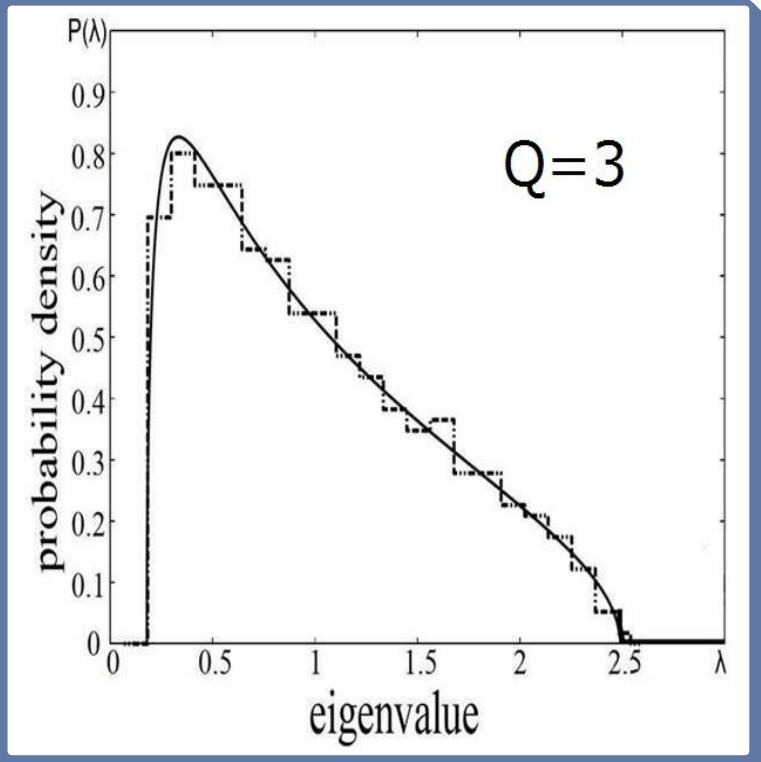
✓物理乱数:東芝製(Toshiba)
日立製(Hitachi)
東京エレクトロン製(TED)

RMTテストの定性評価(高い例)

- 評価結果:



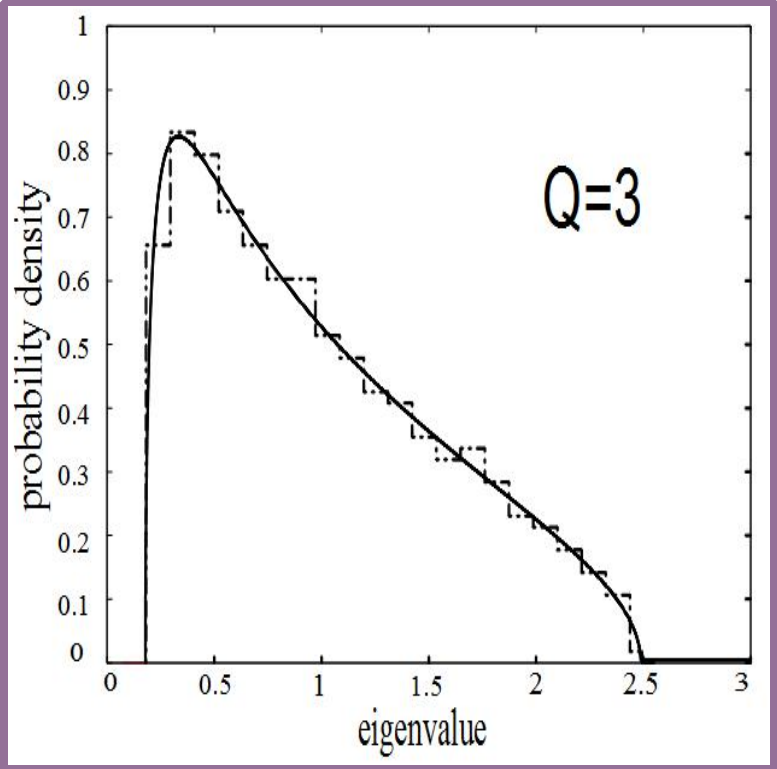
LCG



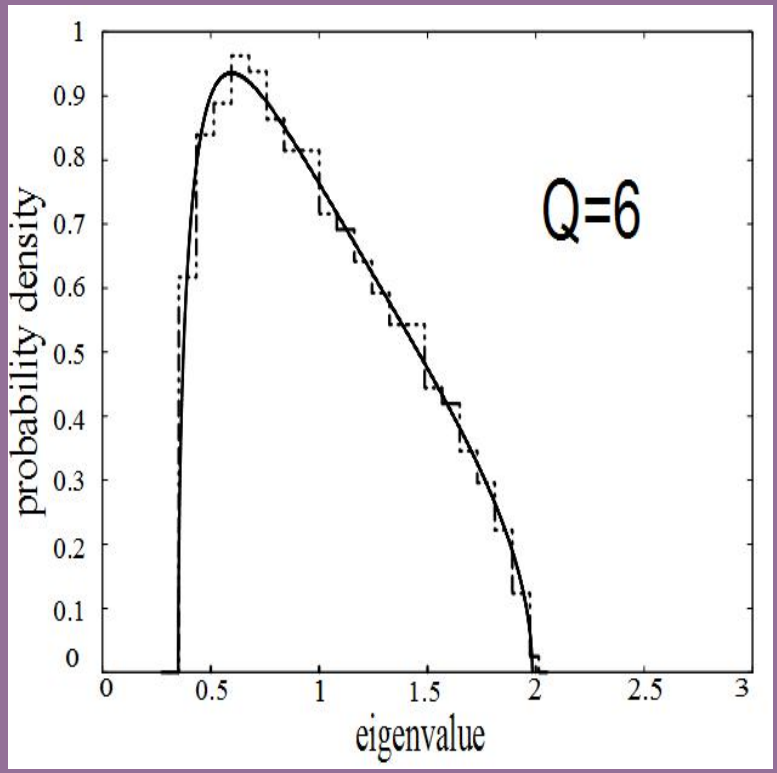
MT

RMTテストの定性評価(高い例)

- 評価結果:



Toshiba



Toshiba

RMTテストの定性評価(低い例)

- 評価結果:

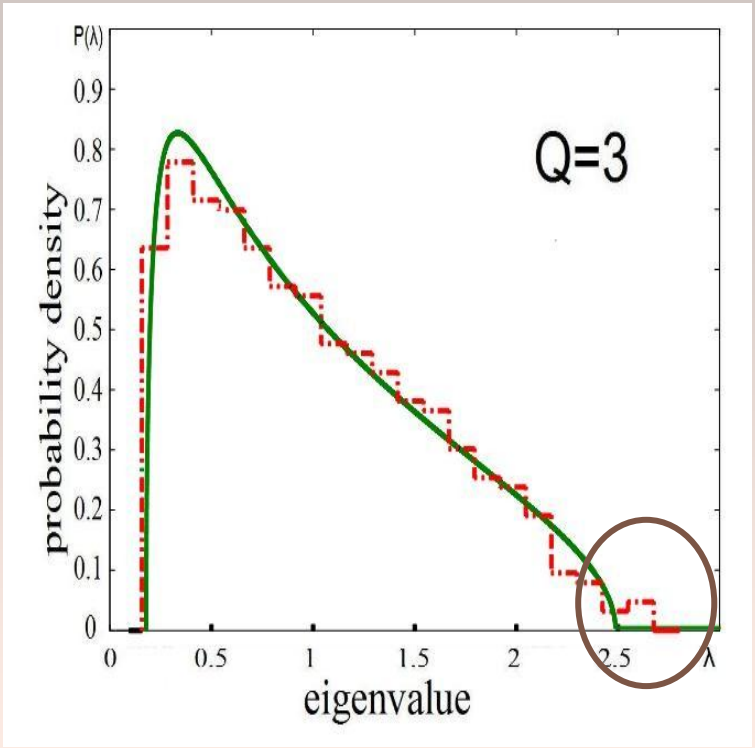
- 評価結果:

- ✓ LCGの初期乱数部分と
- ✓ 乱数時系列の対数収益,

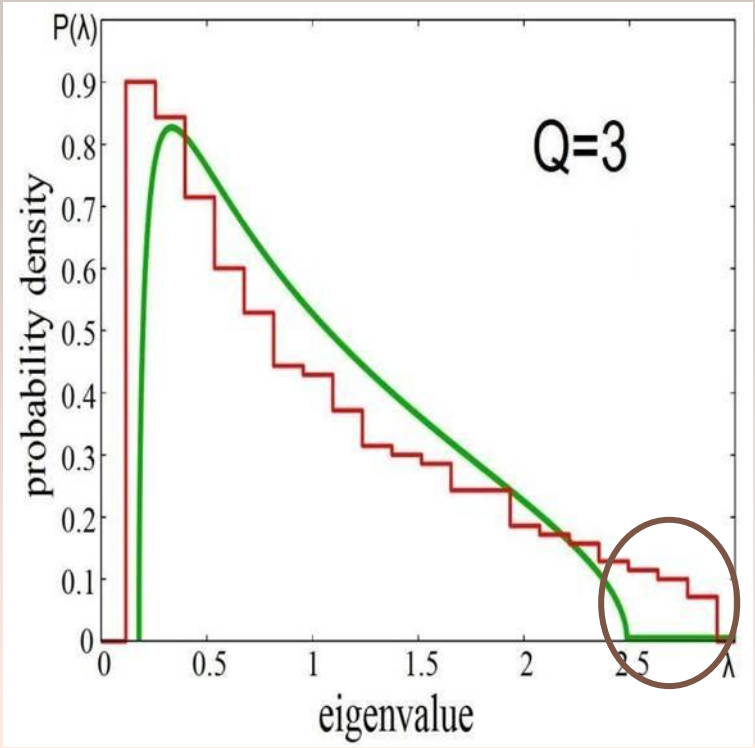
いずれも乱数度が低かった.

RMTテストの定性評価(低い例)

- 評価結果:



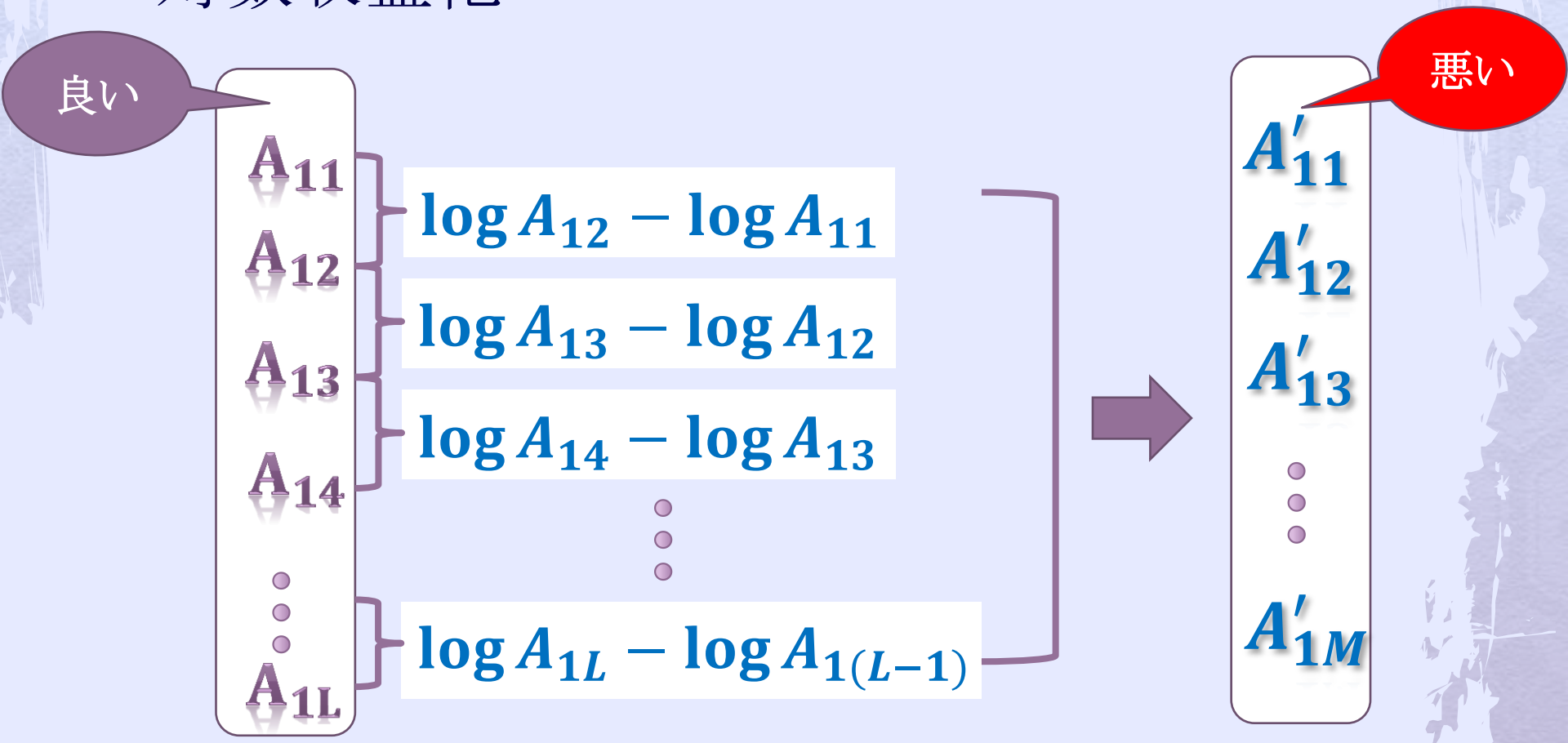
LCGの初期部分



LCG乱数の対数収益列

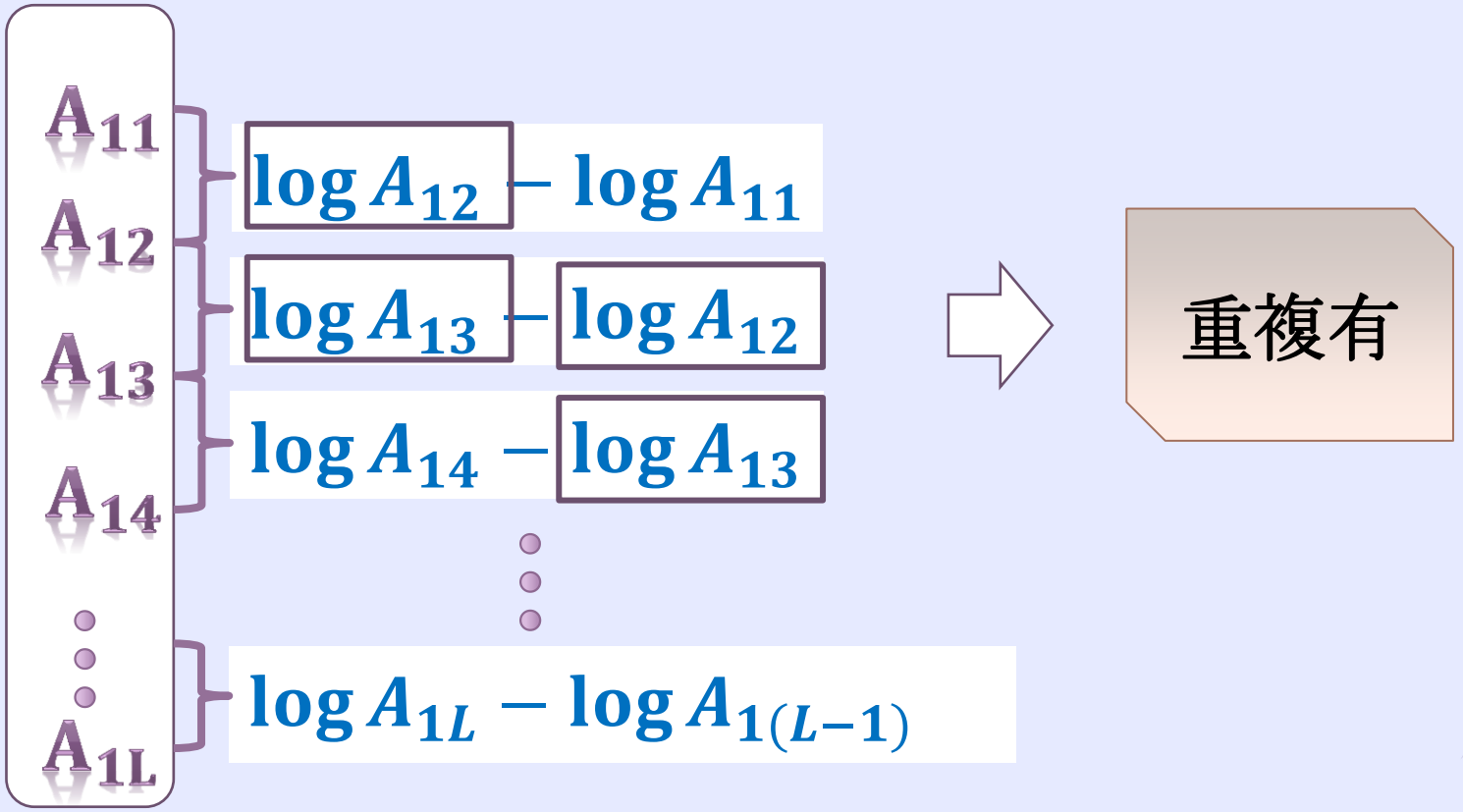
対数収益について

- 株価を処理する時によく使われる，時系列の対数収益化



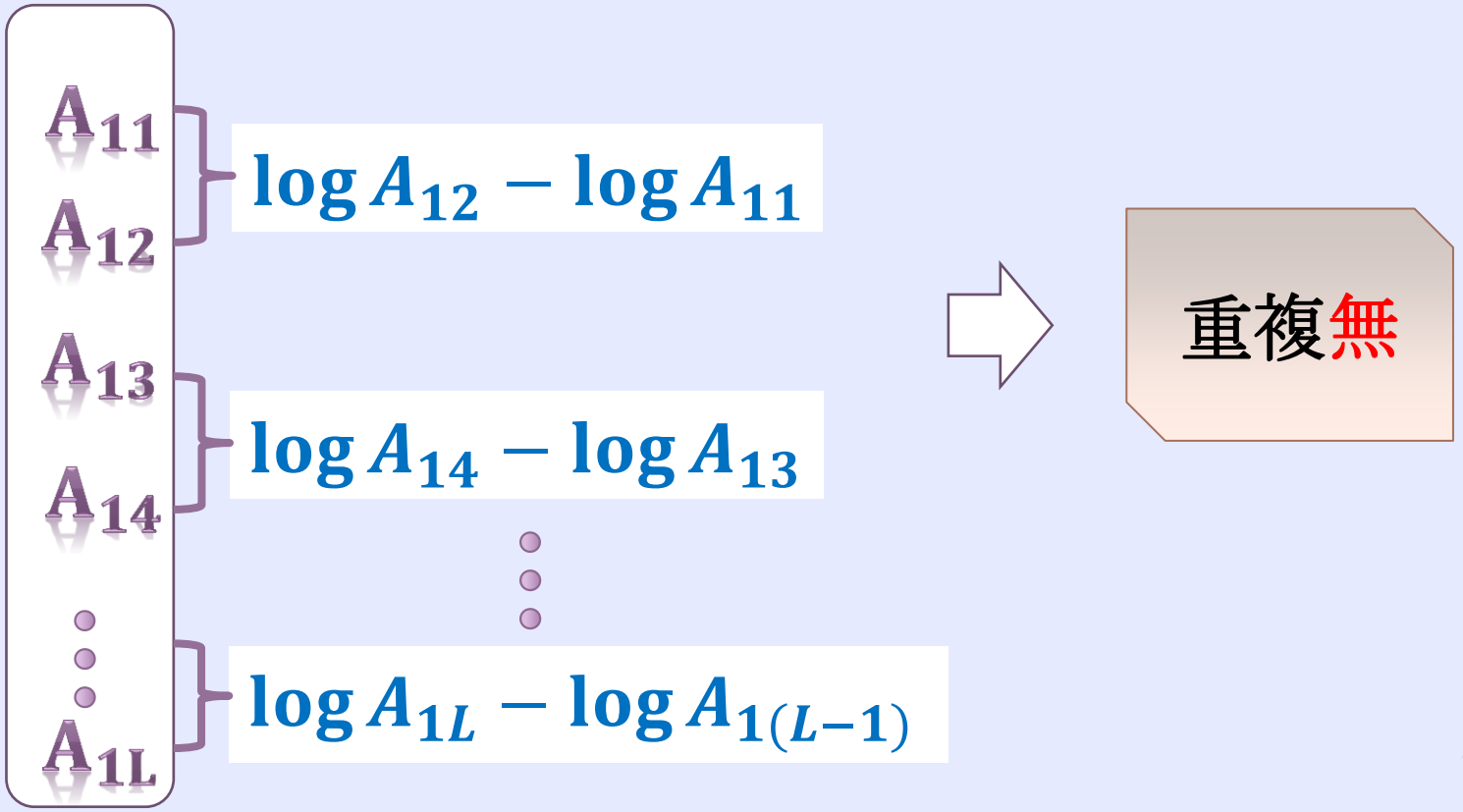
対数収益について

- 対数収益を取ったデータ乱数度低下の原因



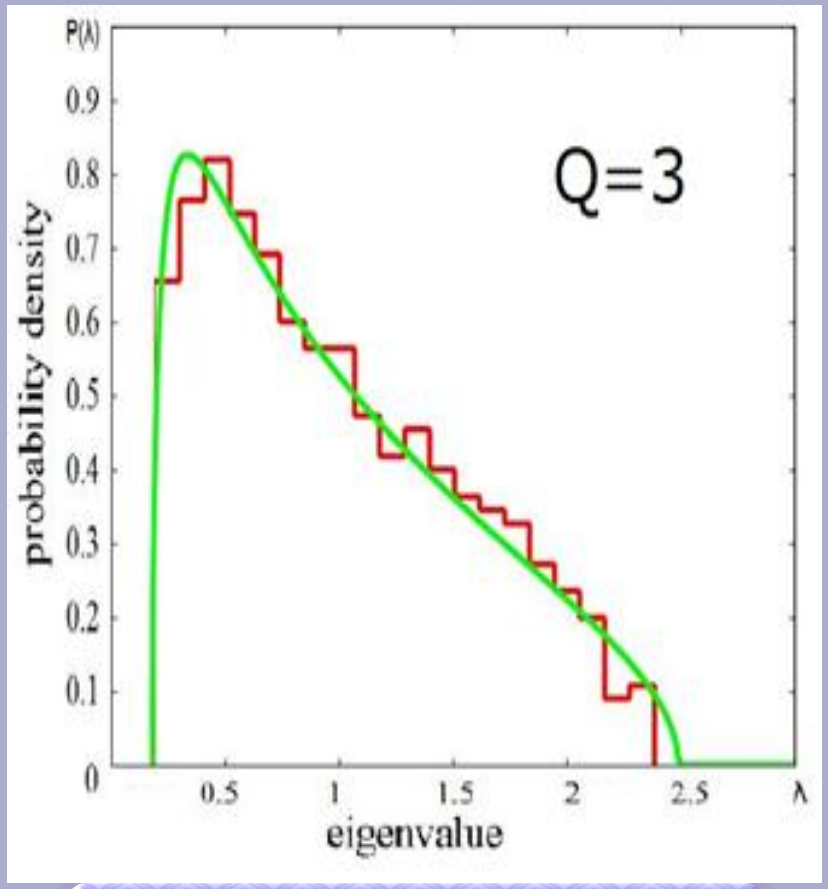
対数収益について

- 対数収益を取ったデータ乱数度低下の原因

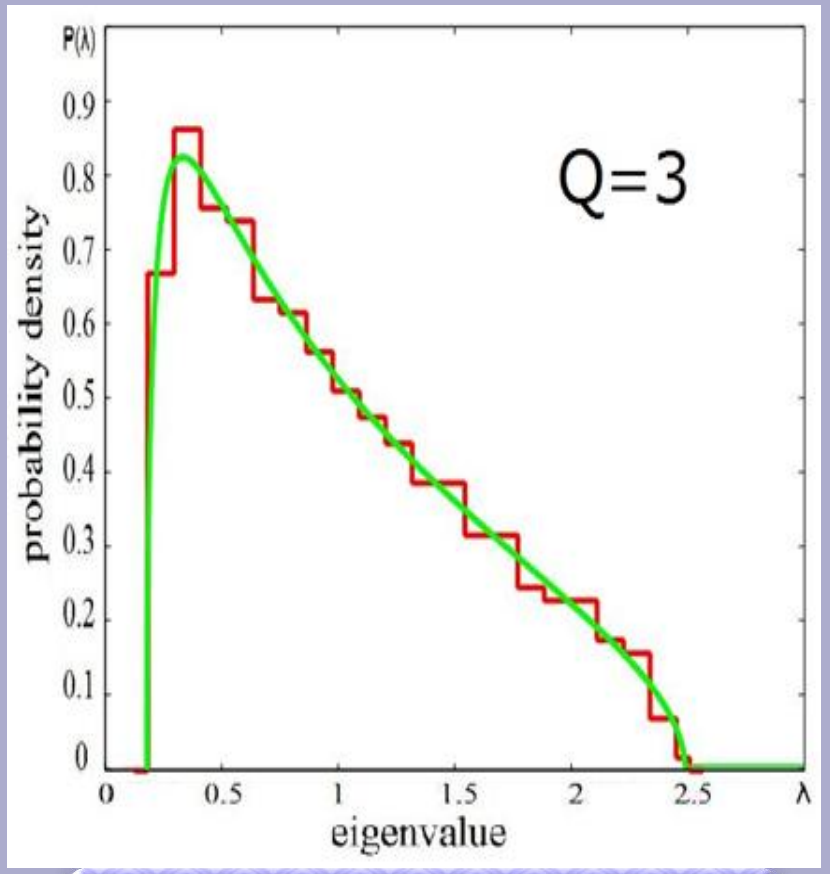


対数収益について

重複無にすると
ランダム性が戻った



LCG



MT

RMTテストの定性評価

- 五種類データのランダム性は良好
- 「LCGの初期部分」,「乱数列を対数収益化したデータ」の乱数度が低いことを検出

モーメント分析法と定量評価

Moment:

$$\mu_k = E(\lambda^k) = \int_{\lambda_-}^{\lambda_+} \lambda^k P_{RMT}(\lambda) d\lambda$$

⇨ 理論値

$$m_k = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^k$$

モーメント分析法と定量評価

$$\mu_1 = 1$$

$$\mu_2 = 1 + 1/Q$$

$$\mu_3 = 1 + 3/Q + 1/Q^2$$

$$\mu_4 = 1 + 6/Q + 6/Q^2 + 1/Q^3$$

$$\mu_5 = 1 + 10/Q + 20/Q^2 + 10/Q^3 + 1/Q^4$$

$$\mu_6 = 1 + 15/Q + 50/Q^2 + 50/Q^3 + 15/Q^4 + 1/Q^5$$

モーメント分析法と定量評価

Moment:

$$\mu_k = E(\lambda^k) = \int_{\lambda_-}^{\lambda_+} \lambda^k P_{RMT}(\lambda) d\lambda$$

⇨ 理論値

$$m_k = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^k$$

⇨ 実測値

実測値と理論値の誤差＝$|m_k/\mu_k - 1| \times 100\%$

定量評価基準の決定

サンプル数＝100

- 五種類のデータの結果:AV.(SD.)(N=500,Q=3,単位：%)

k	LCG	MT	Toshiba	Hitachi	TED
2	0.04(0.10)	0.04(0.09)	0.04(0.10)	0.04(0.10)	0.04(0.09)
3	0.10(0.26)	0.09(0.24)	0.11(0.26)	0.11(0.25)	0.09(0.25)
4	0.18(0.47)	0.14(0.41)	0.19(0.46)	0.19(0.44)	0.15(0.44)
5	0.27(0.72)	0.19(0.62)	0.26(0.70)	0.28(0.66)	0.19(0.67)
6	0.36(1.00)	0.22(0.85)	0.33(0.96)	0.37(0.92)	0.21(0.93)

定量評価基準の決定

サンプル数＝100

- 上表により，五つのデータに対して，5次以下のモーメントの誤差が同程度で比較しにくい
- 6次モーメントでは差が明確に出るため，6次モーメントの誤差により定量評価基準を決める

定量評価基準の決定

- **100サンプル**の結果:[MIN:MAX] (N=500,Q=3,単位:%)

k	LCG	MT	Toshiba	Hitachi	TED
4	[-1.00:0.82]	[-1.05:0.86]	[-1.42:1.45]	[-1.55:0.89]	[-1.25:1.06]
5	[-1.47:1.25]	[-1.43:1.19]	[-2.11:2.08]	[-1.93:1.29]	[-1.74:1.61]
6	[-1.98:1.76]	[-1.83:1.53]	[-2.86:2.77]	[-2.57:1.81]	[-2.26:2.18]

誤差絶対値は3%以下になった

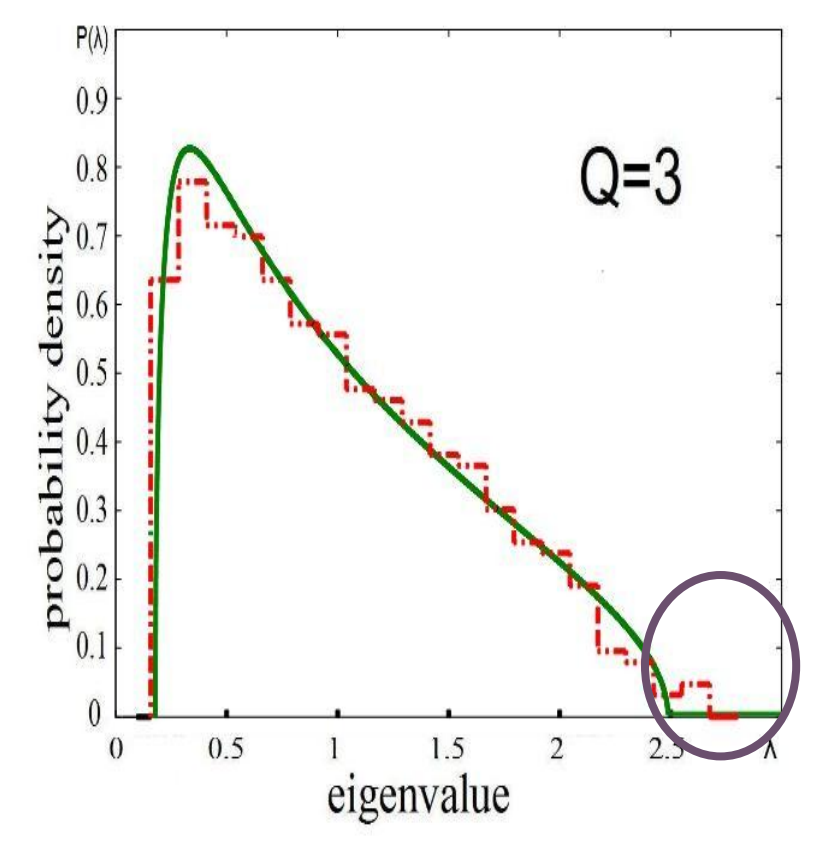
定量評価基準の決定

長さ75万：k=6に対し，X=3 と考えて良い

- kはモーメント次数:K次モーメント，Xは誤差範囲：Xパーセント。

定量評価基準の決定

初期乱数の乱数度評価結果： LCG



K	LCG(Q=3)	LCG(Q=6)
2	0.45%	0.57%
3	1.03%	1.41%
4	1.97%	2.51%
5	3.52%	3.92%
6	5.83%	5.71%

定量評価基準の決定

初期乱数の乱数度評価結果： LCG

6次モーメントの5%程度の誤差は
目視で判断できる

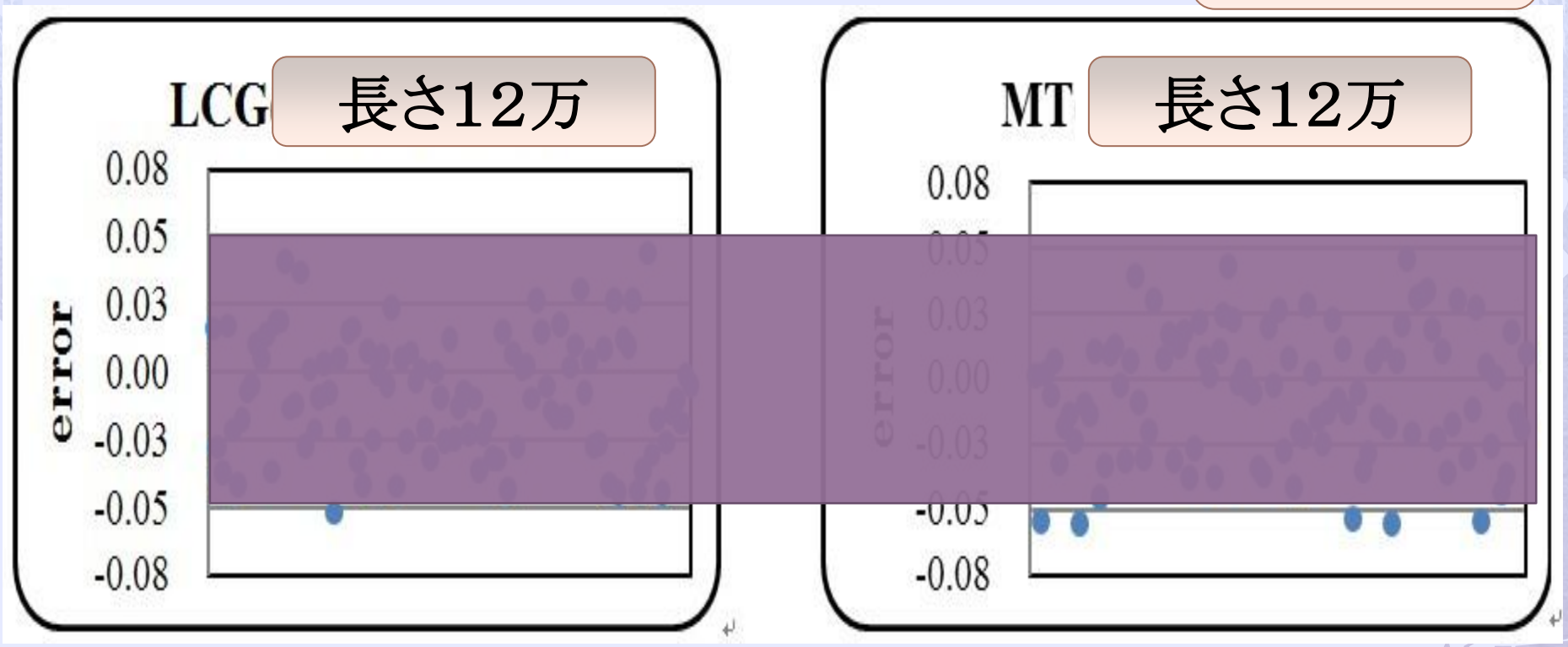
定量評価基準の決定

k=6, X=5を基準とした場合：

- 基準を満たす最短データ長は？
- データ長12万

定量評価基準の決定

$\alpha = 0.05$



- 6次モーメントの誤差：100サンプル

定量評価基準の決定

定量評価基準	k=6, X=3	k=6, X=5
乱数列の 長さ制限	75万以上	12万以上
目視で判断	できない	できる

定量評価基準の決定

定量評価基準	k=6, X=5
乱数列の長さ制限	12万以上
目視で判断	できる

制限の弱い方を選択

応　用

- 暗号学的ハッシュ関数の出力データのランダム性を評価
- 株価tickデータの乱数度と株価変動の関係を調査

応　用⇒ハッシュ値

- 対象:暗号学的ハッシュ関数MD5とSHA-1
- 目的:RMTテストを用いてMD5とSHA-1の安全性を判定できるかをチェック
- 手法:MD5とSHA-1の出力データのランダム性を評価
- 予想: SHA-1の安全性が高いことが既知なので，出力データのランダム性も高い

応　用⇒ハッシュ値

- データの処理

応　用⇒ハッシュ値

- 使用した乱数列：MD5とSHA-1の出力データ
- 乱数列長：12万(サンプル数：10)

応　用⇒ハッシュ値

- 評価結果

AV.(SD.)（N=200,Q=3,単位：%）

K	MD5(Q=3)	SHA-1(Q=3)
2	0.17(0.30)	0.03(0.16)
3	0.44(0.78)	0.06(0.39)
4	0.74(1.39)	0.05(0.69)
5	1.06(2.11)	0.01(1.06)
6	1.37(2.94)	0.11(1.54)

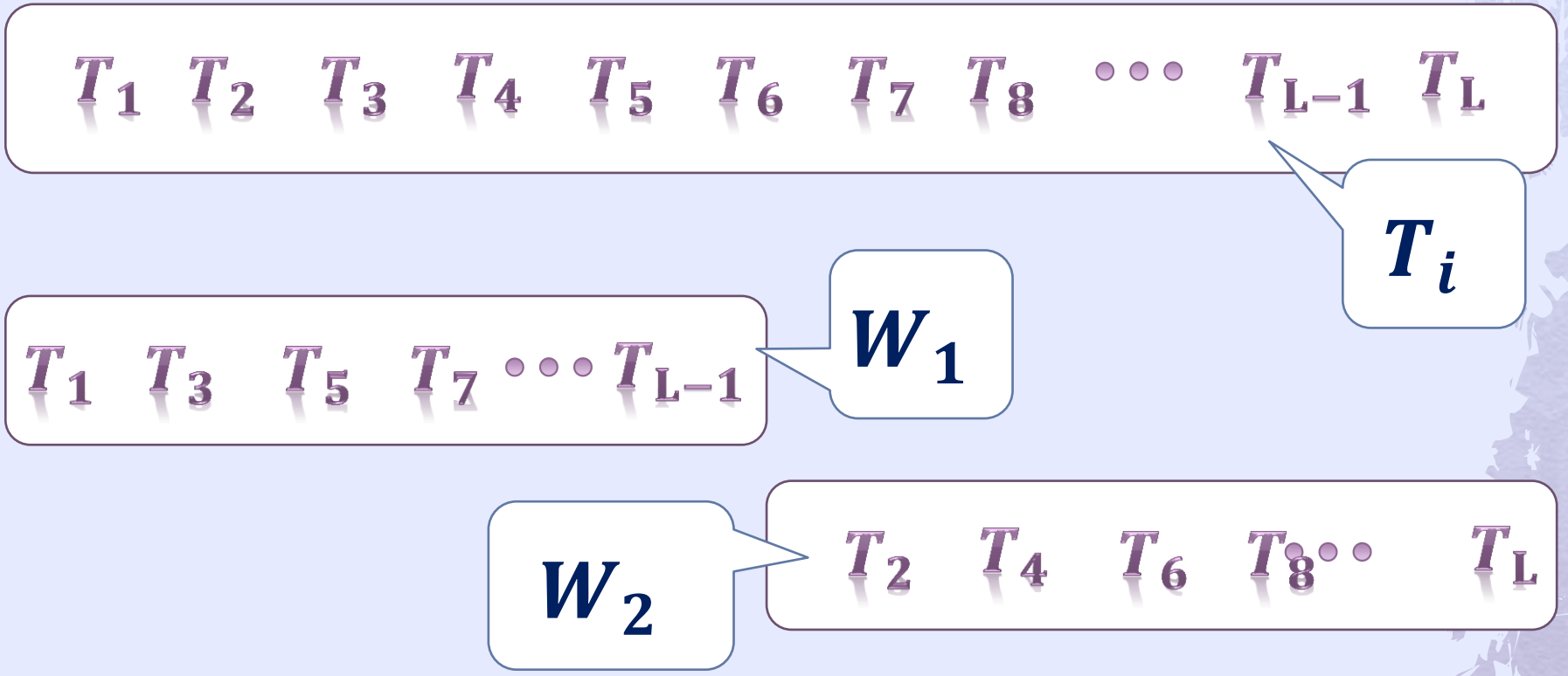
誤差の平均値と標準偏差どちらでもMD5より小さい

応　用⇒ハッシュ値

SHA-1の出力データのランダム性は
MD5より高い

応　用⇒乱数度と株

- 株価tickデータの対数収益時系列:T_i



応　用⇒乱数度と株

- W_1とW_2の6次モーメントの誤差平均値を計算
- 株価tickデータの乱数度を比較

K	2802		2809		2810		2897	
2	-0.30%	-0.04%	-0.45%	-0.21%	-0.12%	-0.28%	-0.22%	-0.60%
3	-0.66%	-0.10%	-1.44%	-0.54%	-0.52%	-0.83%	-0.30%	-1.51%
4	-0.82%	-0.05%	-2.82%	-0.88%	-1.16%	-1.53%	-0.10%	-2.53%
5	-0.68%	0.13%	-4.46%	-1.15%	-1.96%	-2.35%	0.38%	-3.56%
6	-0.20%	0.44%	-6.26%	-1.34%	-2.81%	-3.24%	1.06%	-4.54%
平均	0.32%		3.8%		3.025%		2.8%	

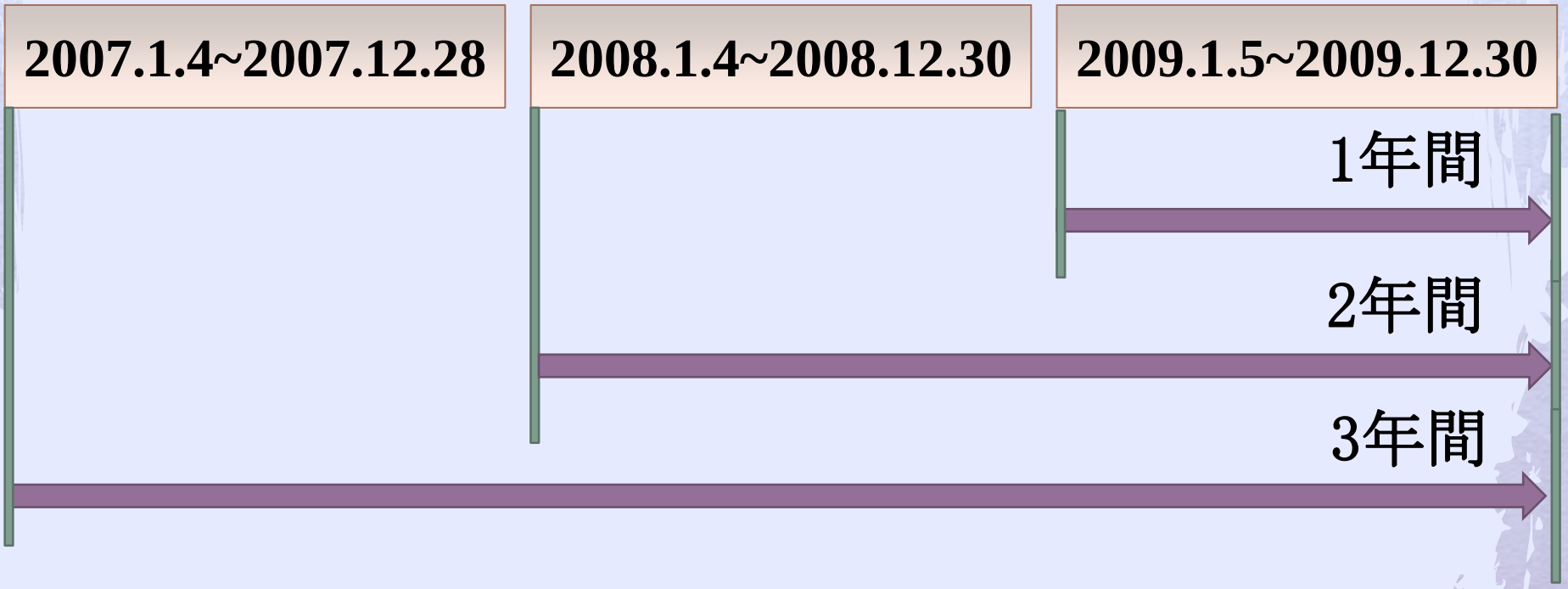
応　用⇒乱数度と株

- 目的:乱数度の視点から見て，乱数度が高い株と低い株のどちらを購入すべきか

- 対象:東証1部の10業種、39会社の株価
  tickデータ

2007年1月4日から
2009年12月30日まで
三年間の1分毎のtickデータ

応　用⇒乱数度と株


2007.1.4~2007.12.28
2008.1.4~2008.12.30
2009.1.5~2009.12.30
1年間
2年間
3年間

応　用⇒乱数度と株

食料品	鉄鋼	電気機器	輸送用機器	卸売業	小売業	銀行	陸運業	情報・通信	電気・ガス
2802	5401	6501	7201	8002	8227	8306	9020	9404	9501
2809	5405	6502	7202	8015	8233	8308	9021	9432	9503
2810	5406	6701	7203	8031	8252	8316	9022	9433	9506
2897	5423	6752	7211	8058	8267	8411		9437	9508

応　用⇒乱数度と株

1
- 2010.1.4の時点で株を購入
- その後の4時点の対数収益を計算(例：2010.2.26)

2
- 業種毎の乱数度が一番高い株と一番低い株を抽出

3
- 乱数度と株損益の関係を調査

応　用⇒乱数度と株(1年間)

1

・2010.1.4の時点で各株を購入
・それ以後の時点の対数収益を計算(例:2010.2.26)

食料品	鉄鋼	電気機器	輸送用機器	卸売業	小売業	銀行	陸運業	情報・通信	電気・ガス
2802	5401	6501	7201	8002	8227	8306	9020	9404	9501
2809	5405	6502	7202	8015	8233	8308	9021	9432	9503
2810	5406	6701	7203	8031	8252	8316	9022	9433	9506
2897	5423	6752	7211	8058	8267	8411		9437	9508

応　用⇒乱数度と株(1年間)

1

2010.2.26の時点，対数収益による業種毎の株の損益順位

食料品	鉄鋼	電気機器	輸送用機器	卸売業	小売業	銀行	陸運業	情報・通信	電気・ガス
2802	5401	6501	7202	8002	8267	8316	9022	9437	9506
2897	5423	6701	7211	8031	8233	8308	9020	9432	9501
2809	5405	6752	7201	8015	8252	8306	9021	9404	9508
2810	5406	6502	7203	8058	8227	8411		9433	9503

応　用⇒乱数度と株(1年間)

2

・業種毎の乱数度が一番高い株を抽出

食料品	鉄鋼	電気機器	輸送用機器	卸売業	小売業	銀行	陸運業	情報・通信	電気・ガス
2802	5401	6501	7202	8002	8267	8316	9022	9437	9506
2897	5423	6701	7211	8031	8233	8308	9020	9432	9501
2809	5405	6752	7201	8015	8252	8306	9021	9404	9508
2810	5406	6502	7203	8058	8227	8411		9433	9503

応　用⇒乱数度と株(1年間)

2

・業種毎の乱数度が一番低い株を抽出

食料品	鉄鋼	電気機器	輸送用機器	卸売業	小売業	銀行	陸運業	情報・通信	電気・ガス
2802	5401	6501	7202	8002	8267	8316	9022	9437	9506
2897	5423	6701	7211	8031	8233	8308	9020	9432	9501
2809	5405	6752	7201	8015	8252	8306	9021	9404	9508
2810	5406	6502	7203	8058	8227	8411		9433	9503

応　用⇒乱数度と株(1年間)

3

・乱数度と株損益の関係を調査(乱数度が一番高い)

食料品	鉄鋼	電気機器	輸送用機器	卸売業	小売業	銀行	陸運業	情報・通信	電気・ガス
2802	5401	6501	7202	8002	8267	8316	9022	9437	9506
2897	5423	6701	7211	8031	8233	8308	9020	9432	9501

乱数度が一番高い株の80%（10社の内に8社）の損益は2位以上

応　用⇒乱数度と株(1年間)

3

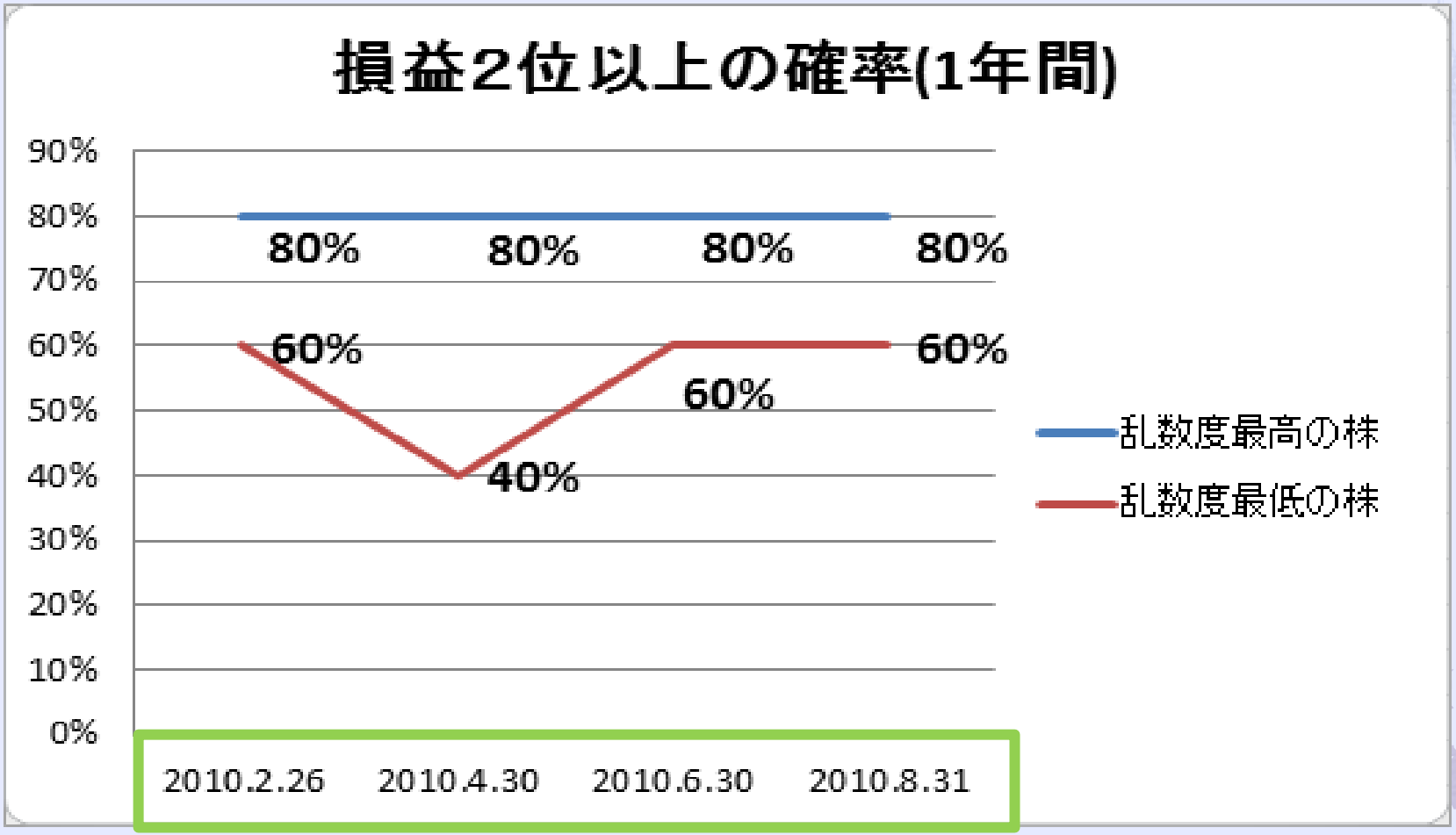
・乱数度と株損益の関係を調査(乱数度が一番低い)

食料品	鉄鋼	電気機器	輸送用機器	卸売業	小売業	銀行	陸運業	情報・通信	電気・ガス
2802	5401	6501	7202	8002	8267	8316	9022	9437	9506
2897	5423	6701	7211	8031	8233	8308	9020	9432	9501

乱数度が一番低い株の60%（10社の内に6社）の損益は2位以上

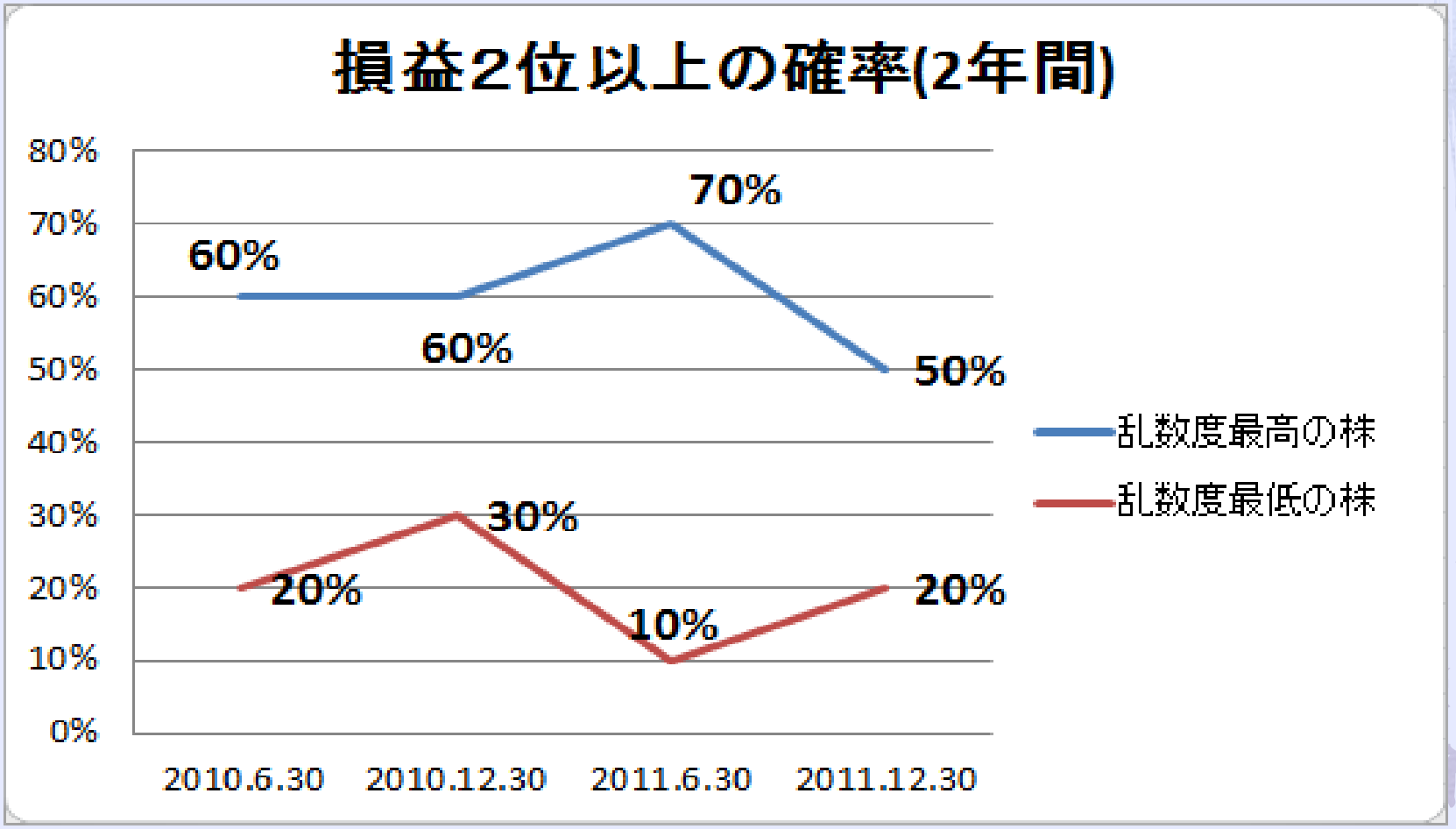
応　用⇒乱数度と株(1年間)

- 2009年1月5日～2009年12月30日



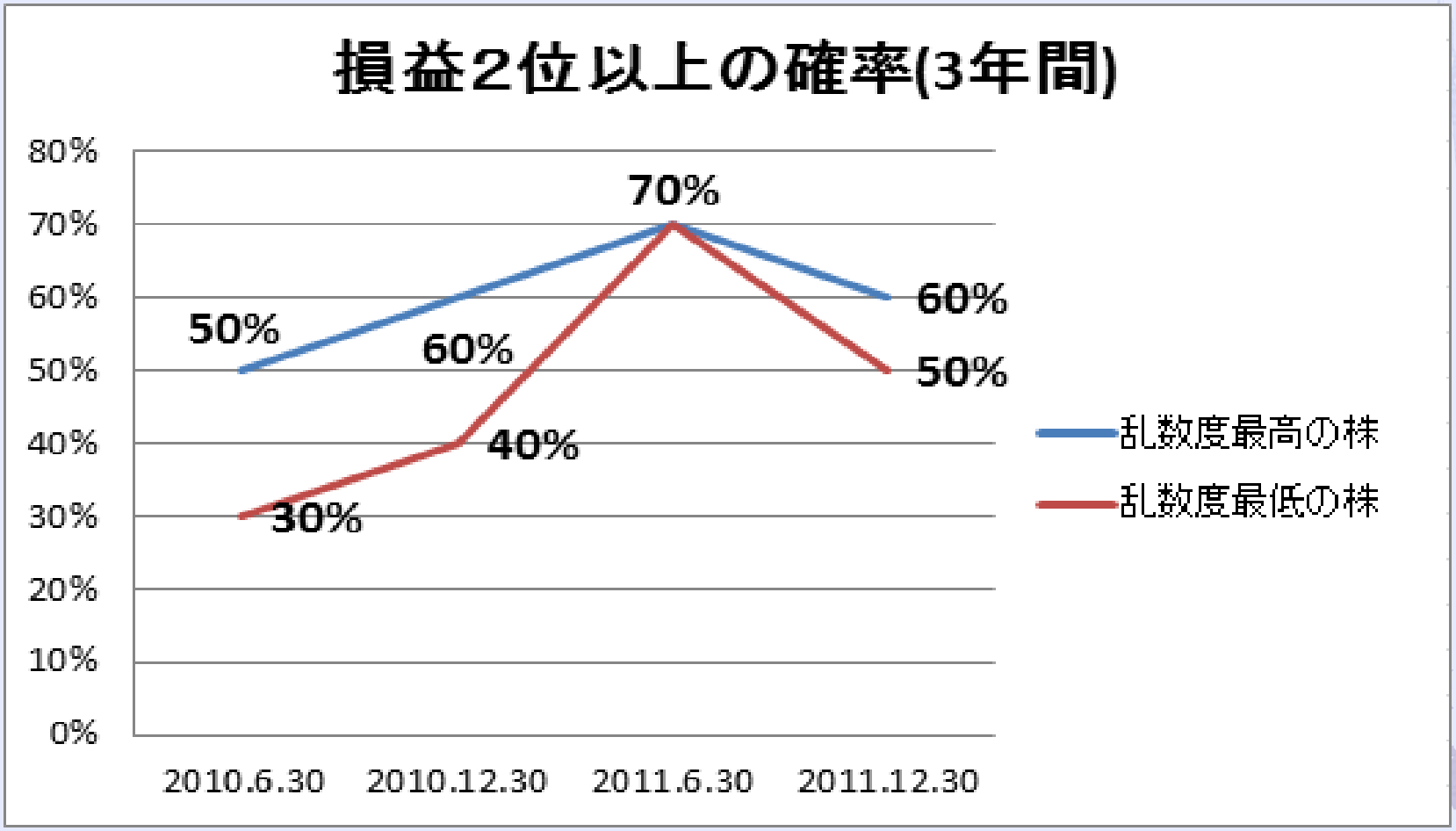
応　用⇒乱数度と株(2年間)

- 2008年1月4日～2009年12月30日



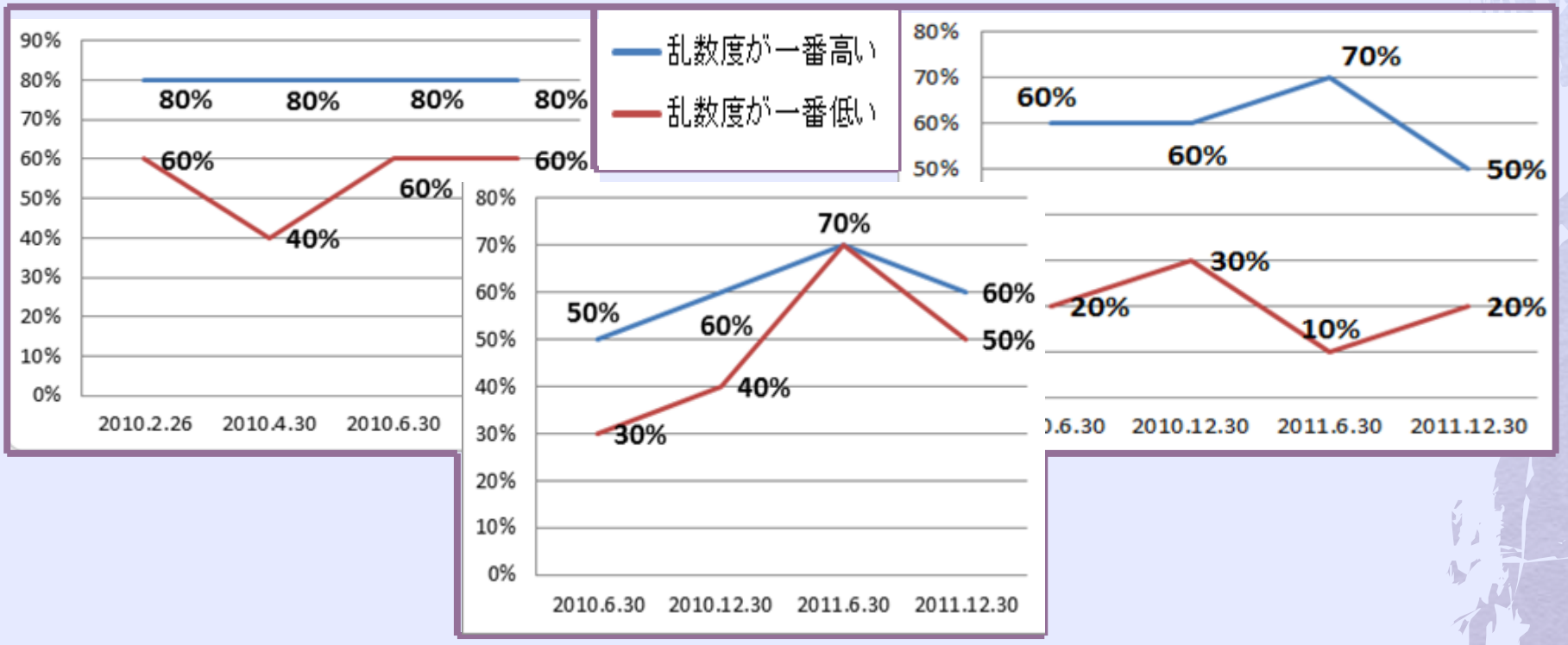
応　用⇒乱数度と株(3年間)

- 2007年1月4日～2009年12月30日



応　用⇒乱数度と株

- 乱数度が高い株を購入する方が安全



今後の課題

- 東証1部の全業種(33)を分析
- 業種内企業数を増やす
- 売る時点を増やす

今までの結論を検証する

まとめ

✓乱数度計測器としてRMTテストを紹介した

✓ RMTテストでハッシュ関数の安全性をチェックできた

✓ RMTテストにより株価データの乱数度を評価し，
乱数度が低い株より乱数度が高い株の安全性が高い