Rate distortion function in betting game system

Masayuki Kumon

Association for Promoting Quality Assurance in Statistics

Abstract

Among various aspects of game theoretic probability, when exploring mathematical structure of the optimal strategies in betting games, Kullback-Leibler divergence is naturally derived as the optimal exponential growth rate of the betting capital process.

This structure had been obtained by Prof. Takeuchi nearly fifty years ago.

Inspired by Claude Shannon's Information Theory, an optimizing betting strategy was also pioneered by John Larry Kelly Jr. in

> **A New Interpretation of Information Rate. *Bell System Technical Journal*, Vol.35, 917-926, 1956.**

The optimalities of Kelly's strategy

- Minimal expected time property
- Asymptotic largest magnitude property

were investigated by Leo Breiman **in**

> Optimal Gambling Systems for Favorable Games. *Fourth Berkeley Symposium on Probability and Statistics I*, 65-78, 1961.

The historical reviews and the recent developments concerning Kelly's strategy such as T. M. Cover's Universal Portfolio are presented in

> **L. C. MacLean, E. O. Thorp, W. T. Ziemba eds.** ***The Kelly Capital Growth Investment Criterion : Theory and Practice.*** **Handbook in Finantial Economic Series, Vol.3, World Scientific, London, 2010.**

In this talk, the following are addressed.

- **Game mutual information which measures information transmission between betting games is introduced.**
- **Two characteristics Game channel capacity and Game rate distortion function are defined from the mutual information, and these meanings are explained.**

- **The effect of the optimal strategy in conditional betting game is verified by using real stock price data.**
- **As an application of Game rate distortion function, an efficient lossy source coding scheme based on the optimal conditional betting strategy is proposed.**

1. Mutual information in betting game system

1.1 Definition of mutual information

- **Mutual information in information theory**

$$X \sim P_X(x) \quad Y \sim P_Y(y) \quad (X, Y) \sim P_{XY}(x, y)$$

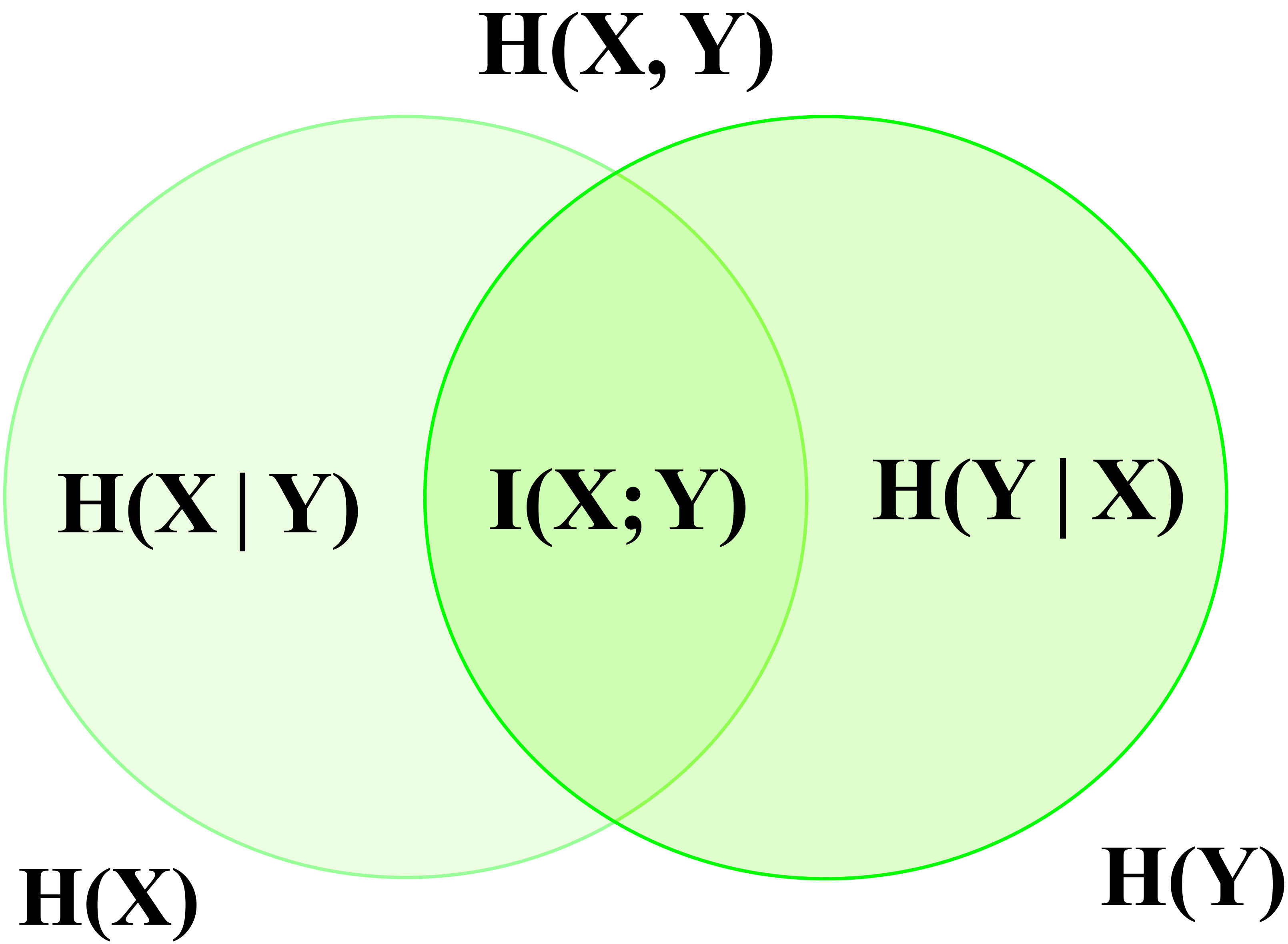
$$H(X) = -E_{P_X}[\log P_X(X)] \quad \text{etc.}$$

- **Shannon's source coding theorem :**

Entropy $H(X)$ is the nearly achievable lower bound on the average length of the shortest description of the random variable X.

$$
\begin{aligned}
I(X;Y) = H(X) - H(X|Y) &= H(Y) - H(Y|X) \\
&= H(X) + H(Y) - H(X,Y) \\
&= D(P_{XY} \| P_X P_Y) \geq 0
\end{aligned}
$$

$$I(X;Y) = 0 \;\Leftrightarrow\; P_{XY}(x,y) = P_X(x)P_Y(y)$$


H(X, Y)
H(X | Y)
I(X; Y)
H(Y | X)
H(X)
H(Y)

mutual information in betting games

$A,\ B$ ~ two betting games

C ~ joint betting game of A and B

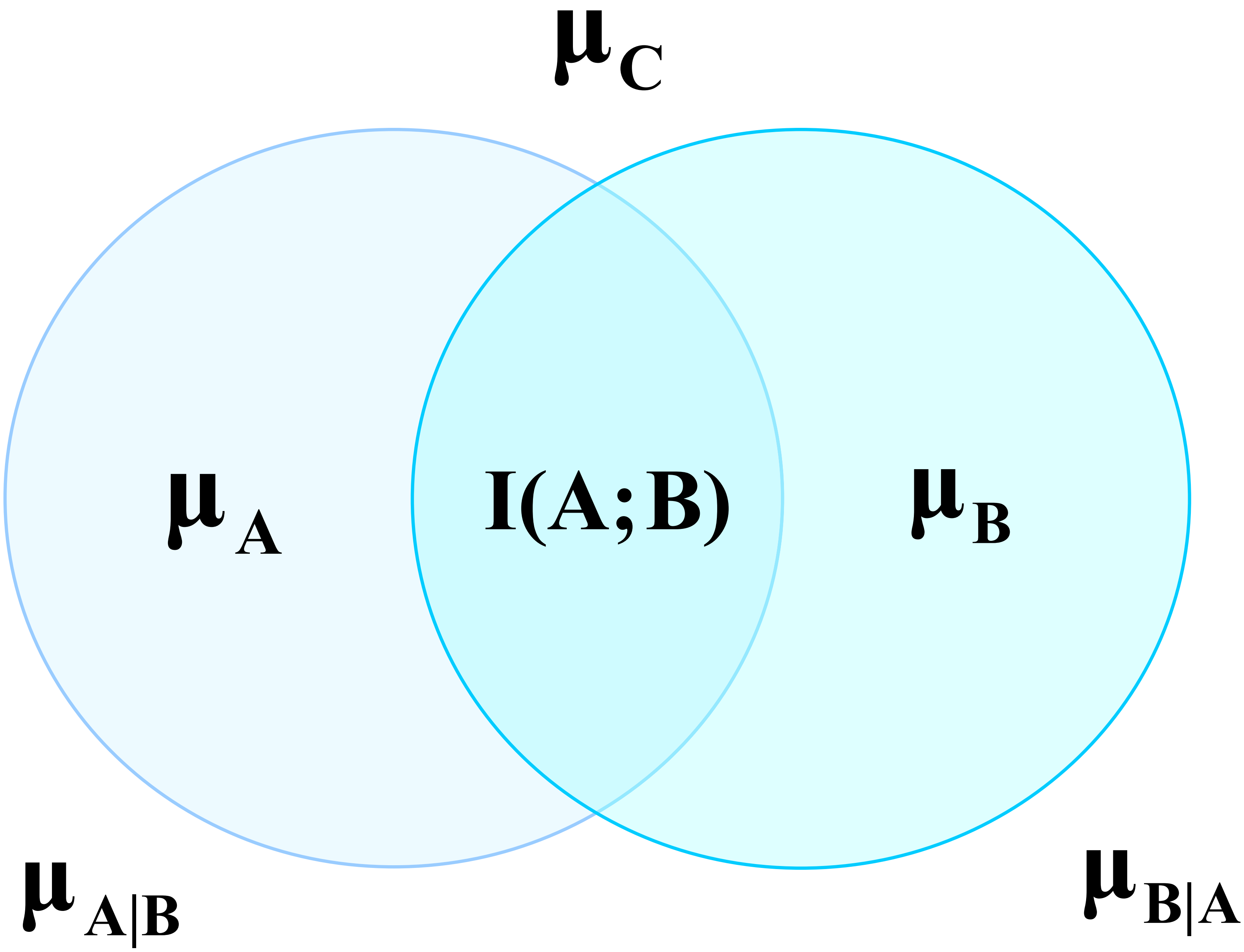
$P_A,\ P_B,\ P_C$: empirical prob. of Reality

$Q_A,\ Q_B,\ Q_C$: risk neutral prob. of Forecaster

$\mu_A := D(P_A \| Q_A)$: quantity of the game A

$\mu_B := D(P_B \| Q_B)$: quantity of the game B

$\mu_C := D(P_C \| Q_C)$: quantity of the game C


μ C
μ A
I(A;B)
μ B
μ A|B
μ B|A

$$I(A;B) := \mu_{B|A} - \mu_B = \mu_{A|B} - \mu_A$$
$$= \mu_C - (\mu_A + \mu_B)$$

$\because$ $\mu_C = \mu_A + \mu_{B|A} = \mu_B + \mu_{A|B}$ (additivity)

$$\mu_{B|A} := \mu_C - \mu_A = D(P_{B|A} \| Q_{B|A} | P_A)$$

$\mu_{B|A}$: **quantity of the conditional betting game $B|A$ given A**

$D(P_{B|A} \| Q_{B|A} | P_A)$:

conditional K-L divergence between $P_{B|A}$ and $Q_{B|A}$ given P_A

Decomposition of $I(A;B)$

$$I(A;B) = I_1(A;B) - I_2(A;B)$$

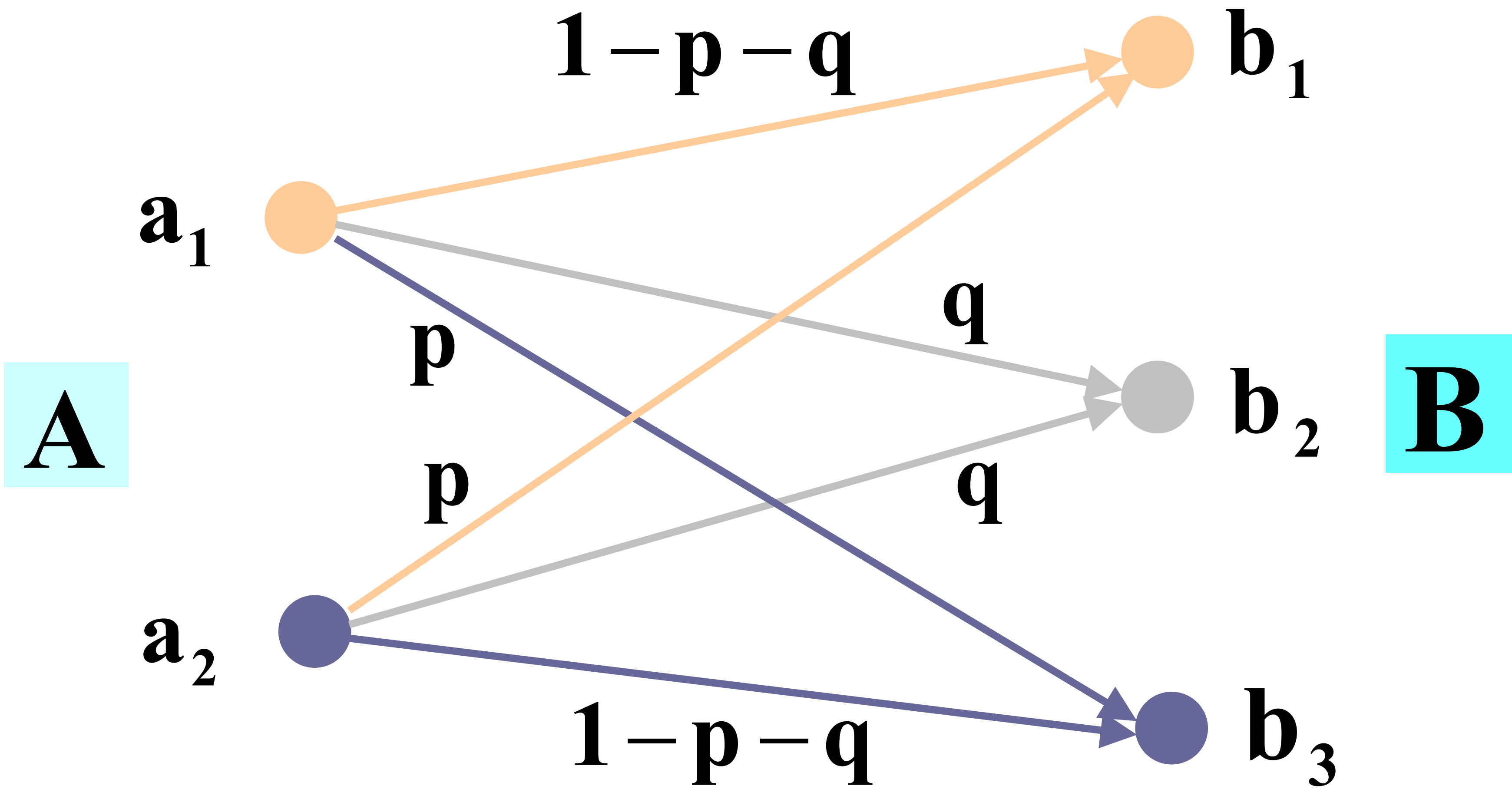
$$I_1(A;B) = D(P_C \| P_A P_B) \geq 0 \ \ :$$

usual mutual information between P_A and P_B

$$I_2(A;B) = E_{P_C}\left[\log \frac{Q_C(X,Y)}{Q_A(X)Q_B(Y)}\right]$$

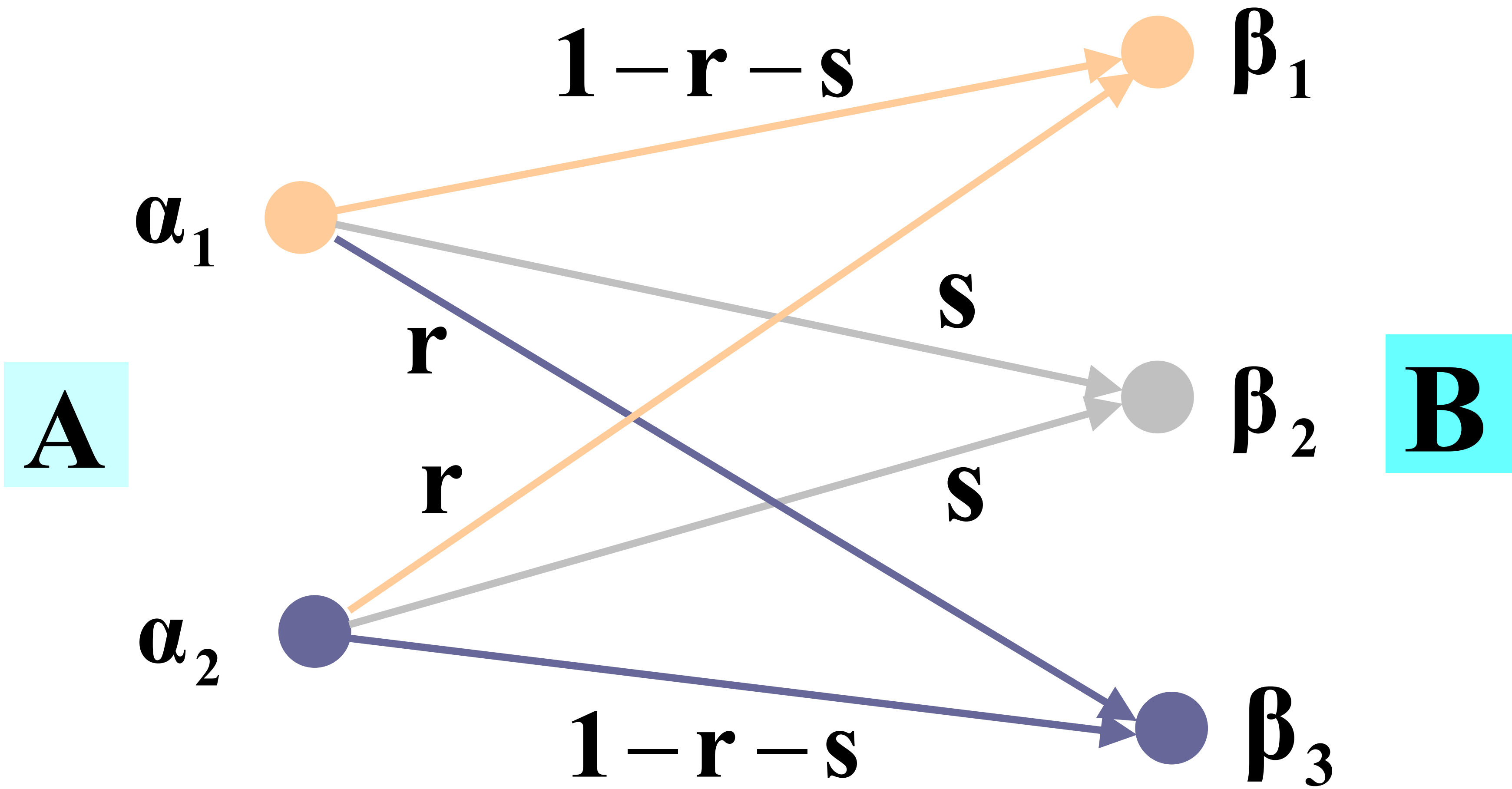
$$Q_C(x,y) = Q_A(x)Q_B(y) \ \Rightarrow \ I_2(A;B) = 0$$

Reality's move



Binary symmetric erasure channel

Forecaster's move



Binary symmetric erasure channel

$$P_{B|A}(y|x) = \delta_{xy} \quad Q_{B|A}(y|x) = P_B(y) \quad \mathcal{X} = \mathcal{Y}$$

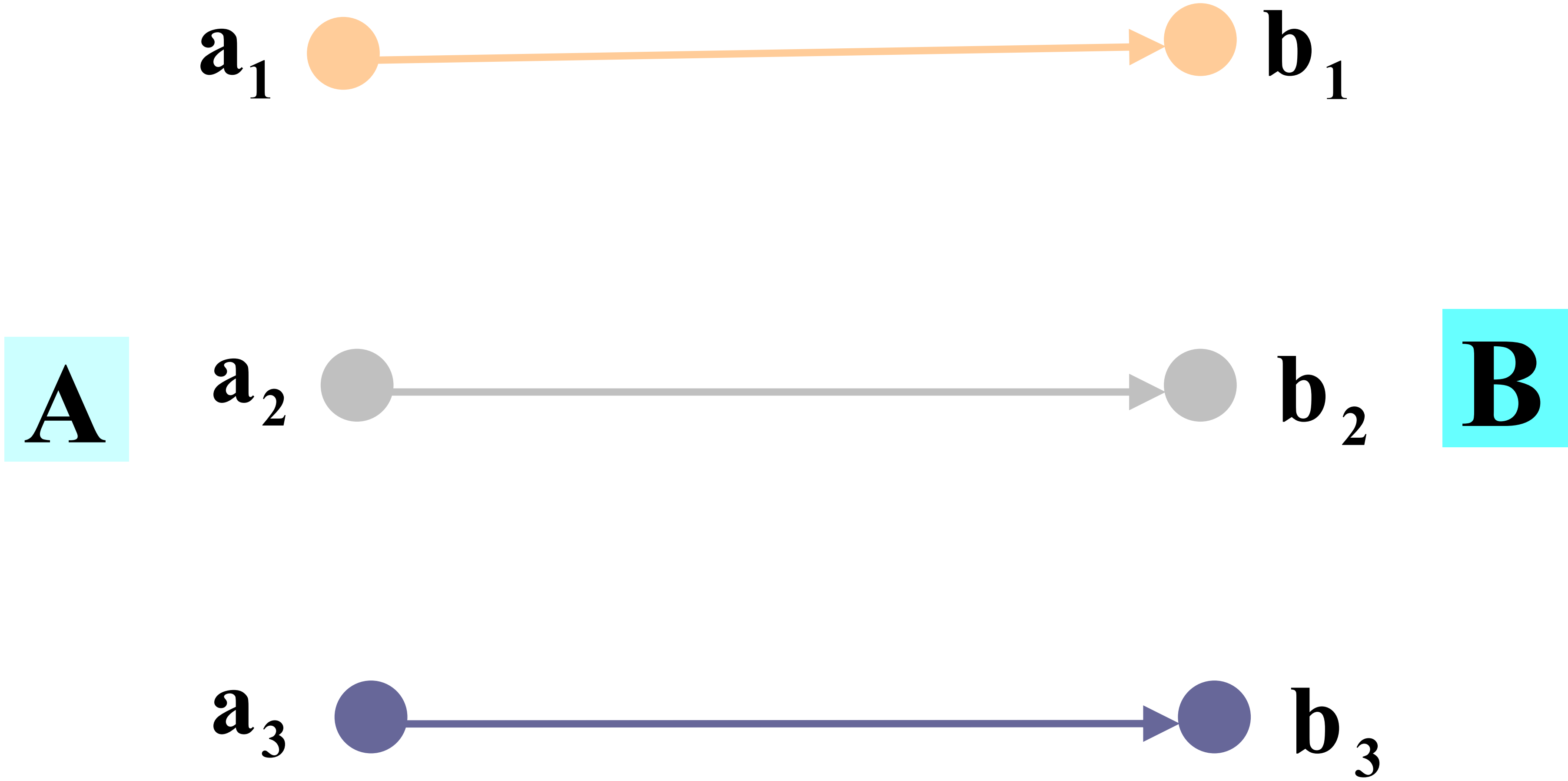
$$\Rightarrow \; D(P_{B|A} \| Q_{B|A} | P_A)$$

$$= \sum_{x \in \mathcal{X}} P_A(x) \sum_{y \in \mathcal{Y}} P_{B|A}(y|x) \log \frac{P_{B|A}(y|x)}{Q_{B|A}(y|x)}$$

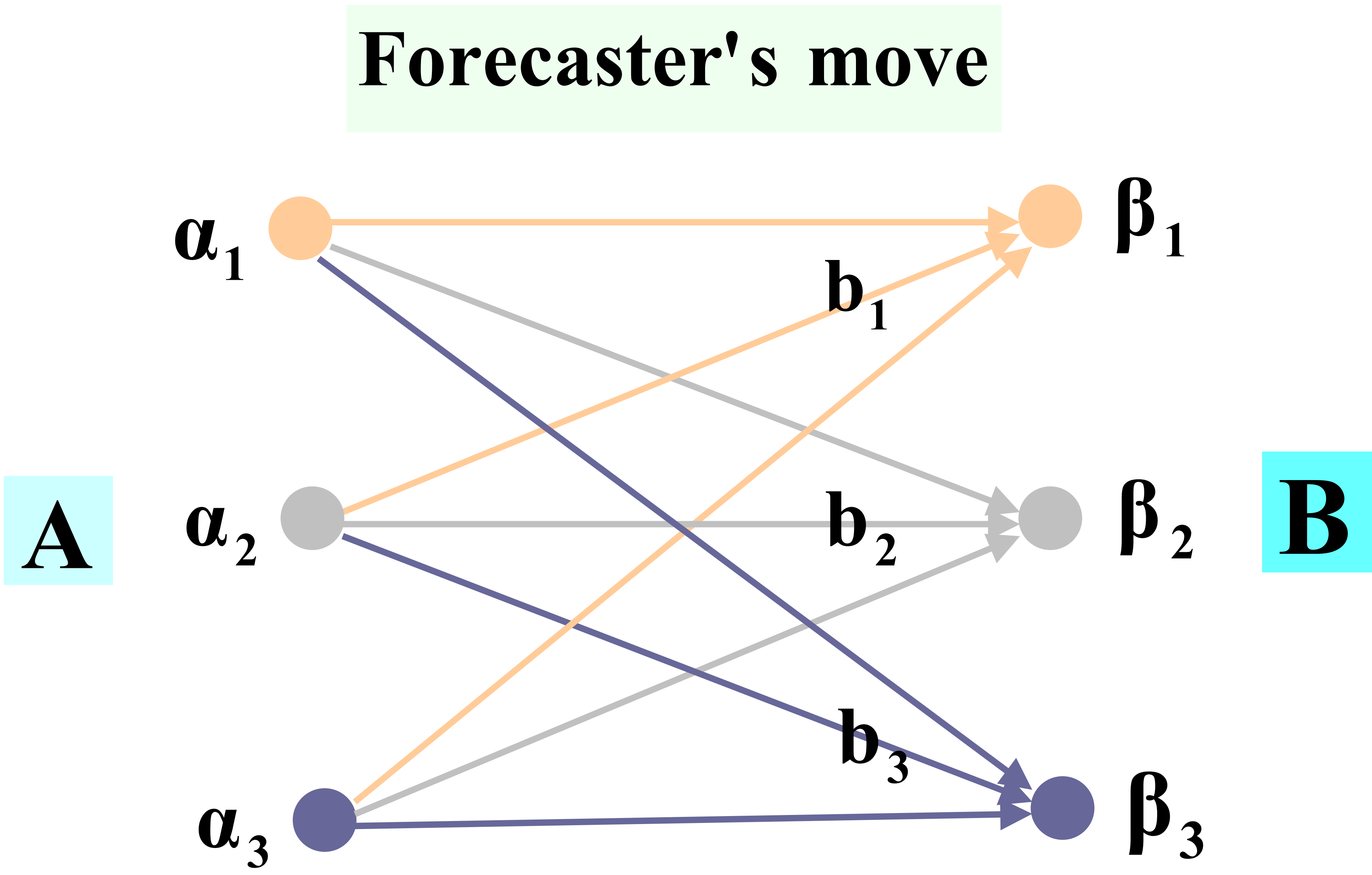
$$= \sum_{x \in \mathcal{X}} P_A(x) \sum_{y \in \mathcal{Y}} \delta_{xy} \log \frac{\delta_{xy}}{P_B(y)}$$

$$= -\sum_{x \in \mathcal{X}} P_A(x) \log P_A(x) = H(X)$$

Reality's move



Entropy channel



Entropy channel

1.2 Game channel capacity

■ **Channel capacity in information theory**

$C = \sup_{P_X} I(X;Y)$: **capacity of channel** $X \Rightarrow Y$

• **Shannon's channel coding theorem :**

Capacity C **is the supremum of rates** R **at which information can be sent with arbitrarily low probability of error.**

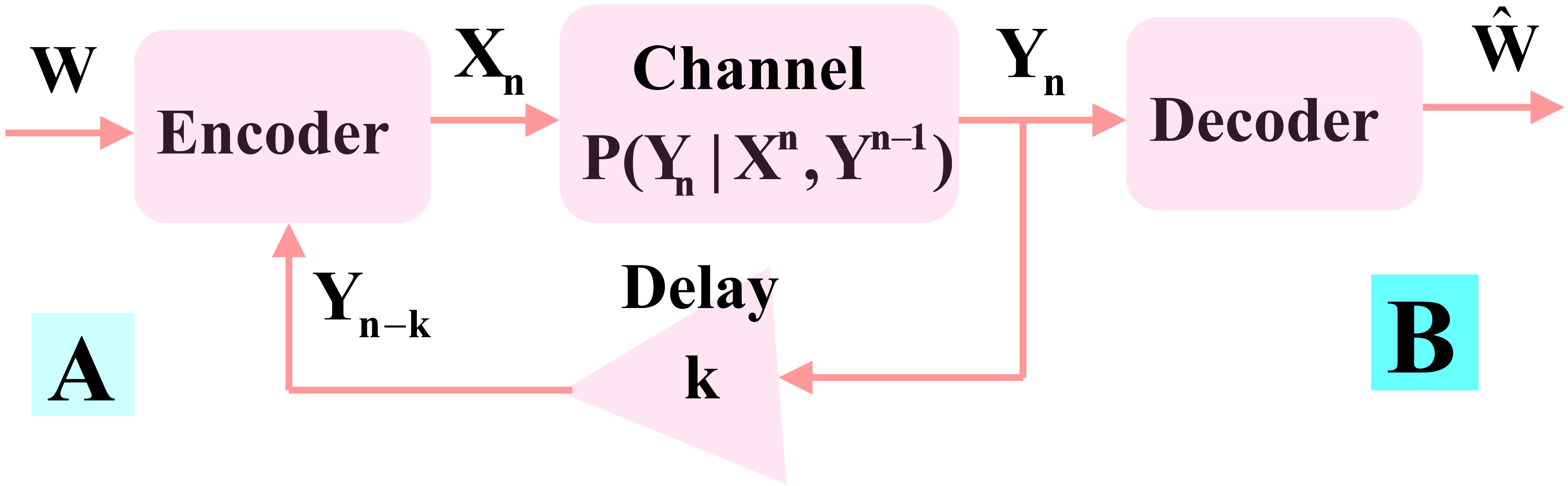
■ Channel capacity in betting games

$$C_g := \sup_{P_A, Q_A = P_A} I(A;B) \ :$$

capacity of betting game channel $A \Rightarrow B$

$$I(A;B) = \mu_{B|A} - \mu_B = \mu_{A|B} - \mu_A$$

$$C_g = \sup_{P_A} \ \mu_{A|B} = \sup_{P_A} \ D(P_{A|B} \| Q_{A|B} | P_B) \geq 0$$



$$C_g = \sup_{P_A} D(P_{A|B} \| Q_{A|B} \mid P_B)$$

$$C = \sup_{P_X} H(Y) - H(Y \mid X)$$

Communication channel with feedback

1.3 Game rate distortion function

■ **Rate distortion function in information theory**

$$R(D) = \inf_{P_{\hat{X}|X} : E_{P_{X\hat{X}}} d(X,\hat{X}) \le D} I(X;\hat{X}) \quad :$$

Rate distortion function of transmission $X \Rightarrow \hat{X}$

- **Shannon's rate distortion theorem :**

Rate distortion function $R(D)$ is the infimum of rates R that asymptotically achieve the distortion D.

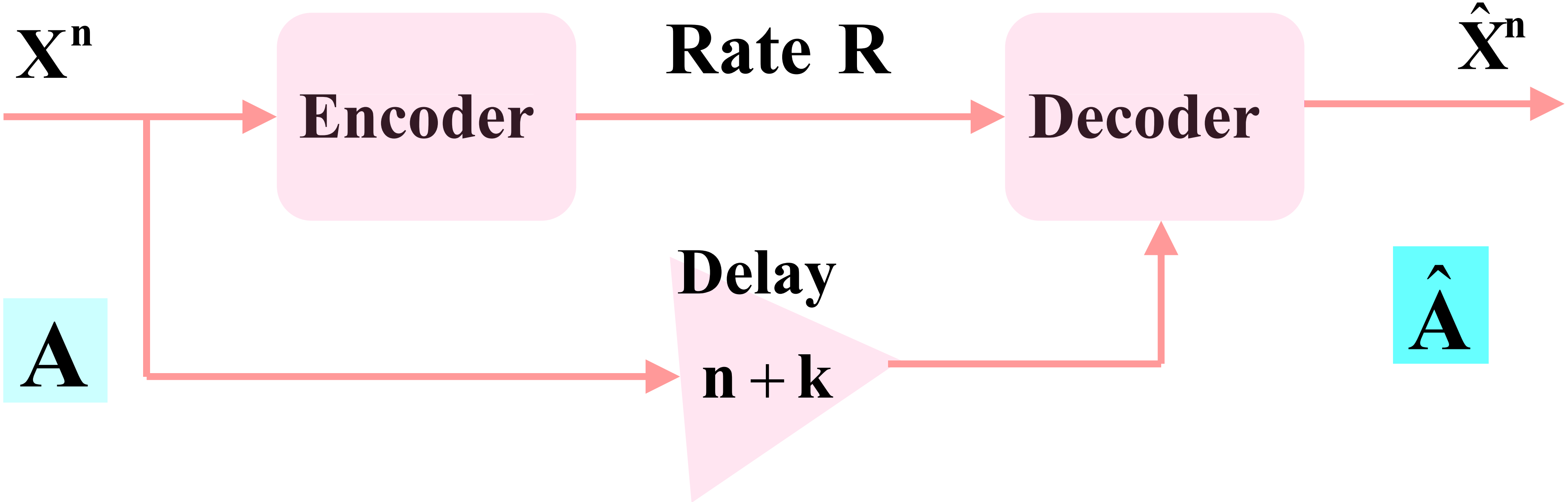
- **Rate distortion function in betting games**

$$R_g(D) := \inf_{P_{\hat{A}|A}: E_{P_{A\hat{A}}} d(X,\hat{X}) \leq D, Q_{\hat{A}|A}: Q_{\hat{A}} = P_{\hat{A}}} I(A;\hat{A}) \quad :$$

Rate distortion function of transmission $A \Rightarrow \hat{A}$

$$I(A;\hat{A}) = \mu_{A|\hat{A}} - \mu_A = \mu_{\hat{A}|A} - \mu_{\hat{A}}$$

$$R_g(D) = \inf_{P_{\hat{A}|A}} \mu_{\hat{A}|A} = \inf_{P_{\hat{A}|A}} D(P_{\hat{A}|A} \| Q_{\hat{A}|A} | P_A) \geq 0$$



$$R_g(D) = \inf_{P_{\hat{A}|A}} D(P_{\hat{A}|A} \| Q_{\hat{A}|A} \mid P_A)$$

$$R(D) = \inf_{P_{\hat{X}|X}} H(X) - H(X \mid \hat{X})$$

Rate distortion with feedforward

2. Optimal conditional betting strategy

2.1 Optimal limit order strategy (cf. [6])

- **Investor** selects $\delta > 0$ and sequentially decides the trading times $0 < t_1 < t_2 < \cdots$ as follows.

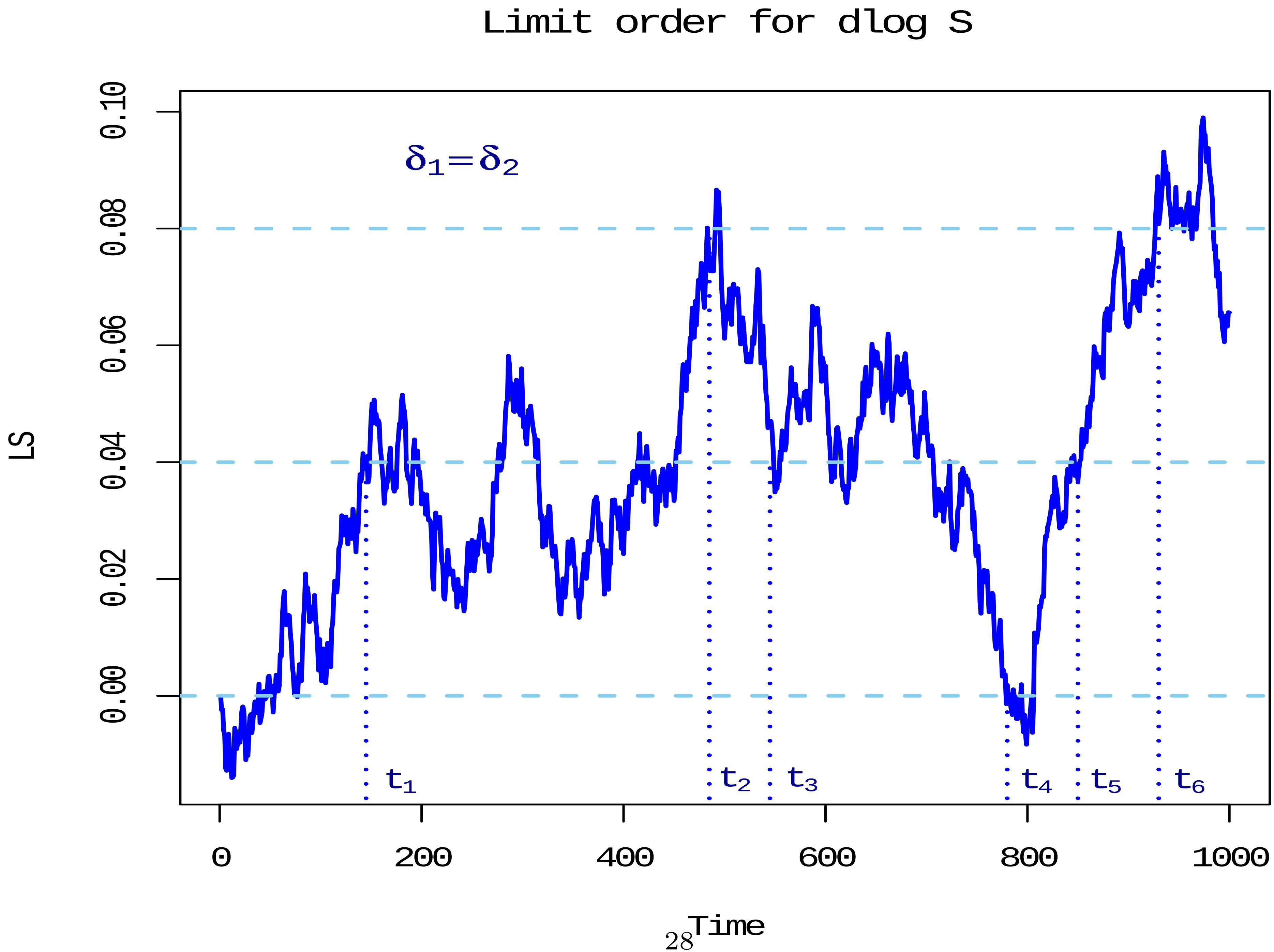
$S(t) > 0$: continuous asset price of Market

t_{i+1} : first time after t_i such that

$$\frac{S(t_{i+1})}{S(t_i)} = 1 + \delta \quad \text{or} \quad \frac{1}{1+\delta}$$

$$\Leftrightarrow \quad \log S(t_{i+1}) - \log S(t_i) = \eta \quad \text{or} \quad -\eta$$

$$\eta = \log(1+\delta)$$


Limit order for dlog S
δ1=δ2
LS
0.00
0.02
0.04
0.06
0.08
0.10
t1
t2
t3
t4
t5
t6
0
200
400
600
800
1000
Time

Embedded Coin-Tossing Game

$\mathcal{K}_0 := 1$

FOR $n = 1, 2, \ldots$:

Investor announces $\alpha_n \in \mathbb{R}$

Market announces $x_n \in \{0, 1\}$

$\mathcal{K}_n = \mathcal{K}_{n-1}(1 + \alpha_n(x_n - \rho))$

END FOR

$\rho = \dfrac{1}{2 + \delta}$: risk neutral prob. set by Investor

- **Notations**

$\chi_1^n,\ \chi_0^n$: number of $x_i = 1,\ 0\ \ (i = 1. \ldots, n)$

$$P(x^n) = \frac{B(\chi_1^n + c_1, \chi_0^n + c_0)}{B(c_1, c_0)} \quad x^n = x_1 \cdots x_n$$

$$B(c_1, c_0) = \frac{\Gamma(c_1)\Gamma(c_0)}{\Gamma(c_1 + c_0)} \quad c_1, c_0 > 0\ :$$

beta binomial distribution modeled by Investor

Maximize $E_P[\log \mathcal{K}_n] \ \Rightarrow \ \{\alpha_i\}_{i=1}^n$

$$\alpha_i = \frac{p_i - \rho}{\rho(1-\rho)} \quad i = 1, \ldots, n$$

$$p_i = P(x_i = 1 | x^{i-1}) = \frac{\chi_1^{i-1} + c_1}{i - 1 + c_1 + c_0}$$

The optimal capital process of Investor is expressed as the likelihood ratio

$$\mathcal{K}_n = \frac{P(x^n)}{Q(x^n)} = \frac{B(\chi_1^n + c_1, \chi_0^n + c_0)/B(c_1, c_0)}{\rho^{\chi_1^n}(1-\rho)^{\chi_0^n}}$$

From the Stirling's formula

$$\log \mathcal{K}_n = nD\left(\hat{p}_n \| q\right) - \frac{1}{2}\log n + O(1)$$

$\hat{p}_n = \left(\frac{\chi_1^n}{n}, \frac{\chi_0^n}{n}\right)$ **: empirical prob. by Market**

$q = (\rho, 1-\rho)$ **: risk neutral prob. by Investor**

$D\left(\hat{p}_n \| q\right)$ **: empirical K-L divergence**

2.2 Optimal conditional limit order strategy (cf. [7])

■ Investor determines the betting ratios $\alpha_n \in \mathbb{R}$ of conditional betting game $B|A$ given A as follows.

$$\alpha_1 = 0, \quad \alpha_n = \begin{cases} \alpha_n^+ & \text{if } \; x_n = 1 \\ \alpha_n^- & \text{if } \; x_n = 0 \end{cases} \qquad n = 2, 3, \ldots$$

- **Notations**

$\chi^n_{x1},\ \chi^n_{x0}$: **number of** $x_i = 1,\ 0\ \ (i = 1.\dots, n)$

$\chi^n_{11},\ \chi^n_{10},\ \chi^n_{01},\ \chi^n_{00}$:

number of $(x_i, y_i) = (1,1),\ (1,0),\ (0,1),\ (0,0)$

$(i = 1, \dots, n)$

$$P^+(y^n|x^n) = \frac{B(\chi^n_{11} + c_1, \chi^n_{10} + c_0)}{B(c_1, c_0)}$$

$$P^-(y^n|x^n) = \frac{B(\chi^n_{01} + c_1, \chi^n_{00} + c_0)}{B(c_1, c_0)} \quad :$$

beta binomial distribution modeled by Investor

Maximize $E_P[\log \mathcal{K}_n]$ $\quad P = P^+ \times P^- \;\Rightarrow\; \{\alpha_i^{\pm}\}_{i=2}^{n}$

$$\alpha_i^+ = \frac{p_i^+ - \rho}{\rho(1-\rho)} \qquad \alpha_i^- = \frac{p_i^- - \rho}{\rho(1-\rho)}$$

$$p_i^+ = P^+(y_i = 1 | x^{i-1}) = \frac{\chi_{11}^{i-1} + c_1}{\chi_{11}^{i-1} + \chi_{10}^{i-1} + c_1 + c_0}$$

$$p_i^- = P^-(y_i = 1 | x^{i-1}) = \frac{\chi_{01}^{i-1} + c_1}{\chi_{01}^{i-1} + \chi_{00}^{i-1} + c_1 + c_0}$$

The optimal capital process of Investor is expressed as the likelihood ratio

$$\mathcal{K}_n = \mathcal{K}_n^+ \times \mathcal{K}_n^- \quad \xi^n = (x_1, y_1) \cdots (x_n, y_n)$$

$$\mathcal{K}_n^+ = \frac{P^+(\xi^n)}{Q^+(\xi^n)} = \frac{B(\chi_{11}^n + c_1, \chi_{10}^n + c_0)/B(c_1, c_0)}{\rho^{\chi_{11}^n}(1-\rho)^{\chi_{10}^n}}$$

$$\mathcal{K}_n^- = \frac{P^-(\xi^n)}{Q^-(\xi^n)} = \frac{B(\chi_{01}^n + c_1, \chi_{00}^n + c_0)/B(c_1, c_0)}{\rho^{\chi_{01}^n}(1-\rho)^{\chi_{00}^n}}$$

$$\log \mathcal{K}_n = nD\left(\hat{p}_{n,y|x} \| q | \hat{p}_{n,x}\right) - \frac{1}{2}\left(\log \chi^n_{x1} + \log \chi^n_{x0}\right) + O(1)$$

$$\hat{p}_{n,y|1} = \left(\frac{\chi^n_{11}}{\chi^n_{x1}}, \frac{\chi^n_{10}}{\chi^n_{x1}}\right) \quad \hat{p}_{n,y|0} = \left(\frac{\chi^n_{01}}{\chi^n_{x0}}, \frac{\chi^n_{00}}{\chi^n_{x0}}\right) \ :$$

empirical conditional prob. by Market

$$q = (\rho, 1-\rho) \ :$$

risk neutral prob. by Investor

$$D\left(\hat{p}_{n,y|x} \| q | \hat{p}_{n,x}\right) \quad \hat{p}_{n,x} = \left(\frac{\chi^n_{x1}}{n}, \frac{\chi^n_{x0}}{n}\right) \ :$$

empirical conditional K-L divergence

- **In the following figures**

  $S_A(t)$: **daily closing prices of Nikkei 225**

  $S_B(t)$: **daily opening prices of**

  Toyota Sony Nintendo

  2003/1/6 - 2007/9/28

  $\mathcal{K}_n$: **capital process of Investor**

$\mathrm{pn1} = \hat{p}_{n,1|1}$ $\mathrm{pn0} = \hat{p}_{n,0|0}$:

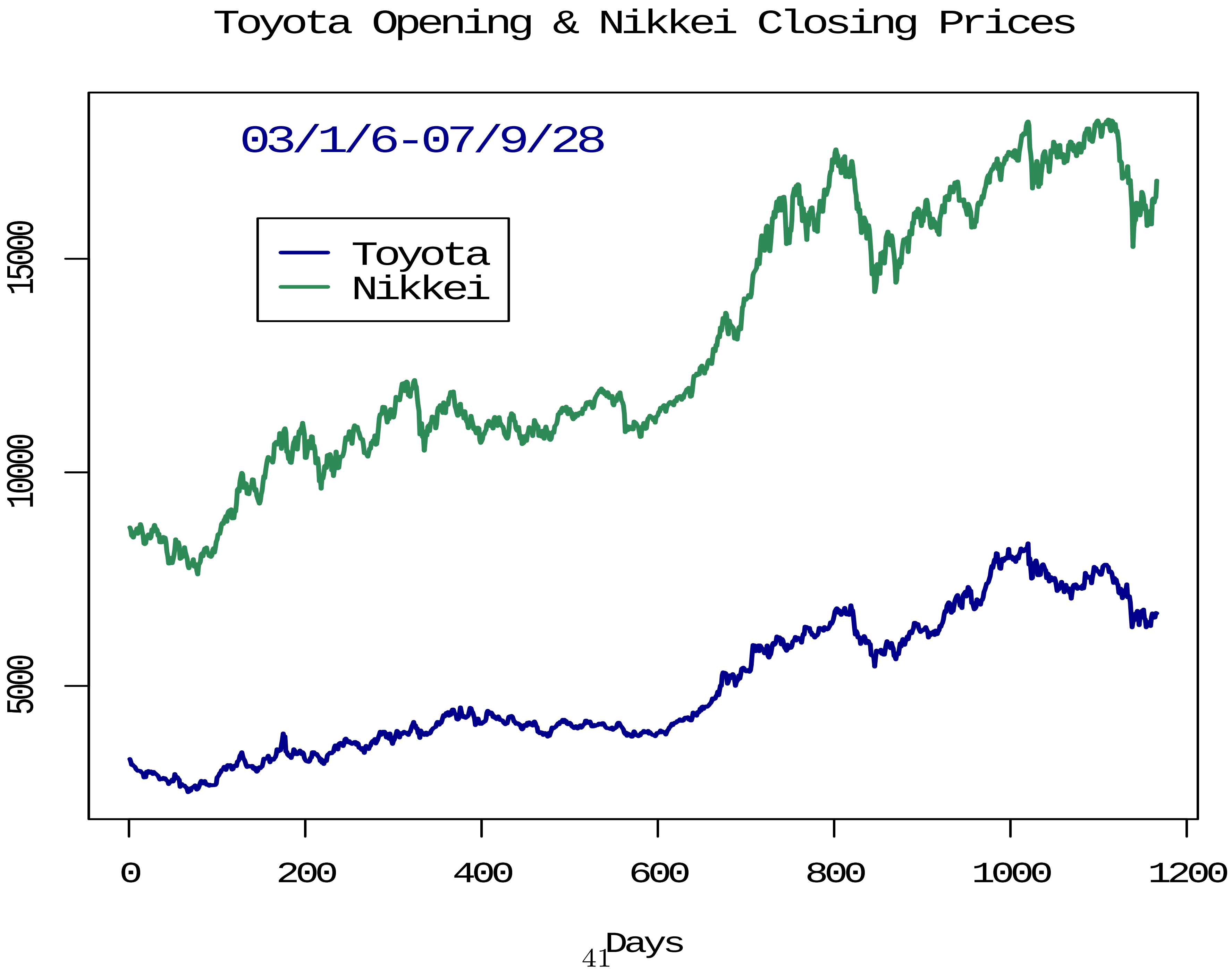
empirical conditional prob. by Market

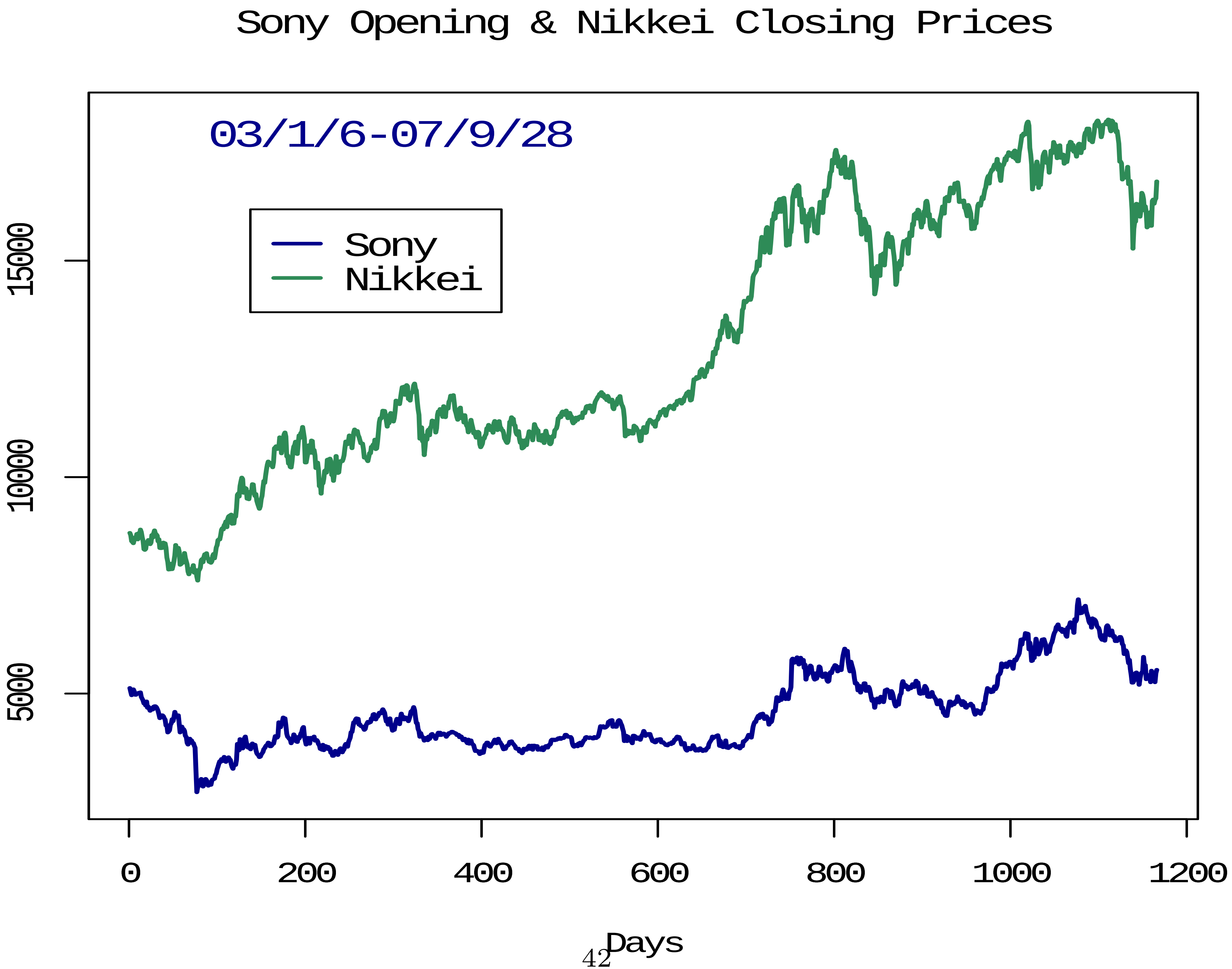
$\mathrm{MDIV} = D\left(\hat{p}_{n,y|x} \| q | \hat{p}_{n,x}\right)$:

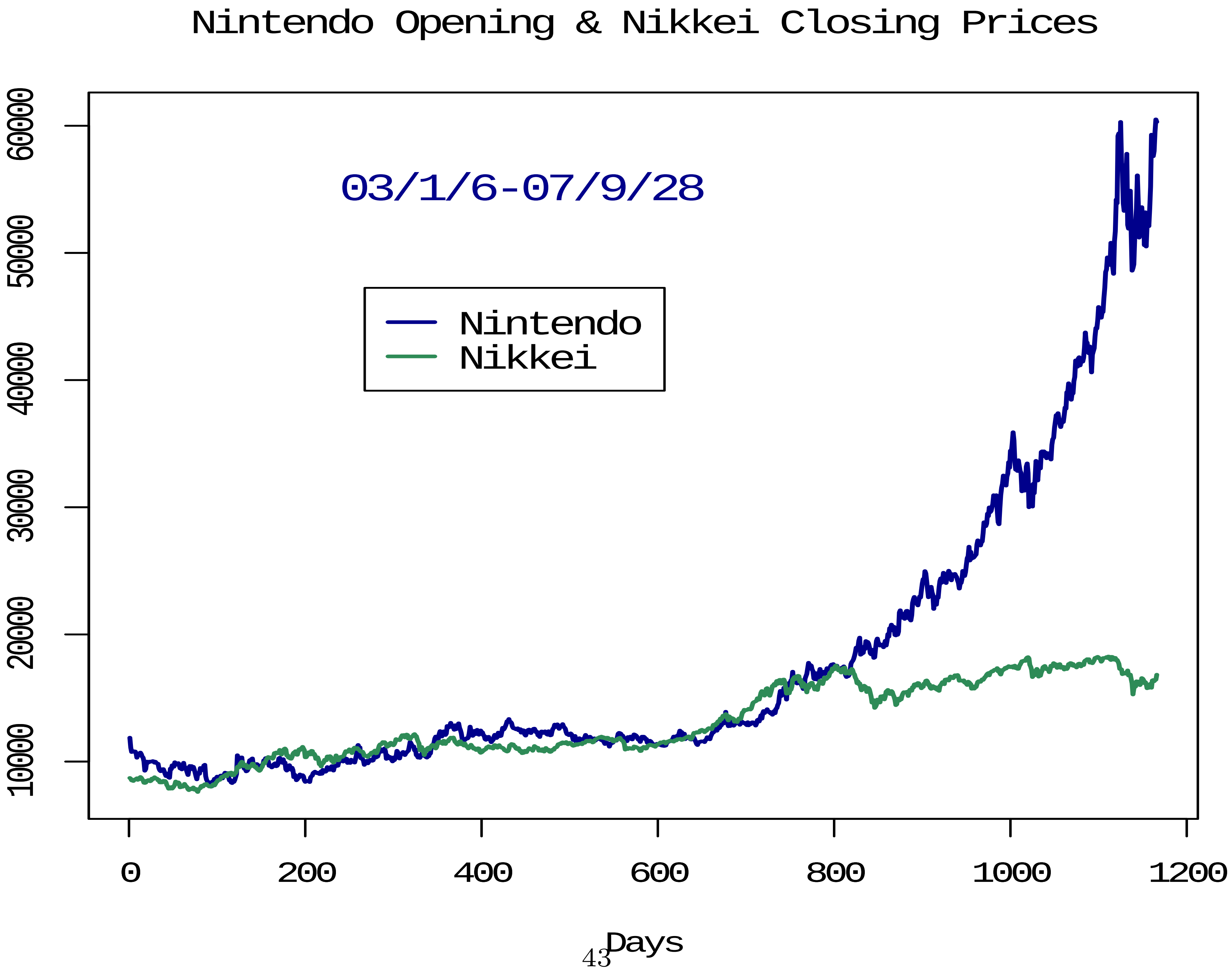
empirical conditional K-L divergence

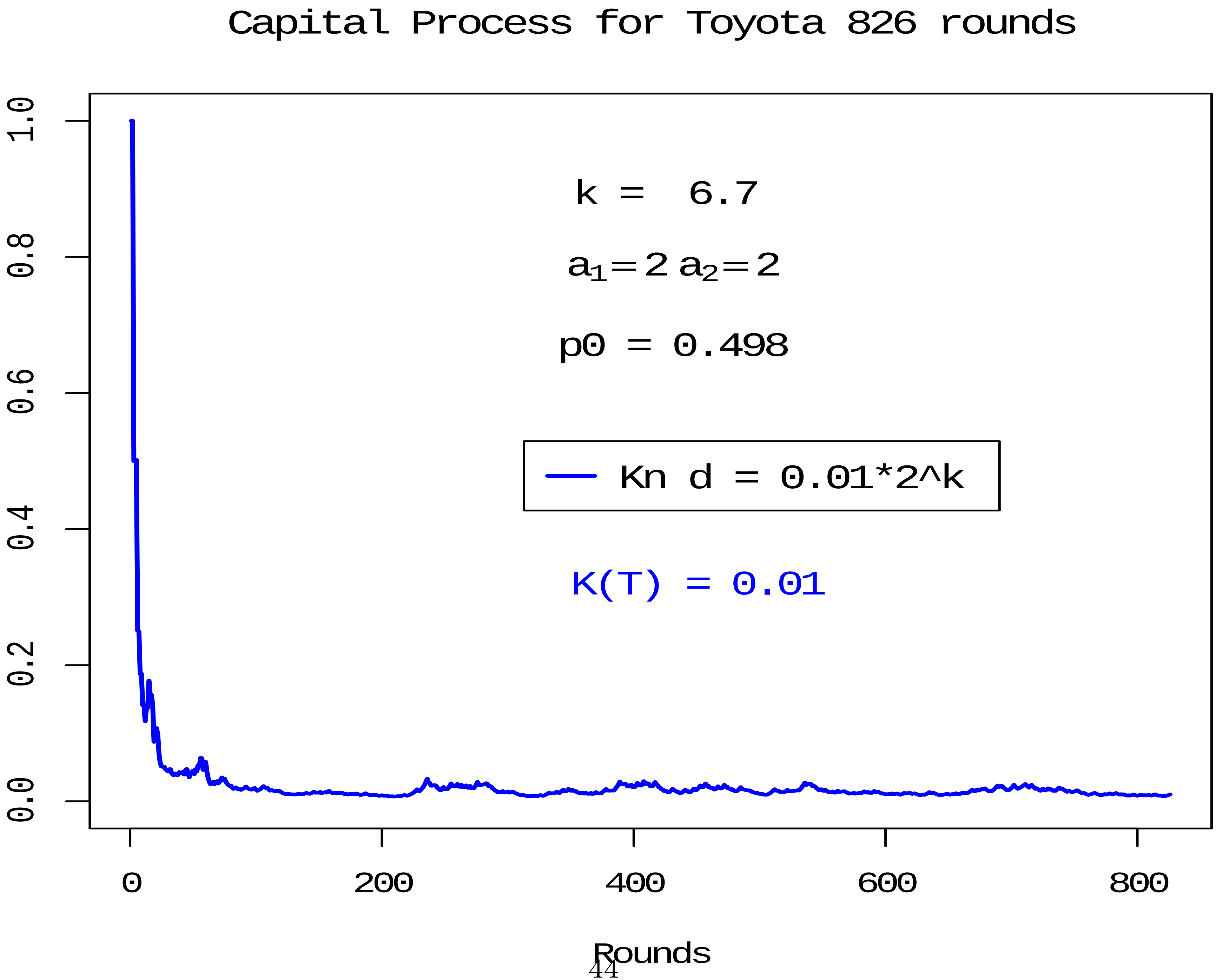
$$\mathrm{mLKn} = \frac{\log \mathcal{K}_n}{n} \quad :$$

empirical exponential growth rate of $\mathcal{K}_n$

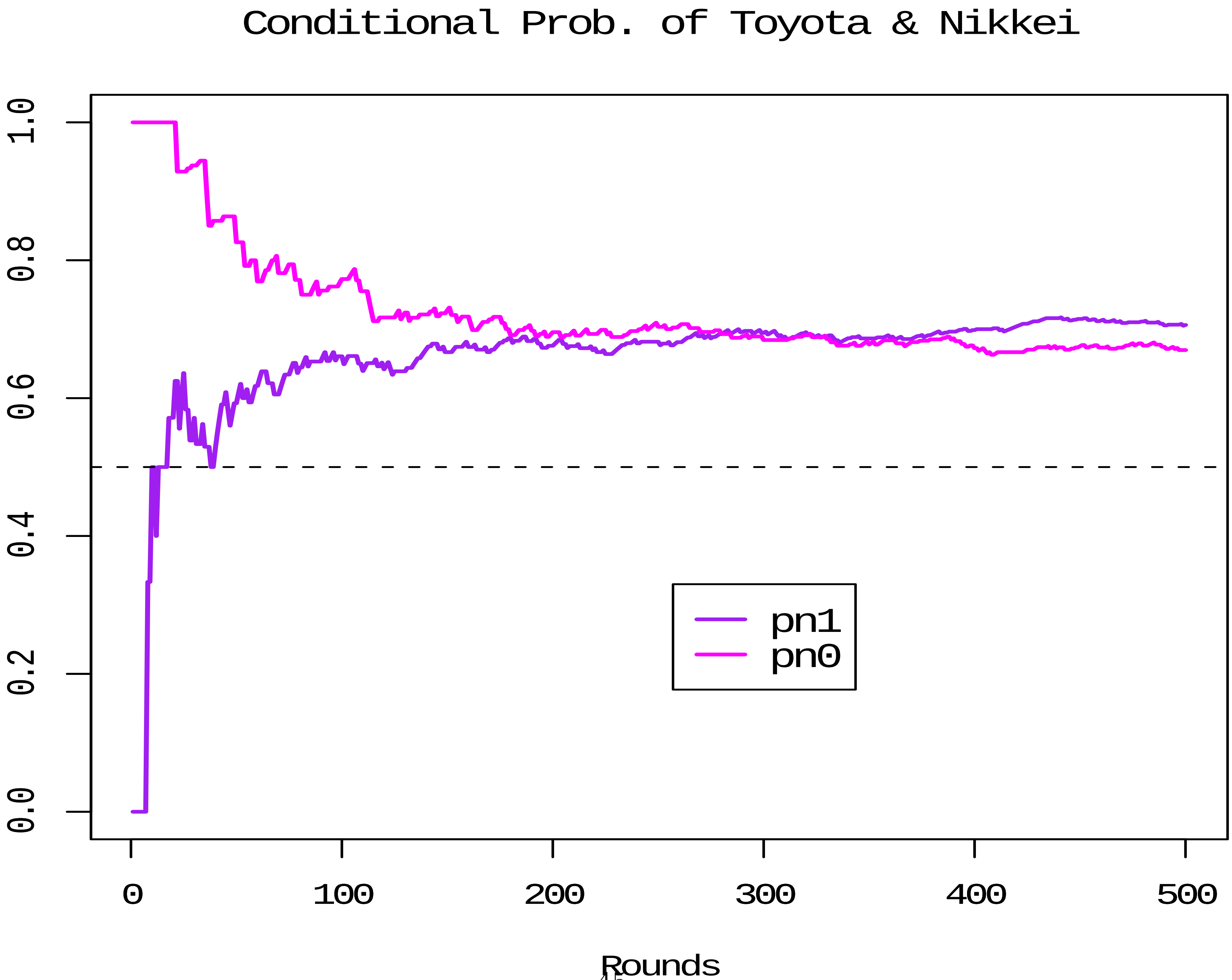

Toyota Opening & Nikkei Closing Prices
03/1/6-07/9/28
Toyota
Nikkei
5000
10000
15000
0
200
400
600
800
1000
1200
Days


Sony Opening & Nikkei Closing Prices
03/1/6-07/9/28
Sony
Nikkei
5000
10000
15000
0
200
400
600
800
1000
1200
Days

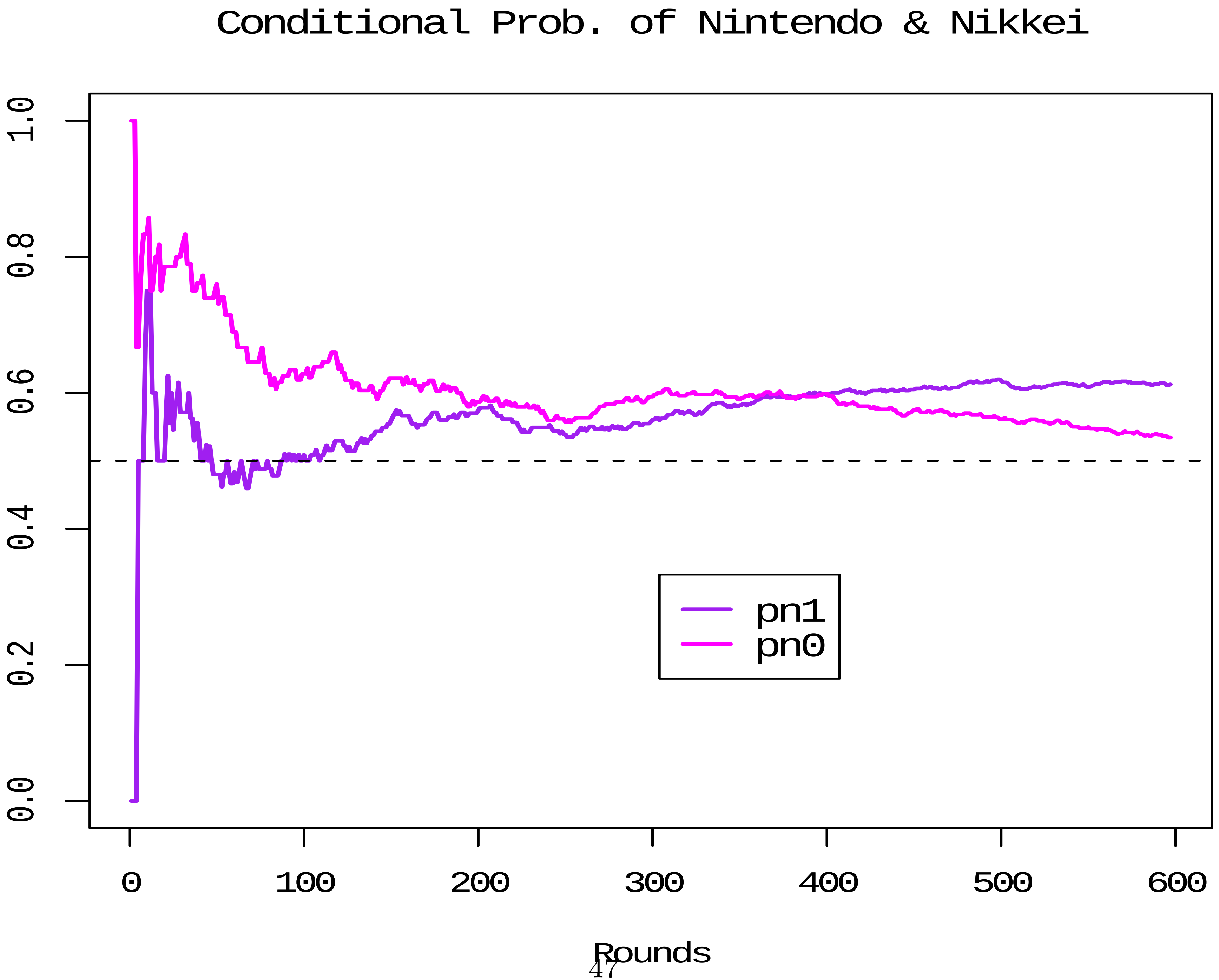

Nintendo Opening & Nikkei Closing Prices
03/1/6-07/9/28
Nintendo
Nikkei
10000
20000
30000
40000
50000
60000
0
200
400
600
800
1000
1200
Days

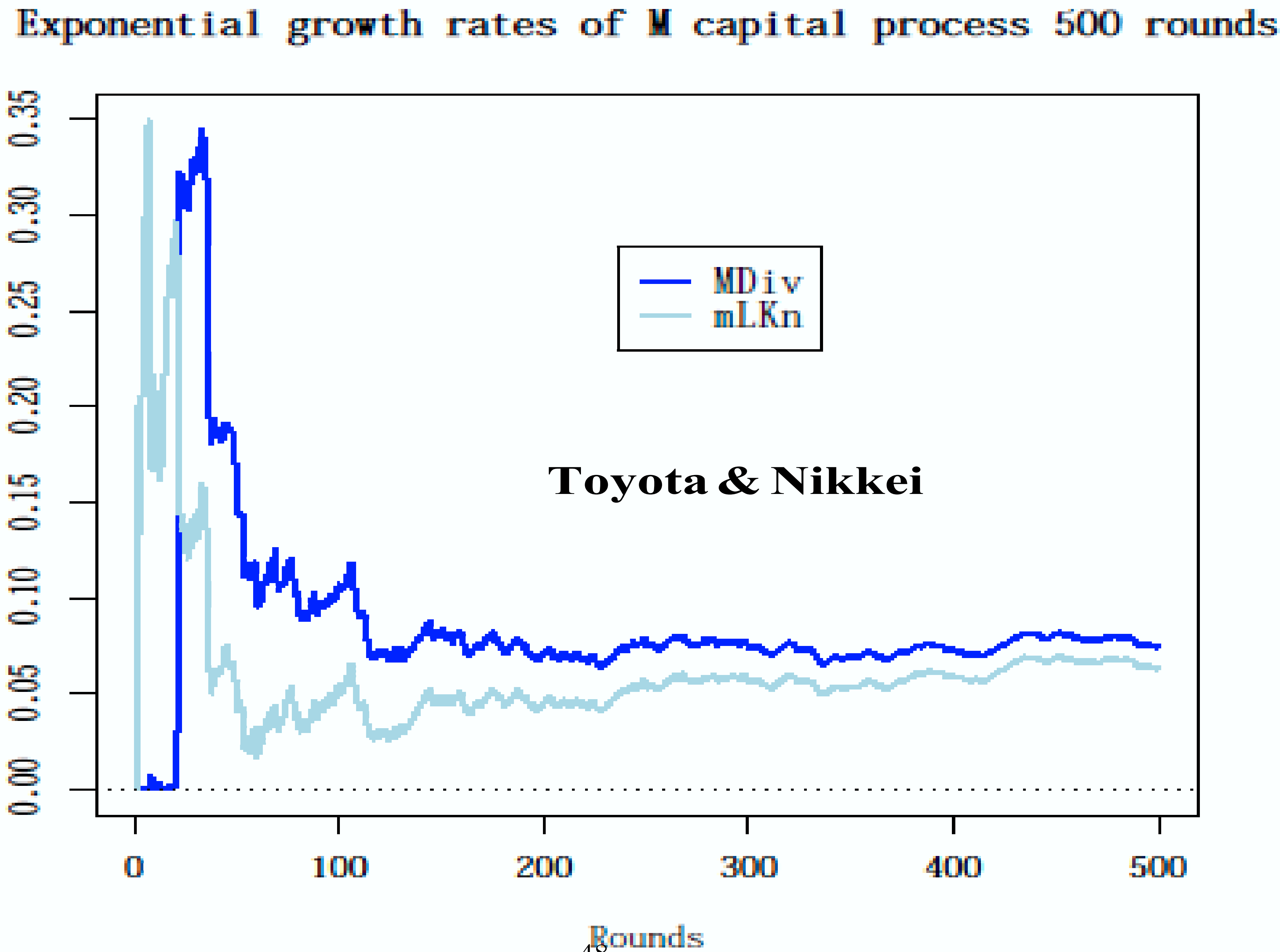


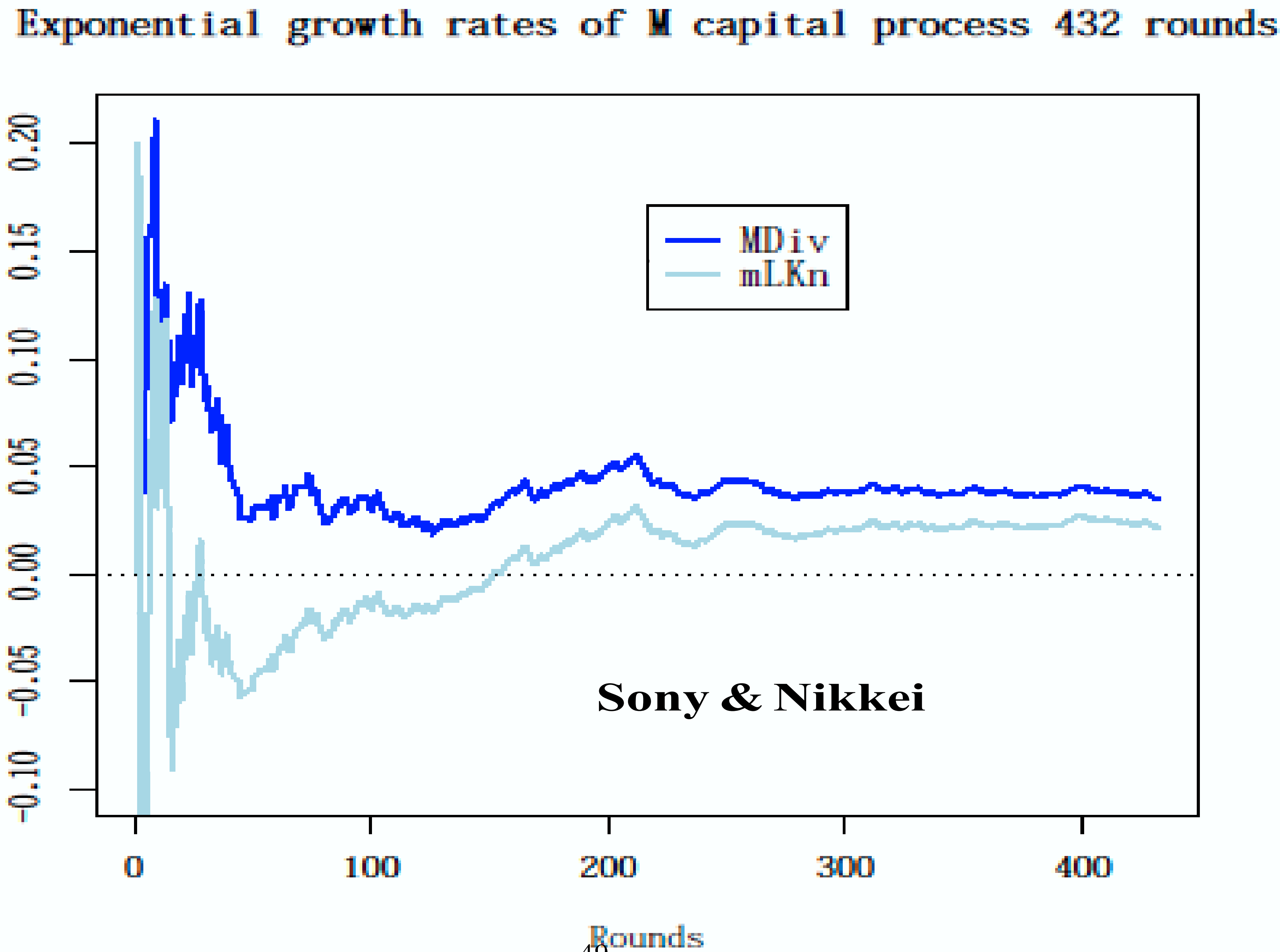
Capital Process for Toyota 826 rounds
k = 6.7
a1=2 a2=2
p0 = 0.498
Kn d = 0.01*2^k
K(T) = 0.01
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
Rounds

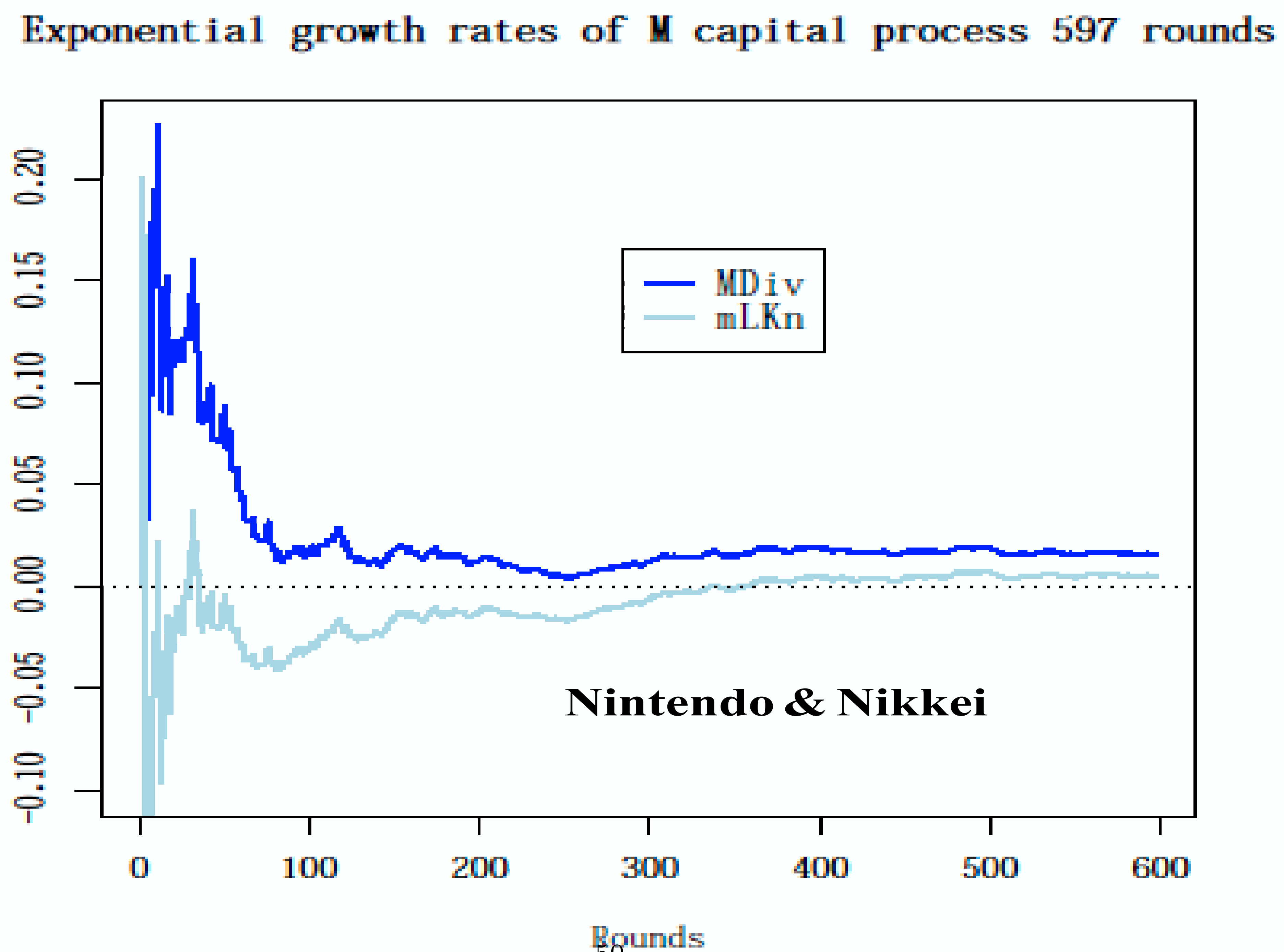

Conditional Prob. of Toyota & Nikkei
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
500
Rounds
pn1
pn0

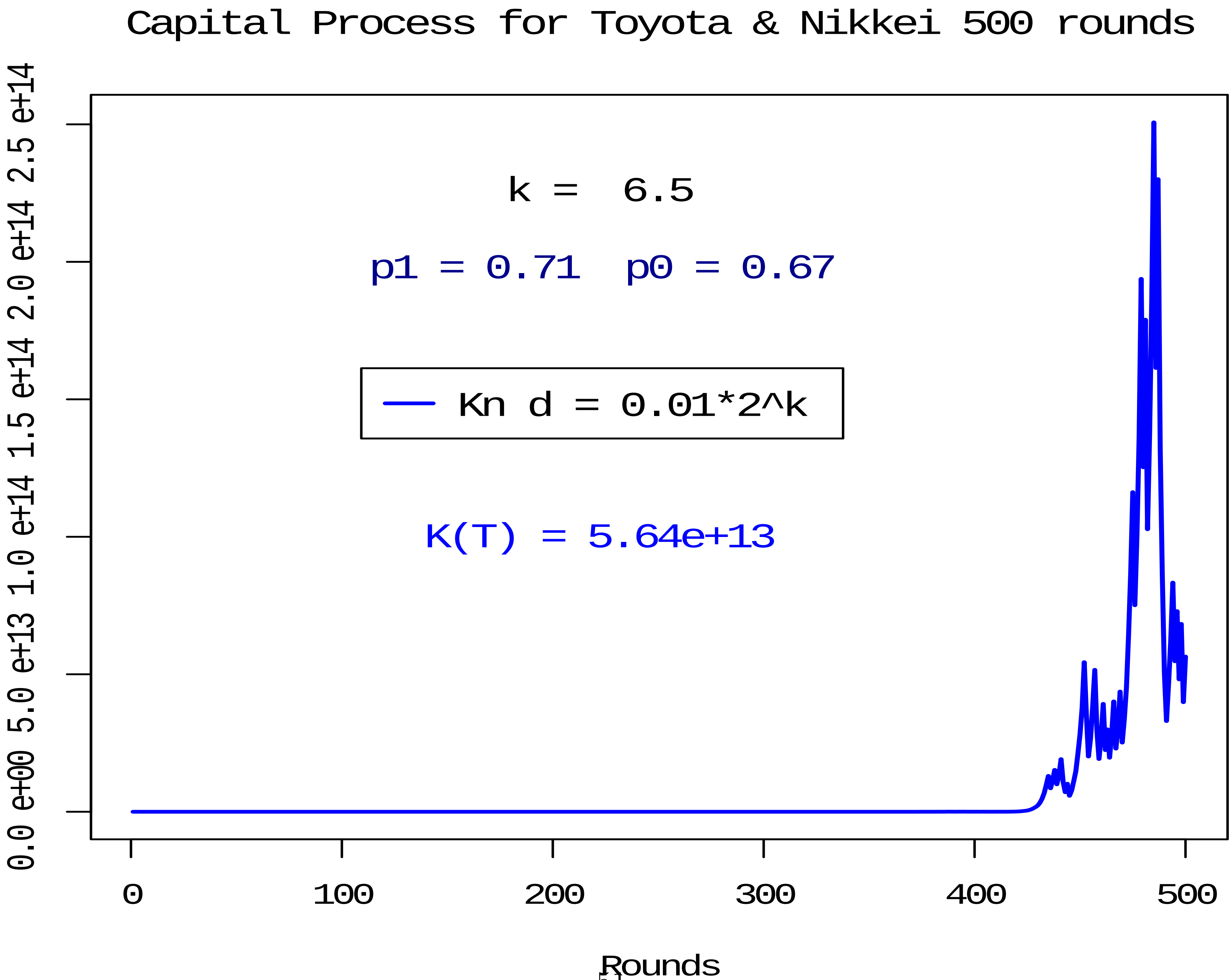
Conditional Prob. of Sony & Nikkei
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
Rounds
pn1
pn0


Conditional Prob. of Nintendo & Nikkei
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
500
600
Rounds
pn1
pn0

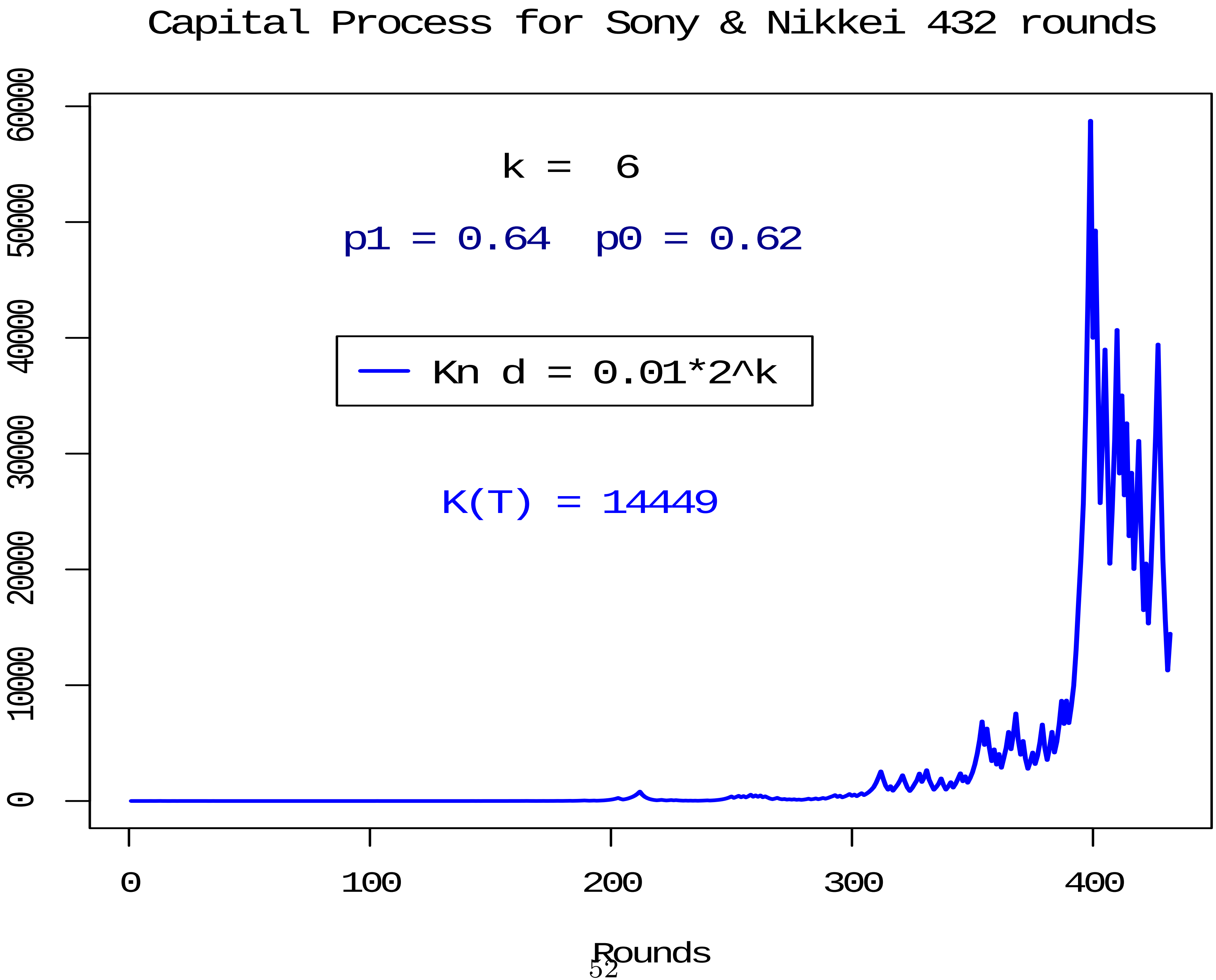

Exponential growth rates of M capital process 500 rounds
MDiv
mLKn
Toyota & Nikkei
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
100
200
300
400
500
Rounds

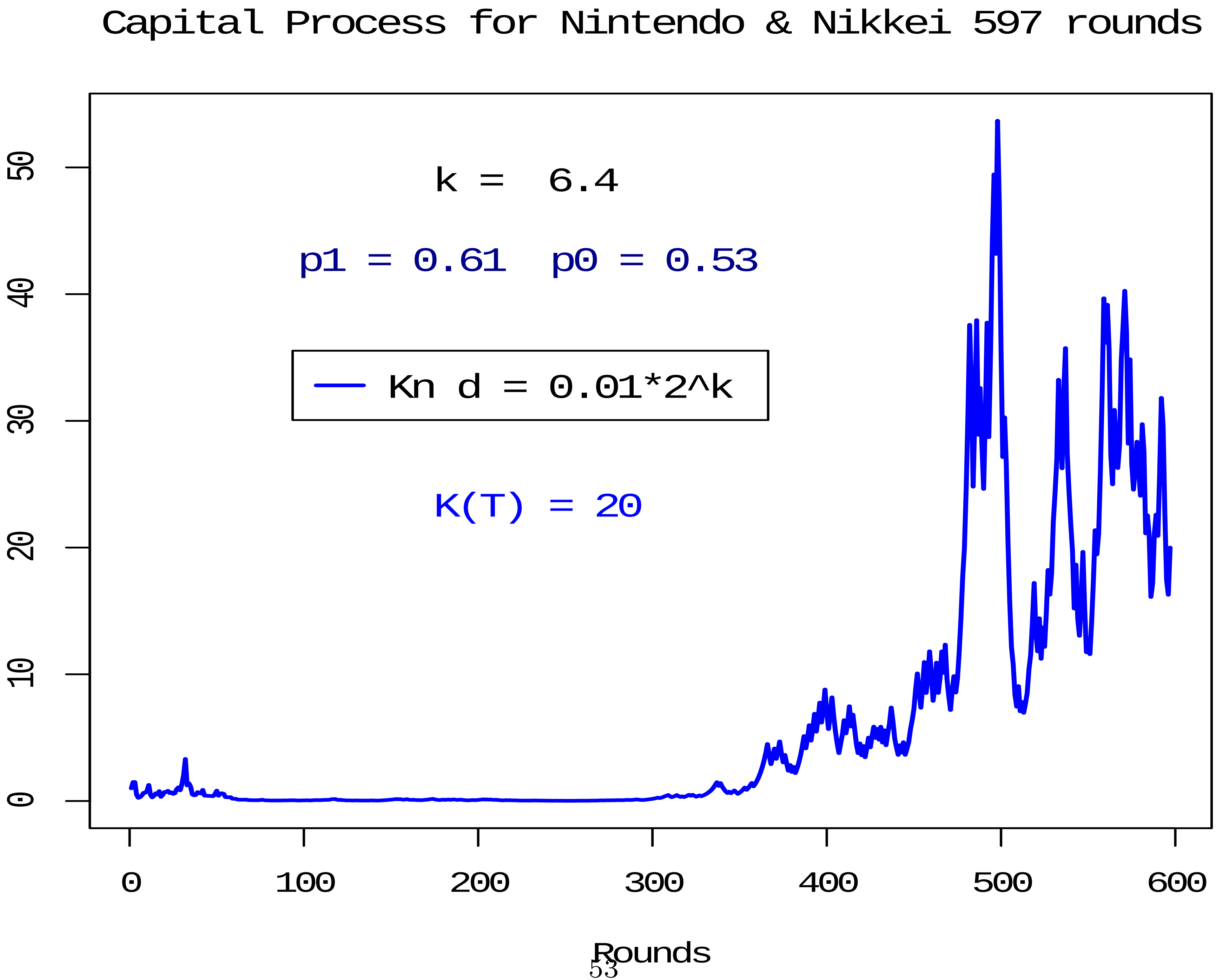

Exponential growth rates of M capital process 432 rounds
MDiv
mLKn
Sony & Nikkei
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
0
100
200
300
400
Rounds


Exponential growth rates of M capital process 597 rounds
MDiv
mLKn
Nintendo & Nikkei
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
0
100
200
300
400
500
600
Rounds



Capital Process for Toyota & Nikkei 500 rounds
k = 6.5
p1 = 0.71 p0 = 0.67
Kn d = 0.01*2^k
K(T) = 5.64e+13
0.0 e+00
5.0 e+13
1.0 e+14
1.5 e+14
2.0 e+14
2.5 e+14
0
100
200
300
400
500
Rounds


Capital Process for Sony & Nikkei 432 rounds
k = 6
p1 = 0.64 p0 = 0.62
Kn d = 0.01*2^k
K(T) = 14449
0
10000
20000
30000
40000
50000
60000
0
100
200
300
400
Rounds


Capital Process for Nintendo & Nikkei 597 rounds
k = 6.4
p1 = 0.61 p0 = 0.53
Kn d = 0.01*2^k
K(T) = 20
0
10
20
30
40
50
0
100
200
300
400
500
600
Rounds

3. Source coding and betting strategy

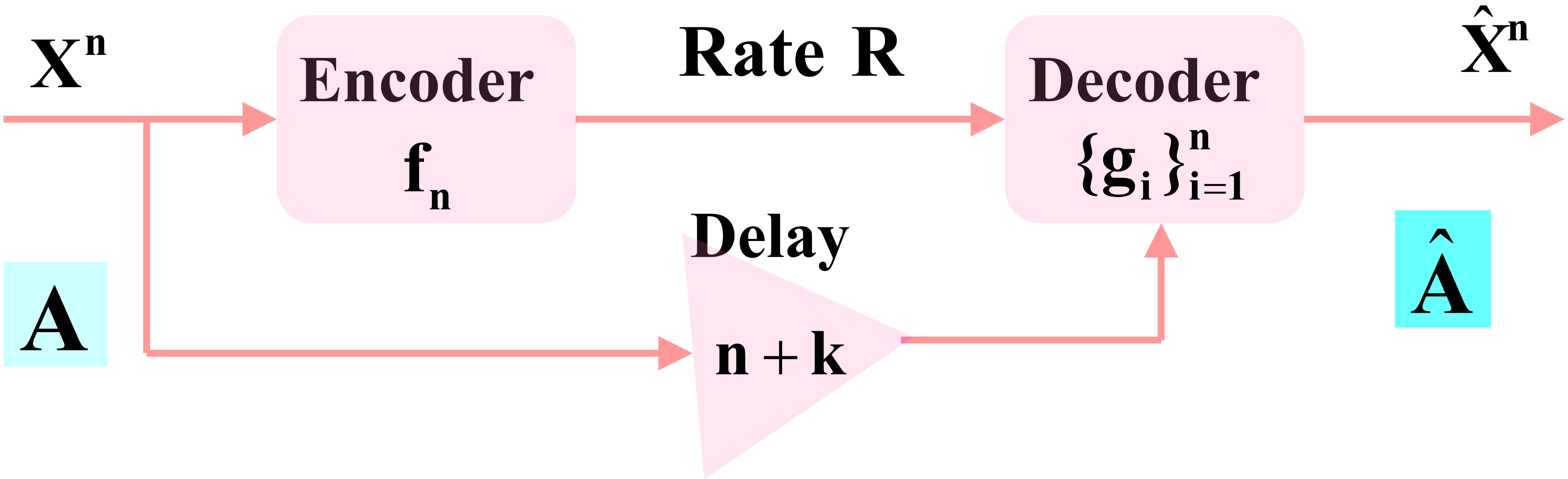
3.1 Lossy source coding with feedforward

■ Source coding model

$X^n = (X_1, \ldots, X_n) \in \mathcal{X}^n$: source sequence

$\hat{X}^n = (\hat{X}_1, \ldots, \hat{X}_n) \in \hat{\mathcal{X}}^n$: estimated sequence

$d_n : \mathcal{X}^n \times \hat{\mathcal{X}}^n \to \mathbb{R}^+$: distortion measure



$f_n : X^n \rightarrow \{1, 2, \ldots, 2^{nR}\}$: encoding function

$g_i : \{1, 2, \ldots, 2^{nR}\} \times X^{i-k} \rightarrow \hat{X}$ $i = 1, \ldots, n$:
sequence of decoding functions

$\hat{X}^n = (\hat{X}_1, \ldots, \hat{X}_n)$: reproduction sequence

Source coding system with feedforward

- $(2^{nR}, n)$ **code with k-delayed feedforward**

$f_n : \mathcal{X}^n \longrightarrow \{1, 2, \ldots, 2^{nR}\}$ **: encoding function**

$g_i : \{1, 2, \ldots, 2^{nR}\} \times \mathcal{X}^{i-k} \longrightarrow \hat{\mathcal{X}} \quad i = 1, \ldots, n$ **:**

sequence of decoding functions

$$f_n(X^n) = w \in \{1, 2, \ldots, 2^{nR}\}$$

$$g_i(w, X^{i-k}) = \hat{X}_i \quad i = 1, \ldots, n$$

$D_n = E_{X^n}\left[d_n(X^n, \hat{X}^n)\right]$ **:**

distortion associated with the $(2^{nR}, n)$ code

$(\boldsymbol{R}, \boldsymbol{D})$: achievable

$$\Leftrightarrow \ \exists(2^{nR}, n) \text{ code with } \limsup_{n\to\infty} D_n \le D$$

- Rate distortion theorem

$$R \ge R^k_{ff}(D) \ \Rightarrow \ (R, D) \text{ is achievable}$$

$$R^k_{ff}(D) = \liminf_{P_{\hat{X}^n|X^n}, D_n \le D} \frac{1}{n} I_k(\hat{X}^n \to X^n) \ :$$

rate distortion function with k-delayed feedforward

$$I_k(\hat{X}^n \to X^n) = \sum_{i=1}^{n} I(\hat{X}^{i+k-1}; X_i | X^{i-1})$$

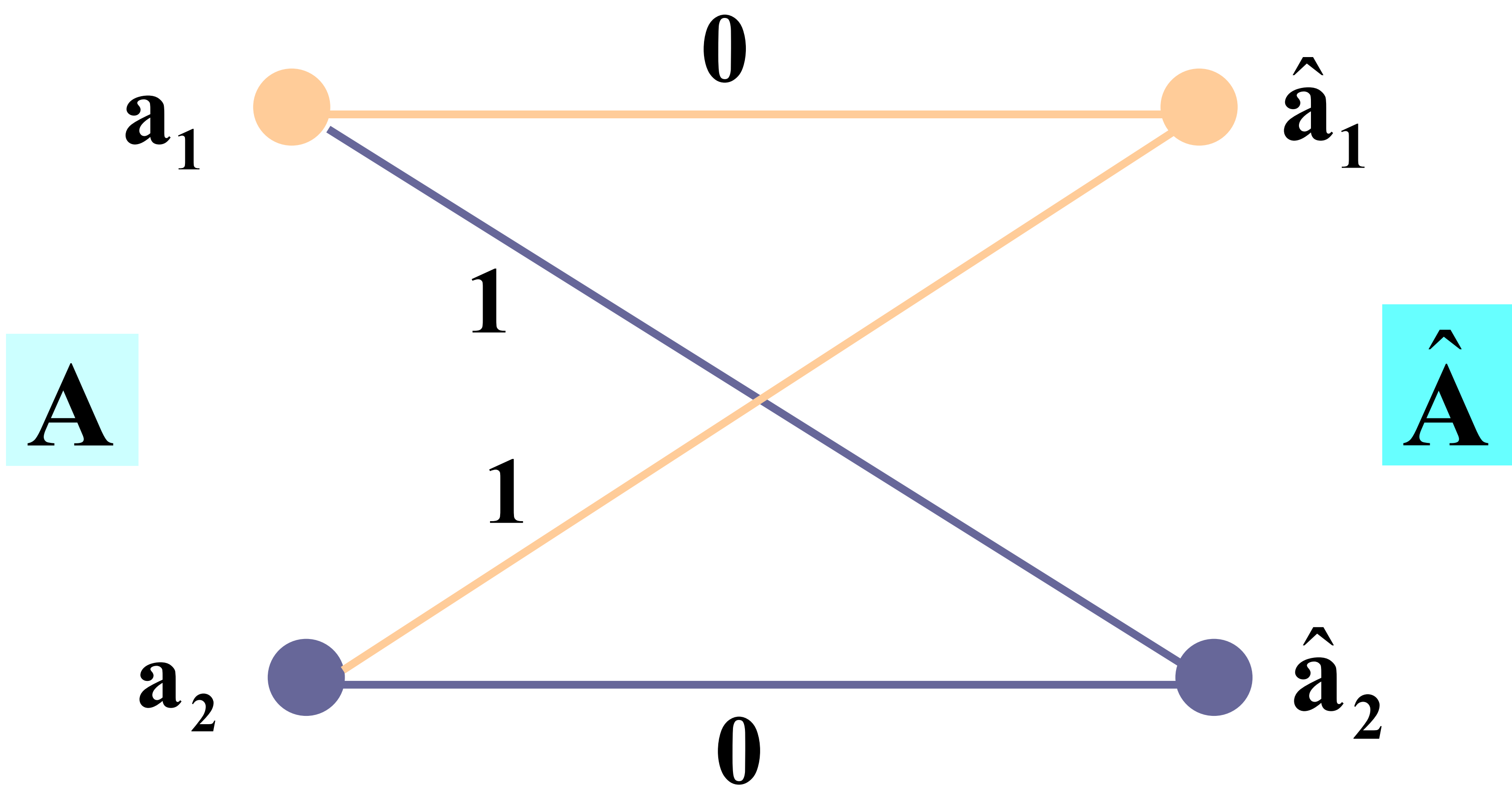
$$= I(\hat{X}^n; X^n) - \sum_{i=k+1}^{n} I(X^{i-k}; \hat{X}_i | \hat{X}^{i-1}) \quad :$$

directed information from $\hat{X}^n$ to X^n with k-delayed feedforward

$$\sum_{i=k+1}^{n} I(X^{i-k}; \hat{X}_i | \hat{X}^{i-1}) \quad :$$

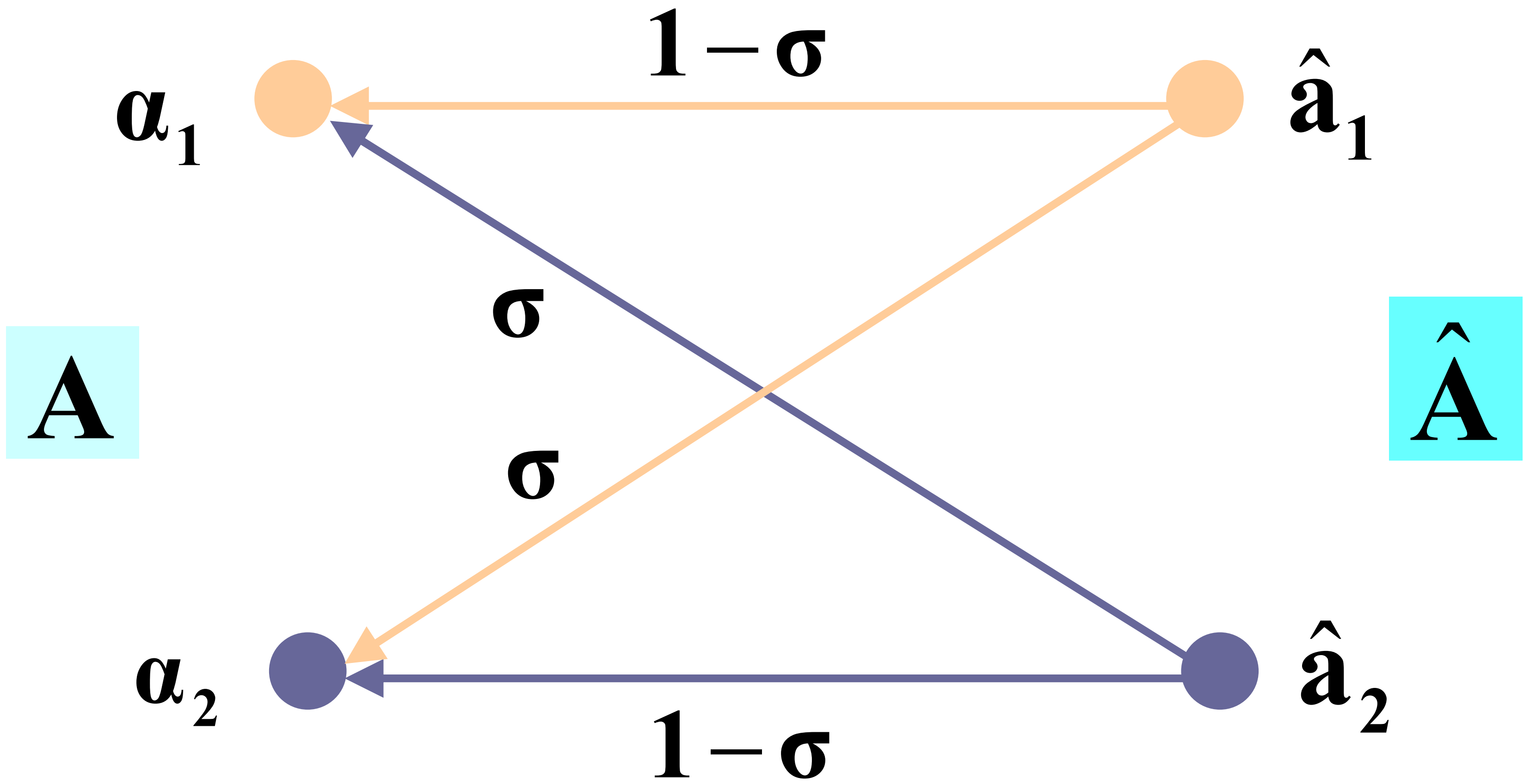
information quantity obtained for free

Reality's move & distortion



Binary symmetric transmission

Forecaster's move



Binary symmetric transmission

Binary symmetric transmission

$$R_g(D) = I(A;\hat{A})$$

$$\Leftrightarrow \quad \rho = P_{A|\hat{A}}(x_2|\hat{x}_1) = P_{A|\hat{A}}(x_1|\hat{x}_2) = D$$

$$\sigma = Q_{A|\hat{A}}(x_2|\hat{x}_1) = Q_{A|\hat{A}}(x_1|\hat{x}_2)$$

$$= \frac{\alpha_1 - a_1 + D(\alpha_2 - \alpha_1)}{a_2 - a_1}$$

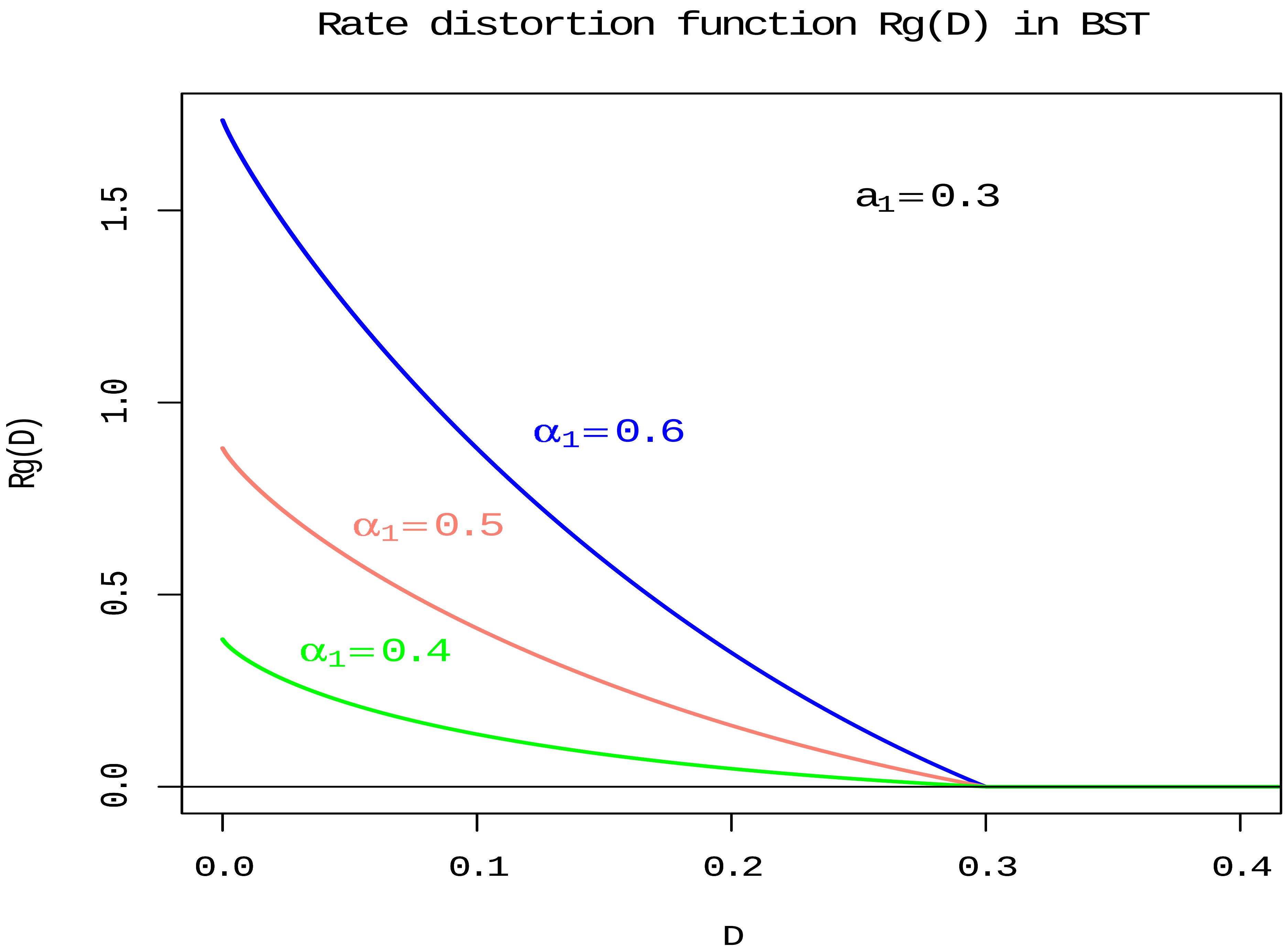
$$R_g(D) = D(P_{\hat{A}|A} \| Q_{\hat{A}|A} | P_A)$$

$$= D(\rho \| \sigma) - D(a_1 \| \alpha_1)$$

$$Q_C = Q_A \times Q_B \;\Leftrightarrow\; \sigma = \alpha_1 = 0.5 \quad Q_{\hat{A}|A} = P_{\hat{A}}$$

$$\begin{aligned} R_g(D) &= D(P_{\hat{A}|A} \| P_{\hat{A}} | P_A) \\ &= D(\rho \| 0.5) - D(a_1 \| 0.5) \\ &= H(a_1) - H(D) = R(D) \end{aligned}$$

$$\begin{aligned} R_g(0) &= D(\delta_{\hat{A}|A} \| P_{\hat{A}} | P_A) \\ &= D(0 \| 0.5) - D(a_1 \| 0.5) \\ &= H(a_1) = R(0) \end{aligned}$$



Rate distortion function Rg(D) in BST
a1=0.3
α1=0.6
α1=0.5
α1=0.4
Rg(D)
0.0
0.5
1.0
1.5
0.0
0.1
0.2
0.3
0.4
D

Conditional betting game $\hat{A}|A$ given A

- **Notations**

$\chi^n_{x1}, \ \chi^n_{x0}$: number of $x_i = 1, \ 0 \ \ (i = 1. \ldots, n)$

$\chi^n_{\hat{x}1}, \ \chi^n_{\hat{x}0}$: number of $\hat{x}_i = 1, \ 0 \ \ (i = 1. \ldots, n)$

$\chi^n_{11}, \ \chi^n_{10}, \ \chi^n_{01}, \ \chi^n_{00}$:

number of $(x_{i-k}, \hat{x}_i) = (1, 1), \ (1, 0), \ (0, 1), \ (0, 0)$

$(i = k + 1, \ldots, n)$

$$\boldsymbol{P^{\pm}_{\hat{A}}(\hat{x}^n|x^n) = P^{+}_{\hat{A}}(\hat{x}^n|x^n) \times P^{-}_{\hat{A}}(\hat{x}^n|x^n)}$$

$$\boldsymbol{P^{+}_{\hat{A}}(\hat{x}^n|x^n) = \frac{B(\chi^n_{11} + c_1, \chi^n_{10} + c_0)}{B(c_1, c_0)}}$$

$$\boldsymbol{P^{-}_{\hat{A}}(\hat{x}^n|x^n) = \frac{B(\chi^n_{01} + c_1, \chi^n_{00} + c_0)}{B(c_1, c_0)}} \quad :$$

conditional beta binomial distribution

modeled by Skeptic

$$Q_{\hat{A}}(\hat{x}^n|x^n) = P_{\hat{A}}(\hat{x}^n) = \frac{B(\chi^n_{\hat{x}1} + c_1, \chi^n_{\hat{x}0} + c_0)}{B(c_1, c_0)} \quad :$$

beta binomial distribution modeled by

Forecaster

Maximize $E_{P_{\hat{A}\pm}}[\log \mathcal{K}_n] \;\Rightarrow\; \{\alpha_i^{\hat{A}\pm}\}_{i=1}^n$

$$\alpha_i^{\hat{A}\pm} = 0 \quad 1 \le i \le k$$

$$\alpha_i^{\hat{A}\pm} = \begin{cases} \alpha_i^{\hat{A}+} & \text{if } \; x_{i-k} = 1 \\ \alpha_i^{\hat{A}-} & \text{if } \; x_{i-k} = 0 \end{cases} \qquad k+1 \le i \le n$$

$$\alpha_i^{\hat{A}+} = \frac{p_i^{P_{\hat{A}}^{+}} - q_i^{Q_{\hat{A}}}}{q_i^{Q_{\hat{A}}}(1 - q_i^{Q_{\hat{A}}})} \qquad \alpha_i^{\hat{A}-} = \frac{p_i^{P_{\hat{A}}^{-}} - q_i^{Q_{\hat{A}}}}{q_i^{Q_{\hat{A}}}(1 - q_i^{Q_{\hat{A}}})}$$

$$p_i^{P_{\hat{A}}^{+}} = P_{\hat{A}}^{+}(\hat{x}_i = 1 | x^{i-k}) = \frac{\chi_{11}^{i-1} + c_1}{\chi_{11}^{i-1} + \chi_{10}^{i-1} + c_1 + c_0}$$

$$p_i^{P_{\hat{A}}^{-}} = P_{\hat{A}}^{-}(\hat{x}_i = 1 | x^{i-k}) = \frac{\chi_{01}^{i-1} + c_1}{\chi_{01}^{i-1} + \chi_{00}^{i-1} + c_1 + c_0}$$

$$q_i^{Q_{\hat{A}}} = Q_{\hat{A}}(\hat{x}_i = 1 | x^{i-k}) = \frac{\chi_{\hat{x}1}^{i-1} + c_1}{i - 1 + c_1 + c_0}$$

The optimal capital process of Skeptic is expressed as the likelihood ratio

$$\mathcal{K}^{P^{\pm}_{\hat{A}}}(\hat{x}^n|x^n) = \prod_{i=1}^{n}\Big(1 + \alpha_i^{\hat{A}\pm}(\hat{x}_i - q_i^{Q_{\hat{A}}})\Big)$$

$$= \prod_{i=1}^{n}\frac{P^{\pm}_{\hat{A}}(\hat{x}_i|x^{i-k})}{Q_{\hat{A}}(\hat{x}_i|x^{i-k})} = \frac{P^{+}_{\hat{A}}(\hat{x}^n|x^n) \times P^{-}_{\hat{A}}(\hat{x}^n|x^n)}{Q_{\hat{A}}(\hat{x}^n|x^n)}$$

$$\log \mathcal{K}^{P^{\pm}_{\hat{A}}}(\hat{x}^n|x^n) = nD\left(\hat{p}_{n,\hat{x}|x} \| \hat{q}_{n,\hat{x}} | \hat{p}_{n,x}\right) - \frac{1}{2}\left(\log \chi^n_{x1} + \log \chi^n_{x0}\right) + O(1)$$

$$\hat{p}_{n,\hat{x}|1} = \left(\frac{\chi^n_{11}}{\chi^n_{x1}}, \frac{\chi^n_{10}}{\chi^n_{x1}}\right) \quad \hat{p}_{n,\hat{x}|0} = \left(\frac{\chi^n_{01}}{\chi^n_{x0}}, \frac{\chi^n_{00}}{\chi^n_{x0}}\right) \ :$$

empirical conditional prob. of Reality

$$\hat{q}_{n,\hat{x}} = \left(\frac{\chi^n_{\hat{x}1}}{n}, \frac{\chi^n_{\hat{x}0}}{n}\right) \ :$$

empirical risk neutral prob. of Forecaster

$$D\left(\hat{p}_{n,\hat{x}|x} \| \hat{q}_{n,\hat{x}} | \hat{p}_{n,x}\right) \ :$$

empirical conditional K-L divergence

3.2 Efficient source coding scheme

■ **Betting strategy and data compression** **(a variant of the arithmetic coding)**

- **Encoding**

$$x^n = x_1 \ldots x_n \in \{0,1\}^n \Rightarrow \hat{x}^n = \hat{x}_1 \ldots \hat{x}_n \in \{0,1\}^n$$

such that $P_{\hat{X}^n|X^n}$ achieves $R^k_{ff}(D)$

$\hat{x}|x(n) = (\hat{x}_1|x_1, \ldots, \hat{x}_n|x_n)$ **:**

observed sequence by the encoder

2^n sequences $\{\hat{x}^n\}$ **: in lexicographical order**

The encoder calculates the cumulative sum

$$G^{\pm}_{\hat{A}}(\hat{x}|x(n)) = \sum_{\hat{x}'|x(n) \leq \hat{x}|x(n) : \text{typical}} R^{\pm}_{\hat{A}}(\hat{x}'|x(n))$$

$$R^{\pm}_{\hat{A}}(\hat{x}'|x(n)) = \frac{1}{\mathcal{K}^{P^{\pm}_{\hat{A}}}(\hat{x}'|x(n))} = \frac{Q_{\hat{A}}(\hat{x}'|x(n))}{P^{\pm}_{\hat{A}}(\hat{x}'|x(n))}$$

$\hat{x}'|x(n)$: **typical**

$$\Leftrightarrow \left| \frac{1}{n} \log \mathcal{K}^{P^{\pm}_{\hat{A}}}(\hat{x}'|x(n)) - R^{k}_{ff}(D) \right| < \epsilon$$

$$G^{\pm}_{\hat{A}}(\hat{x}|x(n)) \in [0, 1] \quad \text{as} \quad n \to \infty$$

$$\ell = \left\lceil \log \mathcal{K}^{P^{\pm}_{\hat{A}}}(\hat{x}|x(n)) \right\rceil + 1 \quad m = \lceil \log n \rceil :$$

specifed numbers of bits

$$\left\lfloor G^{\pm}_{\hat{A}}(\hat{x}|x(n)) \right\rfloor = .c_1 c_2 \ldots c_\ell \;\; c_i \in \{0,1\} \;\; :$$

binary decimal to ℓ place accuracy

$$\chi^n_{x1} = d_1 d_2 \ldots d_m \;\; d_i \in \{0,1\} \;\; :$$

binary number to m digits

$$c(\ell) = (c_1, c_2, \ldots, c_\ell) \;\; d(m) = (d_1, d_2, \ldots, d_m) \;\; :$$

code sequences sent to the decoder

- **Decoding**

When there exists a feedforward $X \longrightarrow \hat{X}$ and χ_{x1}^{n} is known, the decoder can also sequentially calculate the cumulative sum

$$G_{\hat{A}}^{\pm}(\hat{x}|x(n)) = \sum_{\hat{x}'|x(n) \leq \hat{x}|x(n):\text{typical}} R_{\hat{A}}^{\pm}(\hat{x}'|x(n))$$

until $G_{\hat{A}}^{\pm}(\hat{x}|x(n)) \geq .c(\ell)$

$\Rightarrow$ $\hat{x}|x(n)$ **: the encoded sequence**

From the expression

$$\log \mathcal{K}^{P^{\pm}_{\hat{A}}}(\hat{x}^n|x^n) = nD\left(\hat{p}_{n,\hat{x}|x}\|\hat{q}_{n,\hat{x}}|\hat{p}_{n,x}\right) - \frac{1}{2}\left(\log \chi^n_{x1} + \log \chi^n_{x0}\right) + O(1)$$

the required number of bits is

$$\ell + m = \left\lceil \log \mathcal{K}^{P^{\pm}_{\hat{A}}}(\hat{x}|x(n)) \right\rceil + 1 + \lceil \log n \rceil = \left\lceil nD\left(\hat{p}_{n,\hat{x}|x}\|\hat{q}_{n,\hat{x}}|\hat{p}_{n,x}\right) \right\rceil + O(\log n)$$

The empirical codeword length $L_n^* = \frac{\ell+m}{n}$ per source symbol is

$$L_n^* = \frac{\ell + m}{n}$$

$$\leq D\left(\hat{p}_{n,\hat{x}|x} || \hat{q}_{n,\hat{x}} | \hat{p}_{n,x}\right) + O\left(\frac{\log n}{n}\right)$$

$$\rightarrow R_{ff}^{k}(D) \quad \text{as} \quad n \rightarrow \infty$$

References

[1] Leo Breiman. Optimal gambling systems for favorable games. *Fourth Berkeley Symposium on Probability and Statistics I*, 65-78, 1961.

[2] Thomas M. Cover. Universal portfolios. *Mathematical Finance*, 1 (1), 1-29, 1991.

[3] **John Larry Kelly Jr. A new interpretation of information rate. *Bell System Technical Journal*, Vol.35, 917-926, 1956.**

[4] **L. C. MacLean, E. O. Thorp and W. T. Ziemba eds. *The Kelly Capital Growth Investment Criterion : Theory and Practice. Handbook in Finantial Economic Series*, Vol.3, World Scientific, London, 2010.**

[5] M. Kumon, A. Takemura and K. Takeuchi. Capital process and optimality properties of a Bayesian Skeptic in coin-tossing games. *Stochastic Analysis and Applications*, 26, 1161-1180, 2008.

[6] K. Takeuchi, M. Kumon and A. Takemura. A new formulation of asset trading games in continuous time with essential forcing of variation exponent.

Bernoulli, 15, 1243-1258, 2009.

[7] K. Takeuchi, M. Kumon and A. Takemura. Multistep Bayesian strategy in coin-tossing games and its application to asset trading games in continuous time. *Stochastic Analysis and Applications*, 28, 842-861, 2010.

[8] K. Takeuchi, A. Takemura and M. Kumon. New procedures for testing whether stock price processes are

martingales. *Computational Economics*, 37, 67-88, 2011.

[9] M. Kumon, A. Takemura and K. Takeuchi. Sequential optimizing strategy in multi-dimensional bounded forecasting games. *Stochastic Analysis and Applications*, 121, 155-183, 2011.

[10] M. Kumon. Studies of information quantities and information geometry of higher order cumulant spaces.

***Statistical Methodology*, 7, 152-172, 2010.**

[11] M. Kumon, A. Takemura and K. Takeuchi. Conformal geometry of statistical manifold with application to sequential estimation. *Sequential Analysis*, 30, 308-337, 2011.