논문 18-43-12-01 The Journal of Korean Institute of Communications and Information Sciences '18-12 Vol.43 No.12
https://doi.org/10.7840/kics.2018.43.12.1987

마르코프 체인 출력 수열의 최소 엔트로피 추정법에 대한 분석

김 원 태◆, 박 호 중*, 염 용 진**, 강 주 성°

Analysis of Min-Entropy Estimation of Markov Chain Output Sequences

Wontae Kim◆, Hojoong Park*, Yongjin Yeom**, Ju-Sung Kang°

요 약

마르코프 체인은 현재와 과거에 대한 정보가 주어졌을 때, 미래에 대한 정보는 과거의 정보에 의존하지 않고 현재의 정보에 의해서만 결정되는 현상을 위한 확률 모델이다. 암호 시스템에서 암호 키 등의 핵심 요소를 생성하는 암호학적 난수발생기는 충분한 엔트로피를 갖는 예측 불가능한 잡음원을 사용해야 한다. 다양한 잡음원에 대한 최소 엔트로피 추정법 중에서 마르코프 체인 모델을 기반으로 하는 추정법은 출력 난수열 사이에 존재할 수 있는 종속성을 감지해내기 위한 것이다. 본 논문에서는 1 차 마르코프 체인 모델 하에서 출력 비트열의 최소 엔트로피는 3 가지 형태로 분류된 비트열에 의해 추정 가능함을 증명한다. 또한, 증명한 이론을 근거로 난수발생기의 최소 엔트로피 평가 방법을 다룬 NIST 표준인 SP 800-90B의 마르코프 추정량이 이론적 최소 엔트로피에 부합하는지를 실험적으로 확인한다.

Key Words : Markov Chain, Min-Entropy, NIST SP 800-90B, Entropy Estimation, Random Number Generator

ABSTRACT

Markov chain is a stochastic model for the phenomena where information about the present and past is given, information about the future is determined only by the current information without relying on past information. Cryptographically secure random number generators that generate important elements such as encryption keys in cryptographic systems should use unpredictable noise sources with sufficient entropy. Among the various min-entropy estimation methods for noise sources, an estimation method based on the Markov chain model is designed to detect a dependency that may exist between output random numbers. In this paper, we prove that the minimum entropy of the output bitstream under the first Markov chain model can be estimated by the

※ 본 연구는 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2014-6-00908, 난수발생기 및 임베디드 기기 안전성 연구)

◆ First Author : (ORCID:0000-0002-3817-2452)Kookmin University Department of Financial Information Security, kwt123@kookmin.ac.kr, 학생회원

° Corresponding Author : (ORCID:0000-0002-0846-389X)Kookmin University Department of Information Security, Cryptology, and Mathematics, jskang@kookmin.ac.kr, 정회원

* (ORCID:0000-0003-1179-876X)Kookmin University Department of Financial Information Security, ruokay@kookmin.ac.kr, 학생회원

** (ORCID:0000-0002-8240-8661)Kookmin University Department of Information Security, Cryptology, and Mathematics, salt@kookmin.ac.kr, 종신회원

논문번호 : 201809-288-A-RN, Received September 21, 2018; Revised November 16, 2018; Accepted December 3, 2018

bitstream that we have classified into three types. Based on the proposed theory, we experimentally confirm that Markov estimator of SP 800-90B, which is a NIST standard for evaluating the min-entropy of a random number generator, is suitable with the theoretical min-entropy.

Ⅰ. 서 론

마르코프 체인(Markov chain)은 현재와 과거에 대한 정보가 주어졌을 때, 미래 정보의 조건부 확률이 과거의 정보에 의존하지 않고 현재의 정보만 주어진 조건부 확률과 동일한 확률과정(Stochastic process)이다. 마르코프 체인은 IID(Independent and identically distributed) 성질을 만족하지 않는 대표적인 non-IID 확률과정으로 확률적 정의는 다음과 같다.

상태공간(State space) $S=\{0,1,\ldots,s\}$ 위에 정의된 확률변수들의 수열인 확률과정(Stochastic process) $\widehat{X}=\{X_0,X_1,\ldots\}$에 대해 $\Pr[X_i=x_i \mid X_{i-1}=x_{i-1}, X_{i-2}=x_{i-2},\ldots X_0=x_0] = \Pr[X_i=x_i \mid X_{i-1}=x_{i-1},\ldots X_{i-k}=x_{i-k}]$ 를 만족하면 $\widehat{X}$ 를 k 차 마르코프 체인이라 한다. 마르코프 체인 모델은 기상 변화, 온도 변화 등의 물리현상에 적용될 수 있으며 음성, 필기, 동작, 물품 구매 성향 등의 패턴 인식 문제에 활용될 수 있다. 한편, 정보이론에서 Shannon[1]에 의해 정의된 엔트로피(Entropy)는 불확실성에 대한 정량적 평균 수치를 나타낸다. 상태공간 S 를 갖는 확률변수 X 의 Shannon 엔트로피 $H_1(X)$ 는 (1)과 같다.

$$H_1(X) = -\sum_{x\in S}\Pr[X=x]\log_2(\Pr[X=x]). \quad (1)$$

직관적으로 엔트로피가 높으면 확률 분포가 균일하여 출현할 값에 대한 예측 가능성이 낮고, 엔트로피가 낮게 되면 특정 표본값의 발생 확률이 높아 예측 가능성이 높다고 볼 수 있다. 여기서 가장 예측 가능성이 높은 표본값의 불확실성을 측정하는 최소 엔트로피(Min-entropy)는 Rényi[2]에 의해 다음과 같이 정의된다. 상태공간 S 를 갖는 확률변수 X 의 최소 엔트로피 $H_\infty(X)$ 는 (2)와 같이 정의된다.

$$\begin{aligned} H_\infty(X) &= \min_{x\in S}\{-\log_2\Pr[X=x]\} \\ &= -\log_2(\max_{x\in S}\{\Pr[X=x]\}). \end{aligned} \quad (2)$$

이때, 최소 엔트로피는 보수적인 측도로 알려져 있으며 암호학에서는 암호 키(Encryption key)와 같은 비밀 정보의 안전성을 측정하는 기준으로 활용된다[3].

암호학적 난수발생기(Cryptographically secure random number generator)의 난수성 평가 방법을 다루는 대표적인 표준 문서로는 독일 BSI(Bundesamt für Sicherheit in der Informationstechnik)의 AIS.31[4]과 미국 NIST(National Institute of Standards and Technology)의 SP 800-22[5], SP 800-90B[3]가 있다. 이 중 AIS.31과 SP 800-22는 표본열의 난수성 평가를 위하여 각각 유의성 검정 방법으로 설계된 8 가지, 15 가지의 평가방법으로 표본열의 난수성을 평가한다. 이때, 유의성 검정에서는 표본열이 이상적 난수와의 구분 가능•불가능을 판정하기 위해 IID이며 균등한(Uniform) 확률모델이라는 가정 하에서 진행된다. 한편, SP 800-90B에서는 위의 두 표준과 같이 표본열이 이상적인 난수와 구별이 되는지를 판정하는 것이 아니라, 표본열의 샘플 당 최소 엔트로피를 정량적으로 추정하는 것에 목적이 있으며, 엔트로피 추정은 표본열이 미지의 분포(Unknown distribution)라는 가정 하에서 진행된다. SP 800-90B는 총 10 가지 최소 엔트로피 추정법을 이용하여 엔트로피를 추정하며, 그중 마르코프 체인을 고려해 설계한 추정법은 마르코프 추정법(Markov estimate)과 다중 마르코프 모델(MultiMMC, Multi Markov Model with Counting) 예측 추정법, 두 가지가 있다.

마르코프 추정법은 상태공간 S가 $S=\{0,1\}$인 1차 마르코프 체인 모델 하에서 출력된 비트열을 가정하여 조건부 확률 분포를 도출해 최소 엔트로피를 계산한다. 한편, SP 800-90B가 최종안으로 개정되면서 마르코프 추정법에서 수정된 사항으로는 전이확률의 계산에 두 번째 초안[6]에서 적용한 오차확률(ϵ)을 배제한 것과 최대 확률을 갖는 마르코프 체인이 표 1과 같이 총 6 가지 비트열 형태로 귀결시킨 것이 있다. 그러나 이와 관련된 이론적 근거는 제시되지 않고 있다. 이에 대해 3 장의 결과를 근거로 4 장에서 논의할 것이다.

다중 마르코프 모델은 예측 최소 엔트로피 추정법에 대한 연구[7]에서 제안된 추정법 중 하나로 총 16 개의 예측 함수로 구성된다. 각 예측 함수는 표본열

표 1. 마르코프 추정법의 최대 확률 비트열 결정 표
Table 1. Maximum probability bit sequence determination table in Markov estimator

Sequence	Probability
0000...00	$P_0 \times P_{0,0}{}^{127}$
0101...01	$P_0 \times P_{0,1}{}^{64} \times P_{1,0}{}^{63}$
0111...11	$P_0 \times P_{0,1} \times P_{1,1}{}^{126}$
1000...00	$P_1 \times P_{1,0} \times P_{0,0}{}^{126}$
1010...10	$P_1 \times P_{1,0}{}^{64} \times P_{0,1}{}^{63}$
1111...11	$P_1 \times P_{1,1}{}^{127}$

을 처음부터 순서대로 관측하며 다음 값을 예측한다. 1 번부터 16 번까지의 예측 함수는 1 차부터 16 차까지의 마르코프 모델에 대응하여 마르코프 예측 체인들을 구성해 다음 값을 예측한다. 16 개의 예측 함수 중 가장 많이 맞추고 있는 예측 함수의 예측값을 다중 마르코프 모델의 예측값으로 결정하고, 예측 성공 결과 여부에 따라 최소 엔트로피를 계산한다. 일반적으로 예측성공 횟수가 높을수록 최소 엔트로피는 낮게 계산된다. 만약 난수발생기의 출력이 2 차 이상의 마르코프 체인을 따르는 경우, 다중 마르코프 모델 예측 추정법에서 마르코프 추정법보다 적절히 추정할 것으로 예상한다.

본 논문에서는 초기확률(Initial probability)이 $\frac{1}{2}$ 인 1 차 마르코프 체인 비트열의 최소 엔트로피 계산을 연구한 결과[8]를 일반화하여, 임의의 초기확률을 갖는 1 차 마르코프 체인 모델 하에서 출력 비트 당 최소 엔트로피를 3 가지 형태로 분류된 비트열에 의해 추정 가능함을 증명한다. 이를 위하여, 2 장에서는 3 장의 증명을 위한 기호들을 정의한다. 3 장에서는 1 차 마르코프 체인의 최소 엔트로피에 대한 정리들과 그에 대한 증명을 제시한다. 4 장에서는 3 장에서 제시한 이론을 근거로 SP 800-90B의 마르코프 추정법에 대해 최종안과 두 번째 초안의 차이점을 분석한다. 나아가 마르코프 성질을 갖는 비트열을 구성해 이에 대한 SP 800-90B의 마르코프 추정법의 최소 엔트로피 추정값인 마르코프 추정량을 실험적으로 도출하여 이론적 최소 엔트로피와의 부합성을 확인한다. 5 장은 결론으로 논문의 결과를 정리하고 그 의의를 밝힌다.

Ⅱ. 정의 및 기호

2장에서는 길이 $n+1$의 확률과정 $\widehat{X}^{(n)} = \{X_0, X_1, \dots, X_n\}$은 상태공간 $S=\{0,1\}$ 위에 정의된 $n+1$개 확률변수 X_i $(0 \leqq i \leqq n)$로 이루어진 확률변수 열을 의미한다. 확률과정 $\widehat{X}$는 $\lim_{n\to\infty} H_\infty(\widehat{X}^{(n)})$을 의미하도록 하자.

정의 1. 확률과정 $\widehat{X} = \{X_0, X_1, \dots\}$는 (3)이 만족되면 1 차 마르코프 체인 비트열에 대한 확률과정이다.

$$\begin{aligned} &\forall x_0,\dots,x_i \in \{0,1\},\ i \geqq 1, \\ &\Pr[X_i = x_i | X_{i-1} = x_{i-1}, \dots X_0 = x_0] \\ &= \Pr[X_i = x_i | X_{i-1} = x_{i-1}]. \end{aligned} \tag{3}$$

$x, y \in \{0,1\}$에 대해 초기확률 $\Pr[Y=y]$를 π_y로, 전이확률 $\Pr[Y=y|X=x]$를 T_{xy}로 표기하겠다. 이에 따른 전이행렬(Transition matrix)은 (4)와 같다. 본 논문에서는 $0 < T_{xy}, \pi_y < 1$ 를 가정한다.

$$\begin{pmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{pmatrix} = \begin{pmatrix} T_{00} & 1-T_{00} \\ 1-T_{11} & T_{11} \end{pmatrix} = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}. \tag{4}$$

정의 2. 길이 $n+1$의 1 차 마르코프 체인 비트열의 비트 당 최소 엔트로피 $H(\widehat{X}^{(n)})$는 (5)와 같다.

$$H_\infty\left(\widehat{X}^{(n)}\right) = \frac{-\log_2\left(\max_{(x_0,\dots,x_n)\in\{0,1\}^{n+1}}\left(\pi_{x_0}\prod_{i=0}^{n-1} T_{x_i x_{i+1}}\right)\right)}{n+1}. \tag{5}$$

정의 3. 마르코프 체인 비트열의 비트당 최소 엔트로피 $H_\infty(\widehat{X})$는 (6)과 같다.

$$H_\infty(\widehat{X}) = \lim_{n\to\infty} H_\infty(\widehat{X}^{(n)}). \tag{6}$$

정의 4. $n \geqq 1$에 대해 길이 $n+1$의 1 차 마르코프 체인 비트열에서 발생 가능한 모든 출현 확률들의 집합 $\Lambda^{(n)}$는 (7)과 같이 정의한다.

$$\Lambda^{(n)} = \left\{ \pi_{x_0}\prod_{i=0}^{n-1} T_{x_i x_{i+1}} \,|\, (x_0,\dots,x_n) \in \{0,1\}^{n+1} \right\}. \tag{7}$$

3 장에서의 증명을 위해 $\Lambda_0^{(n)}$, $\Lambda_1^{(n)}$를 각각 $\Lambda^{(n)}$에서 n번째 표본(x_n)이 0 또는 1 인 경우로 추가 정의한다.

$$\Lambda_0^{(n)} = \left\{ \pi_{x_0} \prod_{i=0}^{n-1} T_{x_i x_{i+1}} | (x_0, ..., x_{n-1}) \in \{0,1\}^n, x_n = 0 \right\} \quad (8)$$

이고, $\Lambda^{(n)} = \Lambda_0^{(n)} \cup \Lambda_1^{(n)}$의 관계를 갖는다.

정의 5. $n \geqq 1$에 대해 길이 $n+1$의 1 차 마르코프 체인 비트열에서 발생 가능한 모든 출현 확률들에서 초기확률(π_0, π_1)을 제외한 집합 $\Gamma^{(n)}$은 (9)와 같이 정의한다.

$$\Gamma^{(n)} = \left\{ \prod_{i=0}^{n-1} T_{x_i x_{i+1}} | (x_0, \cdots, x_n) \in \{0,1\}^{n+1} \right\}. \quad (9)$$

(9)에서와 같이 $\Gamma^{(n)}$는 $\Lambda^{(n)}$에서 초기확률 π_0 또는 π_1을 제외한 집합이다. 3 장에서의 증명을 위해 $\Gamma_0^{(n)}$, $\Gamma_1^{(n)}$, $\tilde{\Gamma}_0^{(n)}$, $\tilde{\Gamma}_1^{(n)}$을 추가로 정의한다. $\Gamma_0^{(n)}$, $\Gamma_1^{(n)}$는 각각 $\Gamma^{(n)}$에서 n번째 표본(x_n)이 0 또는 1 인 경우로 정의한다. $\tilde{\Gamma}_0^{(n)}$, $\tilde{\Gamma}_1^{(n)}$는 각각 $\Gamma^{(n)}$에서 0번째 표본(x_0)이 0 또는 1 인 경우로 정의한다. 즉,

$$\Gamma_0^{(n)} = \left\{ \prod_{i=0}^{n-1} T_{x_i x_{i+1}} | (x_0, \cdots, x_{n-1}) \in \{0,1\}^n, x_n = 0 \right\}, \quad (10)$$

$$\overline{\Gamma}_0^{(n)} = \left\{ \prod_{i=0}^{n-1} T_{x_i x_{i+1}} | (x_1, \cdots, x_n) \in \{0,1\}^n, x_0 = 0 \right\} \quad (11)$$

이고, $\Gamma^{(n)} = \Gamma_0^{(n)} \cup \Gamma_1^{(n)} = \tilde{\Gamma}_0^{(n)} \cup \tilde{\Gamma}_1^{(n)}$의 관계를 갖는다.

정의 6. 1 차 마르코프 체인 모델 하에서 비트열 "000, 111, 010"의 초기확률을 제외한 출현 확률 V는 (12)와 같이 정의한다.

$$V = \{ T_{00} T_{00},\ T_{11} T_{11},\ T_{01} T_{10} = T_{10} T_{01} \}. \quad (12)$$

즉, 길이 3의 비트열 중 양 끝값이 동일한 경우를 의미한다. 여기에 101이 빠진 이유는 비트열 010과 초기확률을 제외한 출현 확률($T_{01} T_{10} = T_{10} T_{01}$)이 같기 때문이다.

정의 7. 양의 실수 집합 A, B에 대해 연산 $\circ$을 각 집합 원소의 실수 곱들의 집합으로 정의한다.

$$C = A \circ B = \{ ab \,|\, a \in A, b \in B \}. \quad (13)$$

Ⅲ. 1차 마르코프 체인 비트열에 대한 비트당 최소 엔트로피

3.1 정리 및 설명

보조정리 1. $\forall m \geqq 1,\ 1 \leqq n < m,$

$$\Lambda^{(m)} = \Lambda_0^{(m-n)} \circ \tilde{\Gamma}_0^{(n)} \cup \Lambda_1^{(m-n)} \circ \tilde{\Gamma}_1^{(n)}, \quad (14)$$

$$\Gamma^{(m)} = \Gamma_0^{(m-n)} \circ \tilde{\Gamma}_0^{(n)} \cup \Gamma_1^{(m-n)} \circ \tilde{\Gamma}_1^{(n)}. \quad (15)$$

보조정리 2. 유한 양의 실수 집합 A, A_1, A_2, A_3, A_4에 대하여 $A = (A_1 \circ A_2) \cup (A_3 \circ A_4)$라면 (16)을 만족한다.

$$\max(A) = \max\{ \max(A_1)\max(A_2),\ \max(A_3)\max(A_4) \}. \quad (16)$$

보조정리 1.과 보조정리 2.는 보조정리 3.을 증명하는데 활용되는 정리로, 유한 실수 집합(A)의 원소 중 최댓값은 그 집합의 모든 부분 집합($A_1 \circ A_2, A_3 \circ A_4$)들의 최댓값 중 가장 큰 값으로 결정된다는 것을 의미한다.

보조정리 3. $\forall n \geqq 0,$

$$\max(\Gamma^{(2+2n)}) = \max(\Gamma^{(2)}) \max(V)^n. \quad (17)$$

보조정리 3.은 홀수 길이(n : 짝수)를 갖는 마르코프 체인의 출현 확률 중 전이확률들(T_{xy})의 곱만으로 표현되는 부분에 대한 정리이다. 전이확률들의 곱으로 표현되는 부분의 최댓값은 앞의 2 개 전이확률의 곱 이후 $\max(V)$의 반복 제곱 형태로 고정된다는 것을 의미한다.

정리. V 가 $\{T_{00}T_{00},\ T_{11}T_{11},\ T_{01}T_{10}\}$ 일 때, 1 차 마르코프 체인 비트열에 대한 확률과정 $\widehat{X}=\{X_0,X_1,...\}$의 비트당 최소 엔트로피 $H_\infty(\widehat{X})$는 (18)과 같다.

$$H_\infty(\widehat{X})=\frac{-\log_2(\max(V))}{2}. \tag{18}$$

정리.는 1 차 마르코프 체인 비트열의 비트당 최소 엔트로피가 비트열 "000, 111, 010" 세 가지 '중 하나의 형태로 분류되어 추정 가능하다는 것을 의미한다.

3.2 정리 및 증명

보조정리 1. $\forall m \geqq 1,\ 1 \leqq n < m,$

$$\Lambda^{(m)}=\Lambda_0^{(m-n)} \circ \tilde{\Gamma}_0^{(n)} \cup \Lambda_1^{(m-n)} \circ \tilde{\Gamma}_1^{(n)}, \tag{14}$$

$$\Gamma^{(m)}=\Gamma_0^{(m-n)} \circ \tilde{\Gamma}_0^{(n)} \cup \Gamma_1^{(m-n)} \circ \tilde{\Gamma}_1^{(n)}. \tag{15}$$

증명)

$k=m-n$라 하자.

$$\Lambda_0^{(m-n)} \circ \tilde{\Gamma}_0^{(n)} = \Lambda_0^{(k)} \circ \tilde{\Gamma}_0^{(n)}$$

$$= \left\{ \pi_{x_0} \prod_{i=0}^{k-1} T_{x_i x_{i+1}} | (x_0,...,x_{k-1}) \in \{0,1\}^k,\ x_k = 0 \right\}$$

$$\circ \left\{ \prod_{i=k}^{m-1} T_{x_i x_{i+1}} | (x_{k+1},...,x_m) \in \{0,1\}^n,\ x_k = 0 \right\}$$

$$= \{ \pi_{x_0} \prod_{i=0}^{m-1} T_{x_i x_{i+1}} | (x_0,...,x_{k-1}) \in \{0,1\}^k, x_k = 0,$$

$$(x_{k+1},...,x_m) \in \{0,1\}^n \}.$$

$$\Lambda_1^{(m-n)} \circ \tilde{\Gamma}_1^{(n)} = \Lambda_1^{(k)} \circ \tilde{\Gamma}_1^{(n)}$$

$$= \left\{ \pi_{x_0} \prod_{i=0}^{k-1} T_{x_i x_{i+1}} | (x_0,...,x_{k-1}) \in \{0,1\}^k, x_k = 1 \right\}$$

$$\circ \left\{ \prod_{i=k}^{m-1} T_{x_i x_{i+1}} | (x_{k+1},...,x_m) \in \{0,1\}^n, x_k = 1 \right\}$$

$$= \{ \pi_{x_0} \prod_{i=0}^{m-1} T_{x_i x_{i+1}} | (x_0,...,x_{k-1}) \in \{0,1\}^k, x_k = 1,$$

$$(x_{k+1},...,x_m) \in \{0,1\}^n \}.$$

$$\Rightarrow \Lambda_0^{(k)} \circ \tilde{\Gamma}_0^{(n)} \cup \Lambda_1^{(k)} \circ \tilde{\Gamma}_1^{(n)}$$

$$= \{ \pi_{x_0} \prod_{i=0}^{m-1} T_{x_i x_{i+1}} | (x_0,...,x_{k-1}) \in \{0,1\}^k, x_k \in \{0,1\},$$

$$(x_{k+1},...,x_m) \in \{0,1\}^n \}$$

$$= \left\{ \pi_{x_0} \prod_{i=0}^{m-1} T_{x_i x_{i+1}} | (x_0,...,x_m) \in \{0,1\}^{m+1} \right\}.$$

이 식은 $\Lambda^{(n)}$의 정의 4.와 같다.

$$\therefore\ \Lambda^{(m)}=\Lambda_0^{(m-n)} \circ \tilde{\Gamma}_0^{(n)} \cup \Lambda_1^{(m-n)} \circ \tilde{\Gamma}_1^{(n)}. \qquad \square$$

보조정리 2. 유한 양의 실수 집합 A, A_1, A_2, A_3, A_4에 대하여 $A=(A_1 \circ A_2) \cup (A_3 \circ A_4)$라면 (16)을 만족한다.

$$\max(A)=\max\{\max(A_1)\max(A_2),\ \max(A_3)\max(A_4)\}. \tag{16}$$

$\max(A)$는 A의 부분집합 $A_1 \circ A_2$와 $A_3 \circ A_4$ 각각의 최댓값 중 더 큰 값을 의미한다. A_1, A_2는 유한 양의 실수 집합이므로 각각 최댓값이 존재하며, 그 둘을 곱한 값이 $A_1 \circ A_2$ 집합의 최댓값이다. 따라서 (16)과 같이 표현된다. 따로 증명은 하지 않겠다.

보조정리 3. $\forall n \geqq 0,$

$$\max(\Gamma^{(2+2n)})=\max(\Gamma^{(2)})\max(V)^n. \tag{17}$$

증명)

$\max(\Gamma^{(2)})$는 $\{T_{00}T_{00},\ T_{00}T_{01},\ T_{01}T_{10},\ T_{01}T_{11},\ T_{11}T_{11},\ T_{11}T_{10},\ T_{10}T_{01},\ T_{10}T_{00}\}$ 중 하나가 될 수 있다. 이 중에서 $\{T_{00}T_{00},\ T_{00}T_{01},\ T_{01}T_{10},\ T_{01}T_{11}\}$의 모든 경우에 대해 (17)이 만족한다는 것을 보일 것이다. $\{T_{11}T_{11},\ T_{11}T_{10},\ T_{10}T_{01},\ T_{10}T_{00}\}$의 경우는 동일하게 증명된다.

경우 1) $\max(\Gamma^{(2)})=T_{00}T_{00}.$

자명하게, $\max(\Gamma^{(2+2n)})=(T_{00}T_{00})^{n+1}$

$=(T_{00}T_{00})(T_{00}T_{00})^n=\max(\Gamma^{(2)})\max(V)^n.$

$\therefore\ \max(\Gamma^{(2)})=T_{00}T_{00} \rightarrow$

$\max(\Gamma^{(2+2n)})=\max(\Gamma^{(2)})\max(V)^n.$

경우 2) $\max(\Gamma^{(2)})=T_{00}T_{01}.$

$\Rightarrow T_{00}T_{01} \geqq T_{00}T_{00},\ T_{00}T_{01} \geqq T_{01}T_{10}.$

$\Rightarrow T_{01} \geqq T_{00},\ T_{00} \geqq T_{10}.$

$\Rightarrow T_{11} \geqq T_{01} \geqq T_{00} \geqq T_{10}.$

$\because T_{01}=1-T_{00},\ T_{10}=1-T_{11}.$

$\Rightarrow T_{11}T_{11} \geqq T_{00}T_{01}=\max(\Gamma^{(2)}).$

따라서 $T_{11}T_{11}$가 $T_{11}T_{11}=T_{00}T_{01}=\max(\Gamma^{(2)})$를 만족하게 되고, 경우 1)과 같은 과정을 따라 (17)을

만족한다.

$\therefore \max(\Gamma^{(2)}) = T_{00}T_{01} \rightarrow \max(\Gamma^{(2+2n)}) = \max(\Gamma^{(2)})\max(V)^n.$

경우 3) $\max(\Gamma^{(2)}) = T_{01}T_{10}.$

$T_{01}T_{10} \in \Gamma_0^{(2)}, \Gamma_1^{(2)}, \tilde{\Gamma}_0^{(2)}, \tilde{\Gamma}_1^{(2)}$ 이며 $\max(\Gamma^{(2)}) = T_{01}T_{10}$ 이기 때문에,

$T_{01}T_{10} = \max(\Gamma_0^{(2)}) = \max(\Gamma_1^{(2)}) = \max(\tilde{\Gamma}_0^{(2)}) = \max(\tilde{\Gamma}_1^{(2)})$ 이다.

$\max(\Gamma^{(2)}) = T_{01}T_{10}.$

$\max(\Gamma^{(2+2)}) = \max(\max(\Gamma_0^{(2)})\max(\tilde{\Gamma}_0^{(2)}), \max(\Gamma_1^{(2)})\max(\tilde{\Gamma}_1^{(2)}))$

$= T_{01}T_{10}T_{01}T_{10} = \max(\Gamma^{(2)})\max(V).$

$\max(\Gamma^{(2+4)}) = \max(\max(\Gamma_0^{(4)})\max(\tilde{\Gamma}_0^{(2)}), \max(\Gamma_1^{(4)})\max(\tilde{\Gamma}_1^{(2)}))$

$= (T_{01}T_{10})^2 T_{01}T_{10} = \max(\Gamma^{(2)})\max(V)^2.$

$(\because \max(\Gamma^{(4)}) = (T_{01}T_{10})^2 = \max(\Gamma_0^{(4)}) = \max(\Gamma_1^{(4)}) = \max(\tilde{\Gamma}_0^{(4)}) = \max(\tilde{\Gamma}_1^{(4)}).$)

...

$\max(\Gamma^{(2+2n)}) = \max(\Gamma^{(2)})\max(V)^n.$

$\therefore \max(\Gamma^{(2)}) = T_{01}T_{10} \rightarrow \max(\Gamma^{(2+2n)}) = \max(\Gamma^{(2)})\max(V)^n.$

경우 4) $\max(\Gamma^{(2)}) = T_{01}T_{11}.$

$\Rightarrow T_{01}T_{11} \geqq T_{11}T_{11},\ T_{01}T_{11} \geqq T_{01}T_{10}$

$\Rightarrow T_{01} \geqq T_{11},\ T_{11} \geqq T_{10}$

$\Rightarrow T_{01} \geqq T_{11} \geqq T_{10} \geqq T_{00}.$

이 경우 $\max(V)$는 $T_{10}T_{01}$, $T_{11}T_{11}$ 둘 중 하나가 될 수 있다. 각 경우에 대해 나누어 보도록 하겠다.

경우 4-1) $T_{10}T_{01} \geqq T_{11}T_{11}.$

$\max(\Gamma^{(2)}) = T_{01}T_{11}.$

$\max(\Gamma^{(2+2)}) = \max(\max(\Gamma_0^{(2)})\max(\tilde{\Gamma}_0^{(2)}), \max(\Gamma_1^{(2)})\max(\tilde{\Gamma}_1^{(2)}))$

$= \max((T_{01}T_{10})T_{01}T_{11}, (T_{01}T_{11})T_{10}T_{01})$

$(\because \max(\Gamma^{(2)}) = T_{01}T_{11} \in \tilde{\Gamma}_0^{(2)} \cap \Gamma_1^{(2)}$

$\Rightarrow \max(\tilde{\Gamma}_0^{(2)}) = \max(\Gamma_1^{(2)}) = T_{01}T_{11}.$

$T_{01}T_{10} \geqq T_{11}T_{11}$

$\Rightarrow \max(\Gamma_0^{(2)}) = T_{01}T_{10},\ \max(\tilde{\Gamma}_1^{(2)}) = T_{10}T_{01}.$)

$= T_{01}T_{11}T_{10}T_{01} = \max(\Gamma^{(2)})\max(V).$

$\max(\Gamma^{(2+4)}) = \max(\max(\Gamma_0^{(4)})\max(\tilde{\Gamma}_0^{(2)}), \max(\Gamma_1^{(4)})\max(\tilde{\Gamma}_1^{(2)}))$

$= \max((T_{01}T_{10}T_{01}T_{10})T_{01}T_{11}, (T_{01}T_{11}T_{10}T_{01})T_{10}T_{01}).$

$(\because \max(\Gamma^{(4)}) = T_{01}T_{11}T_{10}T_{01} \in \Gamma_1^{(4)}.$

$\Rightarrow \max(\Gamma_1^{(4)}) = T_{01}T_{11}T_{10}T_{01}.$

$\Gamma_0^{(4)} = \Gamma_0^{(2)} \circ \{T_{00}T_{00},\ T_{01}T_{10}\} \cup \Gamma_1^{(2)} \circ \{T_{10}T_{00},\ T_{11}T_{10}\}.$

보조정리 2.에 의해,

$\max(\Gamma_0^{(4)}) = \max\{\max(\Gamma_0^{(2)})\max\{T_{00}T_{00},\ T_{01}T_{10}\}, \max(\Gamma_1^{(2)})\max\{T_{10}T_{00},\ T_{11}T_{10}\}\}$

$= \max\{T_{01}T_{10}T_{01}T_{10}, T_{01}T_{11}T_{11}T_{10}\}$

$= (T_{01}T_{10})^2.$)

$= T_{01}T_{11}(T_{01}T_{10})^2 = \max(\Gamma^{(2)})\max(V)^2.$

$\max(\Gamma^{(2+6)}) = \max(\max(\Gamma_0^{(6)})\max(\tilde{\Gamma}_0^{(2)}), \max(\Gamma_1^{(6)})\max(\tilde{\Gamma}_1^{(2)}))$

$= \max((T_{01}T_{10})^3 T_{01}T_{11}, (T_{01}T_{11}T_{10}T_{01}T_{10}T_{01})T_{10}T_{01}).$

(위의 $\max(\Gamma^{(2+4)})$ 값의 도출 방식과 유사하게 도출할 수 있다.)

$= T_{01}T_{11}(T_{01}T_{10})^3 = \max(\Gamma^{(2)})\max(V)^3.$

...

$\max(\Gamma^{(2+2n)}) = \max(\Gamma^{(2)})\max(V)^n.$

$\therefore \max(\Gamma^{(2)}) = T_{01}T_{11},\ T_{10}T_{01} \geqq T_{11}T_{11} \rightarrow \max(\Gamma^{(2+2n)}) = \max(\Gamma^{(2)})\max(V)^n.$

경우 4-2) $T_{10}T_{01} < T_{11}T_{11}$.

경우 4-1)과 유사한 순서로 다음이 증명된다.

$\max(\Gamma^{(2+2n)}) = \max(\Gamma^{(2)})\max(V)^n$.

$\therefore \max(\Gamma^{(2)}) = T_{01}T_{10},\ T_{10}T_{01} < T_{11}T_{11} \rightarrow \max(\Gamma^{(2+2n)}) = \max(\Gamma^{(2)})\max(V)^n$.

$\max(\Gamma^{(2)})$가 $\{T_{11}T_{11},\ T_{11}T_{10},\ T_{10}T_{01},\ T_{10}T_{00}\}$ 중 하나일 경우에 대해서는 위와 동일하게 증명된다. 따라서 (17)은 항상 만족한다. □

정리. V가 $\{T_{00}T_{00},\ T_{11}T_{11},\ T_{01}T_{10}\}$ 일 때, 1 차 마르코프 체인 비트열에 대한 확률과정 $\widehat{X} = \{X_0, X_1, \ldots\}$의 최소 엔트로피율(Min-entropy rate)은 (18)과 같다.

$$H_\infty(\widehat{X}) = \frac{-\log_2(\max(V))}{2}. \tag{18}$$

증명)

보조정리 1.에 의하여, $\forall n \geqq 0$,

$\Lambda^{(3+2n)} = \Lambda_0^{(1)} \circ \tilde{\Gamma}_0^{(2+2n)} \cup \Lambda_1^{(1)} \circ \tilde{\Gamma}_1^{(2+2n)}$.

보조정리 2.에 의하여,

$\max(\Lambda^{(3+2n)}) = \max\{\max(\Lambda_0^{(1)})\max(\tilde{\Gamma}_0^{(2+2n)}),\ \max(\Lambda_1^{(1)})\max(\tilde{\Gamma}_1^{(2+2n)})\}$.

$\Lambda^{(1)} = \Lambda_0^{(1)} \cup \Lambda_1^{(1)}$.

$\Rightarrow \max(\Lambda^{(1)}) = \max\{\max(\Lambda_0^{(1)}), \max(\Lambda_1^{(1)})\}$.

$\Gamma^{(2+2n)} = \tilde{\Gamma}_0^{(2+2n)} \cup \tilde{\Gamma}_1^{(2+2n)}$.

$\Rightarrow$

$\max(\Gamma^{(2+2n)}) = \max\{\max(\tilde{\Gamma}_0^{(2+2n)}), \max(\tilde{\Gamma}_1^{(2+2n)})\}$.

$\max(\Lambda^{(3+2n)})$는 $\max(\Gamma^{(2+2n)})$와 $\max(\Gamma^{(2+2n)})$에 따라 총 4 가지로 결정될 수 있다. 그러나 4 가지 경우에 모두에서

$$\min\{\max(\Lambda_0^{(1)}), \max(\Lambda_1^{(1)})\}\max(\Gamma^{(2+2n)}) \leqq \max(\Lambda^{(3+2n)}) \leqq \max(\Lambda^{(1)})\max(\Gamma^{(2+2n)}). \tag{19}$$

경우 1) $\max(\Lambda^{(1)}) = \max(\Lambda_0^{(1)})$ 이며,

$\max(\Gamma^{(2+2n)}) = \max(\tilde{\Gamma}_0^{(2+2n)})$.

$\max(\Lambda^{(3+2n)}) = \max(\Lambda_0^{(1)})\max(\tilde{\Gamma}_0^{(2+2n)})$.

$\min\{\max(\Lambda_0^{(1)}), \max(\Lambda_1^{(1)})\} = \max(\Lambda_1^{(1)})$.

$\Rightarrow \max(\Lambda_1^{(1)})\max(\tilde{\Gamma}_0^{(2+2n)}) \leqq \max(\Lambda_0^{(1)})\max(\tilde{\Gamma}_0^{(2+2n)}) \leqq \min(\Lambda^{(1)})\max(\Gamma^{(2+2n)})$.

경우 2) $\max(\Lambda^{(1)}) = \max(\Lambda_0^{(1)})$ 이며, $\max(\Gamma^{(2+2n)}) = \max(\tilde{\Gamma}_1^{(2+2n)})$.

$\min\{\max(\Lambda_0^{(1)}), \max(\Lambda_1^{(1)})\} = \max(\Lambda_1^{(1)})$.

$\Rightarrow \max(\Lambda_1^{(1)})\max(\tilde{\Gamma}_1^{(2+2n)}) \leqq \max\{\max(\Lambda_0^{(1)})\max(\tilde{\Gamma}_0^{(2+2n)}), \max(\Lambda_1^{(1)})\max(\tilde{\Gamma}_1^{(2+2n)})\} \leqq \min(\Lambda^{(1)})\max(\Gamma^{(2+2n)})$.

경우 3) $\max(\Lambda^{(1)}) = \max(\Lambda_1^{(1)})$ 이며, $\max(\Gamma^{(2+2n)}) = \max(\tilde{\Gamma}_0^{(2+2n)})$.

경우 2)와 동일한 과정으로 증명된다.

경우 4) $\max(\Lambda^{(1)}) = \max(\Lambda_1^{(1)})$ 이며, $\max(\Gamma^{(2+2n)}) = \max(\tilde{\Gamma}_1^{(2+2n)})$.

경우 1)과 동일한 과정으로 증명된다.

(19)와 보조정리 3.에 의하여

$\min\{\max(\Lambda_0^{(1)}), \max(\Lambda_1^{(1)})\}\max(\Gamma^{(2)})\max(V)^n \leqq \max(\Lambda^{(3+2n)}) \leqq \max(\Lambda^{(1)})\max(\Gamma^{(2)})\max(V)^n$.

정의 3. 그리고 정의 4.에 의해

$H_\infty(\widehat{X}^{(3+2n)}) = \frac{-\log_2(\max(\Lambda^{(3+2n)}))}{4+2n}$.

$\Rightarrow -\log_2(\min\{\max(\Lambda_0^{(1)}), \max(\Lambda_1^{(1)})\}\max(\Gamma^{(2)})\max(V)^n)/(3+2n) \leqq H_\infty(\widehat{X}^{(3+2n)}) = \frac{-\log_2(\max(\Lambda^{(3+2n)}))}{4+2n} \leqq \frac{-\log_2(\max(\Lambda^{(1)})\max(\Gamma^{(2)})\max(V)^n)}{4+2n}$.

$\Rightarrow \lim_{n\to\infty} -\log_2(\min\{\max(\Lambda_0^{(1)}), \max(\Lambda_1^{(1)})\}\max(\Gamma^{(2)})\max(V)^n)/(4+2n) \leqq \lim_{n\to\infty} H_\infty(\widehat{X}^{(3+2n)}) \leqq \lim_{n\to\infty} \frac{-\log_2(\max(\Lambda^{(1)})\max(\Gamma^{(2)})\max(V)^n)}{4+2n}$.

$\Rightarrow \lim_{n\to\infty} \frac{-\log_2(\min\{\max(\Lambda_0^{(1)}), \max(\Lambda_1^{(1)})\})}{4+2n}$

$$+\frac{-\log_2(\max(\Gamma^{(2)}))}{4+2n}+\frac{-\log_2(\max(V)^n)}{4+2n}$$
$$\leqq \lim_{n\to\infty} H_\infty(\widehat{X}^{(3+2n)})$$
$$\leqq \lim_{n\to\infty}\frac{-\log_2(\max(\Lambda^{(1)}))}{4+2n}+\frac{-\log_2(\max(\Gamma^{(2)}))}{4+2n}$$
$$+\frac{-\log_2(\max(V)^n)}{4+2n}.$$
$$\Rightarrow \lim_{n\to\infty}\frac{-n\log_2(\max(V))}{4+2n}\leqq \lim_{n\to\infty} H_\infty(\widehat{X}^{(3+2n)})$$
$$\leqq \lim_{n\to\infty}\frac{-n\log_2(\max(V))}{4+2n}.$$
$$\Rightarrow -\frac{\log_2(\max(V))}{2}\leqq \lim_{n\to\infty} H_\infty(\widehat{X}^{(3+2n)})$$
$$\leqq -\frac{\log_2(\max(V))}{2}.$$

$\Rightarrow$ 미적분학의 샌드위치 정리에 의하여,

$$\lim_{n\to\infty} H_\infty(\widehat{X}^{(3+2n)}) = -\frac{\log_2(\max(V))}{2}. \quad (20)$$

정의 2.와 정의 3. 그리고 정의 4.에 의해 확률과정 $\widehat{X}=\{X_0,X_1,\ldots\}$의 비트 당 최소 엔트로피 $H_\infty(\widehat{X})$은 다음과 같이 표현된다.

$$H_\infty(\widehat{X}) = \lim_{n\to\infty} H_\infty(\widehat{X}^{(n)})$$
$$= \lim_{n\to\infty}\frac{-\log_2(\max(\Lambda^{(n)}))}{n}.$$

만약 n이 홀수라면, (20)이 성립한다. 즉, $H_\infty(\widehat{X})$
$= \lim_{n\to\infty} H_\infty(\widehat{X}^{(n)}) = -\frac{\log_2(\max(V))}{2}$ 이다.

정의 3.과 정의 4.에 의해, $H(\widehat{X}^{(n)}) = \frac{-\log_2(\max(\Lambda^{(n)}))}{n+1}$이며, 다음은 자명하다.

$$H_\infty(\widehat{X}^{(n)})(n+1)\leqq H_\infty(\widehat{X}^{(n+1)})(n+2)$$
$$\leqq H_\infty(\widehat{X}^{(n+2)})(n+3).$$
$$\Rightarrow$$
$$\lim_{n\to\infty}\frac{H_\infty(\widehat{X}^{(n)})(n+1)}{n}\leqq \lim_{n\to\infty}\frac{H_\infty(\widehat{X}^{(n+1)})(n+2)}{n}$$
$$\leqq \lim_{n\to\infty}\frac{H_\infty(\widehat{X}^{(n+2)})(n+3)}{n}.$$
$$\Rightarrow \lim_{n\to\infty} H_\infty(\widehat{X}^{(n)}) \leqq \lim_{n\to\infty} H_\infty(\widehat{X}^{(n+1)})$$
$$\leqq \lim_{n\to\infty} H_\infty(\widehat{X}^{(n+2)}). \quad (21)$$

(21)에서 n이 홀수라면 $n+2$도 홀수이기 때문에,

$$\lim_{n\to\infty} H_\infty(\widehat{X}^{(n)}) = -\frac{\log_2(\max(V))}{2} = \lim_{n\to\infty} H_\infty(\widehat{X}^{(n+2)})$$

를 만족하고 미적분학의 샌드위치 정리에 의하여

$\lim_{n\to\infty} H_\infty(\widehat{X}^{(n+1)}) = -\frac{\log_2(\max(V))}{2}$ 를 만족한다.

즉, n이 짝수일 때도 $\lim_{n\to\infty} H_\infty(\widehat{X}^{(n)}) = -\frac{\log_2(\max(V))}{2}$ 이다.

따라서 $H_\infty(\widehat{X}) = -\frac{\log_2(\max(V))}{2}$.

Ⅳ. SP 800-90B 마르코프 추정법 분석 및 시뮬레이션 결과

2018년 1월 미국의 NIST는 암호학적 난수발생기에 대한 최소 엔트로피 추정 방법을 다루는 표준 문서인 SP 800-90B를 두 번째 초안(Second draft)[6]에서 정식 문서로 최종 개정하였다. 최종 문서의 마르코프 추정법에서 변경된 사항은 크게 두 가지로, 하나는 초기확률과 전이확률의 추정과정에서 기존 적용되었던 오차확률(ϵ)을 제거한 것이며, 다른 하나는 길이 128의 체인에 대한 최대 출현 확률을 초기확률과 전이확률의 곱에 대한 반복 비교를 통해 도출하던 기존의 방식에서 6 개 비트열의 확률에 대해서만 비교하도록 표 1을 제시한 것이다. 이 두 가지 변경 사항에 대한 이론적 근거는 현재까지 제시되지 않았지만, 기존 제시되어 왔던 엔트로피 저평가에 대한 해결과 계산적 효율성 개선이 목적이라 사료된다.

3 장에서 증명한 결론으로 표 1을 볼 때, 비트열 2 개가 62 회 반복되는 형태를 갖도록 구성하는데 이는 자연스러운 형태로 보인다. 그러나 SP 800-90B의 두 번째 초안과 최종 문서에서의 마르코프 추정법의 변경 사항을 분석한 연구[9]에서는 최종안의 표 1에서 제시한 6 개 비트열은 길이 128 마르코프 체인의 모든 최대 출현 확률을 표현하지 못하고 있다고 지적하였다. 따라서 실제 마르코프 성질을 갖는 비트열에 대한 SP 800-90B 마르코프 추정량의 적절성을 확인할 필요가 있다.

이를 위해 SP 800-90B의 마르코프 추정량과 이론적 최소 엔트로피의 차이를 통해 마르코프 추정법의 적절성을 실험적으로 확인하였다. 실험 과정은 그림 1

과 같다. 먼저 전이확률 T_{00}, T_{01}, T_{10}, T_{11}을 갖는 마르코프 체인 모델 하의 비트열을 그림 2의 의사코드를 이용해 구성하였다. 구성한 비트열을 100만 표본(평가를 위한 권장 입력 표본 개수)에 대해 SP 800-90B의 2차 초안과 최종안 각각의 마르코프 추정량을 도출하여 전이확률에 따른 이론치와 비교하였으며 그 결과는 그림 3과 같다. 여기서 전이확률은 엔트로피 결정에 있어 $-\log_2 T_{11}$와 $-\log_2 T_{00}$의 대칭되는 결과를 배제하고, 최소 엔트로피가 $-\log_2 T_{11}$과 $-\log_2(T_{01}T_{10})/2$로 결정되는 경우를 가능한 균등하

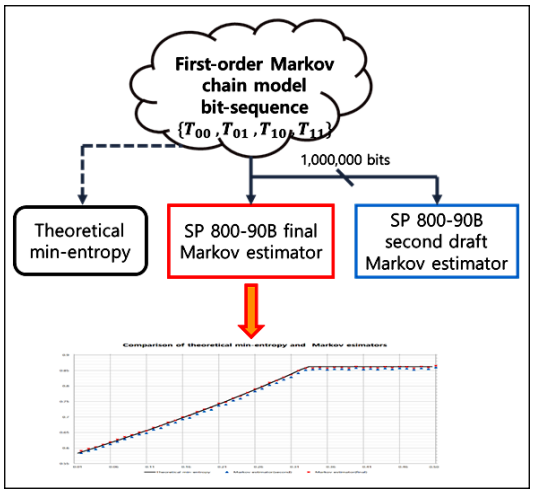

그림 1. 마르코프 추정량에 대한 적절성 분석 실험 전략
Fig. 1. The scheme of the analysis suitability experiment for Markov estimator

Markov source Generator

input : 1,000,000 bytes $\{x_0, \ldots, x_{999,999}\}$
(from real random source)
transition matrix : T_{00}, T_{01}, T_{10}, T_{11}
for $i \in Z_{1,000,000}$
$t_i = x_i/255$
if $t_0 < T_{00}/(T_{10}T_{01})$ then $y_0 = 0$
else $y_0 = 1$
for $i \in Z_{999,999}$
if $y_i = 0$
if $t_{i+1} < T_{00}$ then $y_{i+1} = 0$
else $y_{i+1} = 1$
else
if $t_{i+1} < T_{11}$ then $y_{i+1} = 1$
else $y_{i+1} = 0$
return $\{y_0, \ldots, y_{999,999}\}$

그림 2. 1차 마르코프 체인 비트열 생성기의 의사코드
Fig. 2. The pseudo-code of the first-order Markov chain source generator

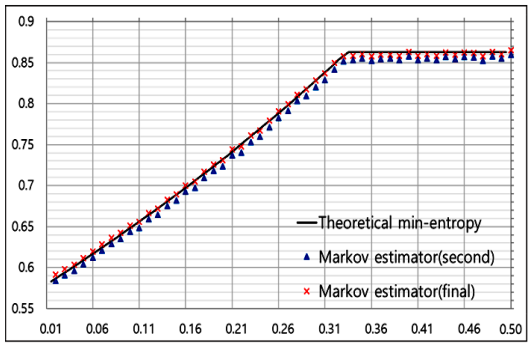

그림 3. SP 800-90B 마르코프 추정량(second draft, final)과 이론치 비교 결과 (x 축 : T_{00}, y 축 : min-entropy)
Fig. 3. Comparison of Markov estimators(second draft and final) and theoretical min-entropy (x-axis : T_{00}, y-axis : min-entropy)

게 관측하기 위하여 T_{11}는 0.55로 고정시키고, T_{00}는 0.01에서 0.5까지 0.01의 간격을 두도록 설계하였다.

실험 결과 SP 800-90B 두 번째 초안과 최종안 각각의 마르코프 추정량은 이론치와의 차이가 최대 0.01로 발생했다. 이는 실질적인 최소 엔트로피 추정에 있어 유의미한 차이로 보기 힘든 값으로, SP 800-90B의 마르코프 추정량은 이론치에 부합한다고 보기에 충분하다. 추가로 실험한 모든 전이확률에서 두 번째 초안의 마르코프 추정량은 최종안의 마르코프 추정량보다 항상 낮게 측정되는 것을 확인할 수 있었는데, 이는 최종안으로 개정되면서 오차확률 적용을 배제한 결과다.

V. 결 론

본 논문에서는 임의의 초기확률을 가지는 1 차 마르코프 체인 비트열의 최소 엔트로피가 "000, 111, 010"의 세 가지 형태로 수렴한다는 것을 증명하였다. 정리.를 논문에서 제시한 이론을 기반으로 기존에 그 이론적 근거가 명확히 제시되지 않은 NIST SP 800-90B의 마르코프 추정법에 대한 이론치와의 부합성을 실험적으로 검증하였다. 구체적으로 증명한 마르코프 체인의 이론적 최소 엔트로피와 SP 800-90B 두 번째 초안과 최종안의 마르코프 추정량의 차이가 최대 0.01임을 실험적으로 확인하였다. 이 결과는 본 논문의 분석이 SP 800-90B 최종안의 마르코프 추정법을 규명함을 보여주며, NIST SP800-90B의 마르코프 추정법의 이론적 근거가 될 수 있음을 의미한다.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical J.*, vol. 27, no. 3, pp. 379-423, Jul. 1948.

[2] A. Rényi, "On measures of entropy and information," in *Proc. The Fourth Berkeley Symp. Mathematics, Statistics and Probability*, vol. 1, pp. 547-561, 1960.

[3] M. S. Turan, et al., "*Recommendation for the entropy sources used for random bit generation*," NIST SP 800-90B, Jan. 2018.

[4] W. Killmann and W. Schindler, "*Functionality classes and evaluation methodology for random number generators*," AIS.20/AIS.31, Sep. 2011.

[5] A. Rukhin, et. al., "*A statistical test suite for random and pseudorandom number generators for cryptographic applications*," NIST SP 800-22 revision 1a, Apr. 2010.

[6] M. S. Turan, et. al., "*Recommendation for the entropoy sources used for random bit generation*," (Second Draft) NIST SP 800-90B, Jan. 2016.

[7] J. Kelsey, K. A. McKay and M. S. Turan, "Predictive models for min-entropy estimation," *CHESS 2015*, vol. 9293, pp. 373-392, Sep. 2015.

[8] W. Kim, Y. Yeom, and J. S. Kang, "On min-entropy estimation for bit-sequences from Markov chain sources," in *Proc. The Int. Conf. Inf. Soc. 2018*, Jul. 2018.

[9] H.E. Kim, W. Kim, J.S. Kang and Y. Yeom, "A study on the Markov statistic in entropy estimation method of NIST SP 800-90B," in *Proc. Symp. KICS*, pp. 25-26, Jun. 2018.

김 원 태 (Wontae Kim)

2017년 2월 : 국민대학교 수학과 학사

2017년 3월~현재 : 국민대학교 금융정보보안학과 석사과정

<관심분야> 난수성 분석 및 평가, 프로그램 고속구현, 대칭키 암호 분석

박 호 중 (Hojoong Park)

2015년 2월 : 국민대학교 수학과 학사

2017년 2월 : 국민대학교 금융정보보안학과 석사

2017년 3월~현재 : 국민대학교 금융정보보안학과 박사과정

<관심분야> 암호이론, 정보보호 알고리즘 및 프로토콜, 난수성 분석

염 용 진 (Yongjin Yeom)

1991년 2월 : 서울대학교 수학과 학사

1994년 2월 : 서울대학교 수학과 석사

1999년 2월 : 서울대학교 수학과 박사

2000년 4월~2012년 2월 : ETRI 부설연구소 책임연구원/팀장

2006년 12월~2007년 12월 : Columbia 대학교 방문연구원

2012년 3월~현재 : 국민대학교 정보보안암호수학과 부교수

2013년~현재 : 국민대학교 BK21+ 미래 금융정보보안 인력양성사업단 교수

<관심분야> 암호구현 및 분석, 보안시스템 평가

강 주 성 (Ju-Sung Kang)

1989년 2월 : 고려대학교 수학과 학사
1991년 2월 : 고려대학교 일반대학원 수학과 석사
1996년 2월 : 고려대학교 일반대학원 수학과 박사
1997년~2004년 : 한국전자통신연구원 선임연구원/팀장
2001년~2002년, 2010년 : 벨기에 루벤대학 COSIC 방문 연구원
2004년~현재 : 국민대학교 정보보안암호수학과 교수
2013년~현재 : 국민대학교 BK21+ 미래 금융정보보안 인력양성사업단 교수
<관심분야> 암호이론, 정보보안 프로토콜, 안전성 분석 및 평가