IT Basics – Typical Sequences

Course: FEO3320 Information Theoretic Security

Tobias Oechtering, Somayeh Salimi

Royal Institute of Technology (KTH),
School of EE and ACCESS Center,
Communication Theory Lab
Stockholm, Sweden

Sep 19, 2013

In this lecture we will consider:

1. Basic Definitions and Properties
2. Strongly Typical Sequences
3. Weakly Typical Sequences
4. Literature

Some Notation

- Upper case letters denote random variables (RV), e.g., X
- Lower case letters denote realizations of RV or constants: X takes realization x
- Script letters denote sets: X is defined on $\mathcal{X}$ and $x \in \mathcal{X}$
- $\mathbb{P}\{$*event*$\}$ denotes probability of the *event*
- p_X denotes probability mass function (pmf) or probability density function (pdf) of RV X: $p_X(x) = \mathbb{P}\{X = x\}$ (subindex $_X$ is dropped where RV is clear)
- $\mathcal{S}_X$ denotes the support of RV X, i.e, $p(x) > 0 \; \forall \; s \in \mathcal{S}_X$
- $\mathbb{E}\{$*event*$\}$ denotes expectation of the *event*
- *LHS* := *RHS* means *RHS defines LHS*
- Sequences: $X_m^n := (X_m, X_{m+1}, \ldots, X_n)$, $m \leq n$, $X^n := X_1^n$
- $\log$ is base 2, unless specified otherwise

Basic Definitions

- **Entropy:** $H(X) \triangleq -\sum_x p_X(x) \log_2 p_X(x)$
  - average uncertainty associated with RV X
- **Conditional entropy:** $H(X|Y) \triangleq -\sum_{x,y} p_{XY}(x,y) \log_2 p_{X|Y}(x|y)$
  - average uncertainty associated with RV X given RV Y
- **Differential entropy:** $h(X) \triangleq -\int_{\mathcal{S}_X} p_X(x) \log_2 p_X(x) \mathrm{d}x$
  - For a Gaussian RV $X \sim \mathcal{N}(\mu, \sigma^2)$ we have $h(X) = \frac{1}{2}\log(2\pi e \sigma^2)$
  - **Cond. diff. entropy:** $h(X|Y) \triangleq -\int_{\mathcal{S}_{XY}} p_{XY}(x,y) \log_2 p_{X|Y}(x|y) \mathrm{d}(x,y)$
- **Mutual information:** $I(X;Y) \triangleq H(X) - H(X|Y) = H(Y) - H(Y|X)$
  - how much observation of RV Y informs us about RV X
  - For continuous RVs: $I(X;Y) \triangleq h(X) - h(X|Y)$

Basic Properties

- **Chain rule:** $H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$
- **Conditioning does not increase entropy:** $H(Y|X) \leq H(Y)$ with equality if and only if (iff) X and Y are independent.
- **Independence bound for entropy:** $H(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i)$ with equality iff X_i, $i = 1, 2, \ldots, n$ are mutually independent.

Similar properties hold for differential entropy.

- Mutual Information $I(X; Y)$ is a **non-negative** function of $p(x, y)$, concave in $p(x)$ for fixed $p(y|x)$, and convex in $p(y|x)$ for fixed $p(x)$.
- **Chain rule:** $I(X, Y; Z) = I(X; Z) + I(Y; Z|X)$

Markov Chain

Definition Markov Chain

RVs X and Y are conditionally independent given Z, write $X - Z - Y$, if

$$P_{X|YZ}(x|y,z) = P_{X|Z}(x|z) \text{ whenever } P_{YZ}(y,z) > 0.$$

Useful properties:

1. *Symmetry:* $X - Z - Y \Rightarrow Y - Z - X$
2. *Decomposition:* $X - Z - (Y, W) \Rightarrow X - Z - Y$
3. *Weak union:* $X - Z - (Y, W) \Rightarrow X - (Z, W) - Y$
4. *Contraction:* $X - Z - Y$ and $X - (Z, Y) - W \Rightarrow X - Z - (Y, W)$
5. *Intersection:* If $P_{W,X,Y,Z}(w,x,y,z) > 0$ for all w, x, y, z and $X - (Z, Y) - W$ and $X - (Z, W) - Y \Rightarrow X - Z - (Y, W)$

- **Data Proc. Ineq.:** If $X - Z - Y$, then $I(X;Z) \geq I(X;Y)$.

Introduction Typical Sequences

4 sequences of length 18:

$$
\begin{array}{ll}
(a) & 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0 \\
(b) & 1,0,1,1,0,1,0,1,1,1,0,0,0,0,1,0,1,0 \\
(c) & 0,0,0,1,1,0,0,1,0,0,1,1,0,0,0,1,1,0 \\
(d) & 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1.
\end{array}
$$

One sequence was generated by a random number generator, binary RV X with $P_X(0) = 2/3$ and $P_X(1) = 1/3$.
Probability of each sequence:

$$
(a)\ (2/3)^{18} \quad (b)\ (2/3)^9 \cdot (1/3)^9 \quad (c)\ (2/3)^{11} \cdot (1/3)^7 \quad (d)\ (1/3)^{18}
$$

Question: Which sequence would you intuitively guess? Why?

Strongly Typical Sequences

$N(a;x^n)$ denotes the number of occurrences of $a \in \mathcal{X}$ in $x^n \in \mathcal{X}^n$.

- Ex.: For $x^n = (0,0,1,0,1)$ we have $N(0;x^n) = 3$ and $N(1;x^n) = 2$.
- $\pi(a|x^n) := \frac{N(a;x^n)}{n}$ denotes the **empirical pmf (or type)** of x^n.
- Ex.: $x^n = (0,0,1,0,1)$ and $y^n = (1,0,0,0,1)$ are of the same type

Strongly Typical Sequences

$x^n \in \mathcal{X}^n$ is *strongly ε-typical* with respect to pmf $P_X(x)$ if $N(a;x^n) = 0$ for $a \notin \mathcal{S}_X$ and

$$\left|\frac{N(a;x^n)}{n} - p_X(a)\right| \leq \frac{\varepsilon}{|\mathcal{X}|} \quad \text{for } a \in \mathcal{S}_X.$$

$\mathcal{T}_\varepsilon^{(n)}(X)$ (or $\mathcal{T}_\varepsilon^{(n)}(P_X)$) denotes the set of strongly ε-typical sequences.

Asymptotic Equipartition Property (AEP)

Theorem: Strong Asymptotic Equipartition Property (AEP)

For ε sufficiently small, $x^n \in \mathcal{T}_\varepsilon^{(n)}(X)$, and X^n iid$\sim P_X$ we have

1. $2^{-n(1+\varepsilon)H(X)} \leq p_{X^n}(x^n) \leq 2^{-n(1-\varepsilon)H(X)}$
2. $(1-\delta_\varepsilon(n))2^{n(1-\varepsilon)H(X)} \leq |\mathcal{T}_\varepsilon^{(n)}(X)| \leq 2^{n(1+\varepsilon)H(X)}$
3. $(1-\delta_\varepsilon(n)) < \mathbb{P}\{X^n \in \mathcal{T}_\varepsilon^{(n)}(X)\} \leq 1$

where $\delta_\varepsilon(n) \to 0$ for fixed $\varepsilon > 0$ as $n \to \infty$.

Joint Strongly Typical Sequences

- Sequences x^n and y^n are **jointly strongly ε-typical** wrt P_{XY} if $N(a,b;x^n,y^n) = 0$ for $(a,b) \notin \mathcal{S}_{XY}$ and

$$\left|\frac{N(a,b;x^n,y^n)}{n} - p_{XY}(x,y)\right| \leq \frac{\varepsilon}{|\mathcal{X}\mathcal{Y}|} \quad \text{for } (a,b) \in \mathcal{S}_{XY}.$$

  - $\mathcal{T}_\varepsilon^{(n)}(X,Y)$ (or $\mathcal{T}_\varepsilon^{(n)}(P_{XY})$) denotes the jointly strongly ε-typical set

Joint Typicality Lemma

For $0 < \varepsilon_1 < \varepsilon_2$ sufficiently small, $x^n \in \mathcal{T}_{\varepsilon_1}^{(n)}(P_X)$, and Y^n iid$\sim P_Y$ with P_X and P_Y marginal distributions of P_{XY}, then we have

$$(1-\delta_{\varepsilon_1,\varepsilon_2}(n))2^{-n(I(X;Y)+2\varepsilon_2 H(Y))} \leq \mathbb{P}\{(x^n,Y^n) \in \mathcal{T}_{\varepsilon_2}^{(n)}(P_{XY})\} \leq 2^{-n(I(X;Y)-2\varepsilon_2 H(Y))}$$

where $\delta_{\varepsilon_1,\varepsilon_2}(n) \to 0$ for fixed $\varepsilon > 0$ as $n \to \infty$.

Conditional Typicality Lemma

- The next lemma and it derivations (Berger's Markov Lemma) cannot be proved for weakly typical sequences.

Conditional Typicality Lemma

Given (X,Y) with pmf P_{XY}. Let $x^n \in \mathcal{T}_{\varepsilon_1}^{(n)}(X)$ and Y^n drawn according to $\prod_{i=1}^{n} P_{Y|X}(y_i|x_i)$. Then for every $\varepsilon > \varepsilon_1$,

$$\mathbb{P}\{(x^n, Y^n) \in \mathcal{T}_{\varepsilon}^{(n)}(X,Y)\} \to 1 \quad \text{as } n \to \infty.$$

- **Markov Lemma:** Given RVs (X,Y,Z) with $X - Y - Z$. Let $(x^n, y^n) \in \mathcal{T}_{\varepsilon_2}^{(n)}(X,Y)$. If $Z^n \sim \prod_{i=1}^{n} p_{Z|Y}(z_i|y_i)$, then for $\varepsilon_1 > \varepsilon_2$

$$\mathbb{P}\{(x^n, y^n, Z^n) \in \mathcal{T}_{\varepsilon_1}^{(n)}(X,Y,Z)\} \to 1 \quad \text{as } n \to \infty.$$

Weakly Typical Sequences

- **AEP:** Let X_i iid$\sim P_X$, then the weak law of large numbers gives

$$-\frac{1}{n}\log P(X_1, X_2, \ldots, X_n) \to H(X) \quad \text{in probability}$$

Definition: Weakly Typical Sequences

For $\varepsilon > 0$ sequence $x^n \in \mathcal{X}^n$ is *weakly ε-typical* wrt P_X if

$$\left| -\frac{1}{n}\log P_{X^n}(x^n) - H(X) \right| \leq \varepsilon. \tag{1}$$

$\mathcal{A}_\varepsilon^{(n)}(X)$ (or $\mathcal{A}_\varepsilon^{(n)}(P_X)$) denotes the set of weakly ε-typical sequences.

- Empirical entropies should be ε-close to the true entropies

Properties

- Equivalently to (1): $2^{-n(H(X)+\varepsilon)} \leq P_{X^n}(x^n) \leq 2^{-n(H(X)-\varepsilon)}$

Properties of weakly typical sequences

1. $\mathbb{P}\{(X_1, X_2, \ldots, X_n) \in \mathcal{A}_\varepsilon^{(n)}(X)\} > 1 - \varepsilon$ for n sufficiently large
2. $|\mathcal{A}_\varepsilon^{(n)}(X)| \leq 2^{n(H(X)+\varepsilon)}$
3. $|\mathcal{A}_\varepsilon^{(n)}(X)| \geq (1 - \varepsilon)2^{n(H(X)-\varepsilon)}$ for n sufficiently large

Jointly Weakly Typical Sequences

- (x^n, y^n) are **jointly weakly typical sequences** wrt joint pmf P_{XY} with marginal pmfs P_X and P_Y if $x^n \in \mathcal{A}_\varepsilon^{(n)}(P_X)$, $y^n \in \mathcal{A}_\varepsilon^{(n)}(P_Y)$, and

$$\left| -\frac{1}{n} \log P_{X^n Y^n}(x^n, y^n) - H(X.Y) \right| \leq \varepsilon.$$

- $\mathcal{A}_\varepsilon^{(n)}(P_{XY})$ denotes the set of jointly weakly typical sequences.

Properties

1. $\mathbb{P}\{(X^n, Y^n) \in \mathcal{A}_\varepsilon^{(n)}(P_{XY})\} \to 1$ as $n \to \infty$
2. $|\mathcal{A}_\varepsilon^{(n)}(P_{XY})| \leq 2^{n(H(X,Y)+\varepsilon)}$
3. Let $\tilde{X}^n$ and $\tilde{Y}^n$ be iid according to marginal P_X and P_Y of joint pmf P_{XY}, then for sufficiently large n we have

$$(1-\varepsilon)2^{-n(I(X;Y)+3\varepsilon)} \leq \mathbb{P}\{(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{A}_\varepsilon^{(n)}(P_{XY})\} \leq 2^{-n(I(X;Y)-3\varepsilon)}$$

Resource and Reading Assignment

- **Resources:**
  - Appendix A of *Information Theoretic Security*, by Y. Liang, V. Poor, S. Shamai, NOW Foundations and Trends.
  - *Elements of Information Theory*, by T. Cover, J. Thomas, Wiley.
- **Further reading:**
  - *Network Information Theory*, by A. El Gamal, Y.-H. Kim, Cambridge.
  - *Information Theory and Network Coding*, by R. Yeung, Springer.
  - *Topics in Multi-User Inforamtion Theory* by G. Kramer, NOW Foundations and Trends.
- **Reading Assignment:**
  - chapter 7 of Bloch's book (system aspects), if you have...