Non-asymptotic Equipartition Properties With Respect to Information Quantities

En-Hui Yang [1] Jin Meng [1]

[1]University of Waterloo, Canada

November 26, 2011

Outline

Outline

Asymptotic Equipartition Properties (AEP)

AEP w.r.t Entropy

For independent and identically distributed (IID) source $X = \{X_i\}_{i=1}^{\infty}$,

$$-\frac{1}{n}\ln p(X_1X_2\cdots X_n) \to H(X) \tag{1}$$

Typicality

With high probability, the outcomes of $X_1X_2\cdots X_n$ are approximately equiprobable with their respective probability

from $e^{-n(H(X)+\epsilon)}$ to $e^{-n(H(X)-\epsilon)}$

for small fixed number ϵ. Those sequences are called typical sequences, and the set of typical sequences are called the typical set, denoted by $\mathcal{T}^{(n)}$.

Application of AEP to Source Coding

Simple Proof on Source Coding using AEP

Implication of AEP suggests that only typical sequence need to be represented, while error probability (the probability that a sequence appears with no representation) is arbitrarily small. Therefore, the rate of source code is related to $|\mathcal{T}^{(n)}|$, which can be easily bounded by the following argument.

$$
\begin{aligned}
|\mathcal{T}^{(n)}| e^{-n(H(X)+\epsilon)} &\leq \sum_{x^n \in \mathcal{T}^{(n)}} p(x^n) \leq 1 \\
\Rightarrow \quad R_n = \frac{1}{n} \ln |\mathcal{T}^{(n)}| &\leq H(X) + \epsilon
\end{aligned}
$$

Asymptotic Equipartition Properties (AEP)

AEP w.r.t Joint Entropy

For independent and identically distributed (IID) source pair $(X, Y) = \{(X_i, Y_i)\}_{i=1}^{\infty}$,

$$-\frac{1}{n}\ln p(X^n, Y^n) \to H(X, Y) \tag{2}$$

Joint Typicality

With high probability,

$$e^{-n(H(X)+\epsilon)} \leq p(X^n) \leq e^{-n(H(X)-\epsilon)}$$
$$e^{-n(H(Y)+\epsilon)} \leq p(Y^n) \leq e^{-n(H(Y)-\epsilon)}$$
$$e^{-n(H(X,Y)+\epsilon)} \leq p(X^n, Y^n) \leq e^{-n(H(X,Y)-\epsilon)}$$

Call the set of those sequences as joint typical set $\mathcal{T}_J^{(n)}$.

Application of AEP to Channel Coding

Simple Proof on Channel Coding using AEP

It can be shown that while $(X^n, Y^n) \in \mathcal{T}_J^{(n)}$ with high probability,

$$\Pr\left\{(\tilde{X}^n, Y^n) \in \mathcal{T}_J^{(n)}\right\} \leq e^{-n(I(X;Y)-3\epsilon)}$$

if $\tilde{X}$ and Y are independently. Then as long as channel code rate

$$R \leq I(X;Y) - 3\epsilon$$

$\mathcal{T}_J^{(n)}$ contains only one codeword X^n.

Impact and Limitation of AEP

Impact

Establish asymptotic coding theorem in

- source coding;
- channel coding ;
- and multi-user information theory.

Limitation

- AEP applies only to large block length n;
- AEP can yield only first order results; and
- more importantly, AEP is not applicable to the non-asymptotic regime, which is the real game in practice.

Impact and Limitation of AEP

Impact

Establish asymptotic coding theorem in

- source coding;
- channel coding ;
- and multi-user information theory.

Limitation - Solution: Non-asymptotic Equipartition Property (NEP)!

- AEP applies only to large block length n;
- AEP can yield only first order results; and
- more importantly, AEP is not applicable to the non-asymptotic regime, which is the real game in practice.

Outline

Entropy NEP

Random variable: $-\frac{1}{n}\ln p(X^n)$

Distribution: $?$

Entropy NEP $\approx$ characterization of $-\frac{1}{n}\ln p(X^n)$ for any n

Weak Right Entropy NEP

Chernoff Bound Result

For independent and identically distributed (IID) source $X = \{X_i\}_{i=1}^{\infty}$,

$$\Pr\left\{-\frac{1}{n}\ln p(X^n) \geq H(X) + \delta\right\} \leq e^{-nr_X(\delta)} \tag{3}$$

where

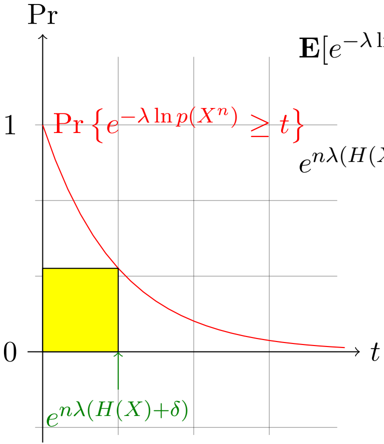
$$r_X(\delta) \triangleq \sup_{\lambda \geq 0}\left[\lambda(H(X) + \delta) - \ln\int p^{-\lambda+1}(x)dx\right].$$

Proof of Weak Right Entropy NEP

Proof

$$\begin{aligned}
& \Pr\left\{-\frac{1}{n}\ln p(X_1X_2\cdots X_n) \geq H(X)+\delta\right\} \\
= \;& \Pr\left\{-\ln p(X_1X_2\cdots X_n) \geq n(H(X)+\delta)\right\} \\
\leq \;& \inf_{\lambda\geq 0} \frac{\mathbf{E}[e^{-\lambda\ln p(X_1X_2\cdots X_n)}]}{e^{n\lambda(H(X)+\delta)}} \\
= \;& \inf_{\lambda\geq 0} e^{-n\left[\lambda(H(X)+\delta)-\ln \mathbf{E}[p^{-\lambda}(X_1)]\right]} \\
= \;& \inf_{\lambda\geq 0} e^{-n\left[\lambda(H(X)+\delta)-\ln\int p^{-\lambda+1}(x)dx\right]} \\
= \;& e^{-nr_X(\delta)}\,. \qquad (4)
\end{aligned}$$

Graphical Interpretation



$$\mathbf{E}[e^{-\lambda \ln p(X^n)}] = \int \Pr\left\{e^{-\lambda \ln p(X^n)} \geq t\right\} dt$$

$$\geq$$

$$e^{n\lambda(H(X)+\delta)} \Pr\left\{-\ln p(X^n) \geq n(H(X)+\delta)\right\}$$

Example of $r_X(\delta)$

For I.I.D binary source with $\Pr\{X=0\} = p \leq 0.5$,

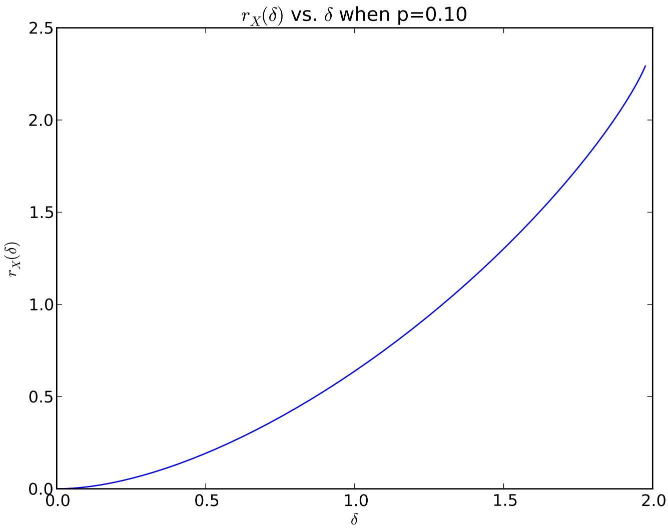
$$r_{X|Y}(\delta) = D\left(p + \frac{\delta}{\ln\frac{1-p}{p}} \middle\| p\right) \tag{5}$$

where

$$D(q||p) \triangleq (1-q)\ln\frac{1-q}{1-p} + q\ln\frac{q}{p}$$

- $D\left(p + \frac{\delta}{\ln\frac{1-p}{p}} \middle\| p\right)$ is convex and non-decreasing.
- $$D\left(p + \frac{\delta}{\ln\frac{1-p}{p}} \middle\| p\right) = \frac{1}{2p(1-p)\ln^2\frac{1-p}{p}}\delta^2 + O(\delta^3)$$

Plot of $r_X(\delta)$



Properties of $r_X(\delta)$

- $r_X(\delta)$ is convex and non-decreasing.
- Parametric Form

$$\delta(\lambda) = \int \frac{p^{-\lambda+1}(x)}{\left[\int p^{-\lambda+1}(y)dy\right]}\left[-\ln p(x)\right] dx - H(X) \qquad (6)$$

$$r_X(\delta(\lambda)) = \lambda(H(X) + \delta(\lambda)) - \ln \int p^{-\lambda+1}(x)dx \; . \qquad (7)$$

- $r'_X(\delta) = \lambda$, $r''_X(\delta) = \frac{1}{\delta'(\lambda)}$.
- $$r_X(\delta) = \frac{1}{2\sigma_H^2(X)}\delta^2 + O(\delta^3) \qquad (8)$$

Strong Right Entropy NEP

For independent and identically distributed (IID) source $X = \{X_i\}_{i=1}^{\infty}$,

$$\Pr\left\{-\frac{1}{n}\ln p(X^n) \geq H(X) + \delta\right\} \leq \frac{1}{1-e^{-\lambda}}\left[\frac{1}{\sqrt{2\pi}\sigma_H(X,\lambda)} + \frac{2M_H(X,\lambda)}{\sigma_H^3(X,\lambda)}\right] e^{-nr_X(\delta)-\frac{1}{2}\ln n} \quad (9)$$

$$\Pr\left\{-\frac{1}{n}\ln p(X^n) \geq H(X) + \delta\right\} \geq e^{-\lambda d}\left[\frac{de^{-\frac{d^2}{2n\sigma_H^2(X,\lambda)}}}{\sqrt{2\pi}\sigma_H(X,\lambda)} - \frac{2M_H(X,\lambda)}{\sigma_H^3(X,\lambda)}\right] e^{-nr_X(\delta)-\frac{1}{2}\ln n} \quad (10)$$

for any $d > 0$, where $\lambda = r'_X(\delta) > 0$.

Derivation of Strong Entropy NEP

Define

$$f_\lambda(x) \triangleq \frac{p^{-\lambda}(x)}{\int p^{-\lambda+1}(y)dy} \tag{11}$$

and

$$B_k \triangleq \left\{ x^n : H(X) + \delta + \frac{k}{n} \leq -\frac{1}{n} \ln p(x^n) < H(X) + \delta + \frac{k+1}{n} \right\}.$$

Then

$$\begin{aligned} &\Pr\left\{ -\frac{1}{n} \ln p(X^n) \geq H(X) + \delta \right\} \\ &= \int_{-\frac{1}{n} \ln p(x^n) \geq H(X)+\delta} p(x^n) dx^n \\ &= \sum_{k=0}^{\infty} \int_{x^n \in B_k} p(x^n) dx^n \end{aligned}$$

Derivation of Strong Entropy NEP

$$
\begin{aligned}
&= \sum_{k=0}^{\infty} \int_{x^n \in B_k} f_\lambda^{-1}(x^n) f_\lambda(x^n) p(x^n) dx^n \\
&= \sum_{k=0}^{\infty} \int_{x^n \in B_k} e^{\left\{-n\left[-\frac{1}{n}\lambda \ln p(x^n) - \ln \int p^{-\lambda+1}(y)dy\right]\right\}} f_\lambda(x^n) p(x^n) dx^n \\
&\leq \sum_{k=0}^{\infty} e^{\left\{-n\left[\lambda(H(X)+\delta+\frac{k}{n}) - \ln \int p^{-\lambda+1}(y)dy\right]\right\}} \int_{x^n \in B_k} f_\lambda(x^n) p(x^n) dx^n \\
&= e^{-nr_X(\delta)} \sum_{k=0}^{\infty} e^{-\lambda k} \int_{x^n \in B_k} f_\lambda(x^n) p(x^n) dx^n \qquad (12)
\end{aligned}
$$

Central Limit Theorem

Lemma (Berry and Esseen)

Let $V_1, V_2, \cdots$ be independent real random variables with zero means and finite third moments, and set

$$\sigma_n^2 = \sum_{i=1}^{n} \mathbf{E} V_i^2.$$

Then there exists a universal constant $C < 1$ such that for any $n \geq 1$,

$$\sup_{-\infty < t < +\infty} \left| \Pr\{\sum_{i=1}^{n} V_i \leq \sigma_n t\} - \Phi(t) \right| \leq C \sigma_n^{-3} \sum_{i=1}^{n} \mathbf{E}|V_i|^3,$$

where $\Phi(t) = (2\pi)^{-1/2} \int_{-\infty}^{t} e^{-u^2/2} du$.

Derivation of Strong Entropy NEP - Cont.

Consider $Z_1, Z_2, \cdots, Z_n$ with pmf or pdf $f_\lambda(z)p(z)$, and applying central limit theorem to the IID sequence

$$\{-\ln p(Z_i) - (H(X) + \delta)\}_{i=1}^n$$

yields

$$\begin{aligned}
&\int_{x^n \in B_k} f_\lambda(x^n)p(x^n)dx^n \\
&\leq \frac{1}{\sqrt{2\pi}} \int_0^{\frac{1}{\sqrt{n}\sigma_H(X,\lambda)}} e^{-\frac{t^2}{2}} dt + 2C\frac{1}{\sqrt{n}} \frac{M_H(X,\lambda)}{\sigma_H^3(X,\lambda)} \\
&\leq \frac{1}{\sqrt{n}} \left(\frac{1}{\sqrt{2\pi}\sigma_H(X,\lambda)} + \frac{2M_H(X,\lambda)}{\sigma_H^3(X,\lambda)} \right) \qquad (13)
\end{aligned}$$

for any $k \geq 0$.

Derivation of Strong Entropy NEP - Cont.

Combining (13) with (12) yields

$$\begin{aligned}
&\Pr\left\{-\frac{1}{n}\ln p(X^n) \geq H(X) + \delta\right\} \\
&\leq \quad e^{-nr_X(\delta)-\frac{1}{2}\ln n}\left(\frac{1}{\sqrt{2\pi}\sigma_H(X,\lambda)} + \frac{2M_H(X,\lambda)}{\sigma_H^3(X,\lambda)}\right)\sum_{k=0}^{\infty} e^{-\lambda k} \\
&= \quad \frac{1}{1-e^{-\lambda}}\left(\frac{1}{\sqrt{2\pi}\sigma_H(X,\lambda)} + \frac{2M_H(X,\lambda)}{\sigma_H^3(X,\lambda)}\right) e^{-nr_X(\delta)-\frac{1}{2}\ln n} \,.
\end{aligned}$$

This completes the proof of (9).

Derivation of Strong Entropy NEP - Cont.

To prove (10), note that for any $d > 0$

$$
\begin{aligned}
\Pr\left\{-\frac{1}{n}\ln p(X^n) \geq H(X) + \delta\right\}
&\geq \int_{0 \leq -\frac{1}{n}\ln p(x^n) - (H(X)+\delta) < \frac{d}{n}} p(x^n)dx^n \\
&= \int_{0 \leq -\frac{1}{n}\ln p(x^n) - (H(X)+\delta) < \frac{d}{n}} f_\lambda^{-1}(x^n) f_\lambda(x^n) p(x^n) dx^n \\
&\geq e^{-nr_X(\delta) - \lambda d} \int_{0 \leq -\frac{1}{n}\ln p(x^n) - (H(X)+\delta) < \frac{d}{n}} f_\lambda(x^n) p(x^n) dx^n \quad (14)
\end{aligned}
$$

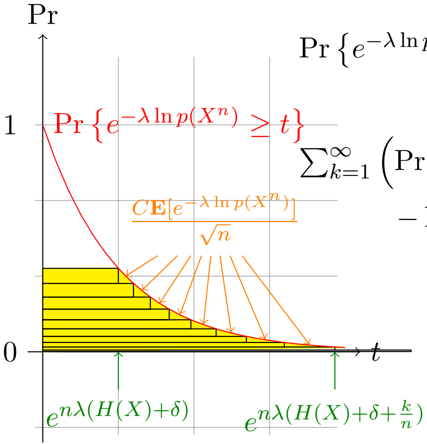
Derivation of Strong Entropy NEP - Cont.

Applying Lemma 1 to the IID sequence $\{-\ln p(Z_i) - (H(X)+\delta)\}_{i=1}^n$ again, we have

$$\begin{aligned}
&\int_{0 \le -\frac{1}{n}\ln p(x^n) - (H(X)+\delta) < \frac{d}{n}} f_\lambda(x^n) p(x^n) dx^n \\
&\ge \frac{1}{\sqrt{2\pi}} \int_0^{\frac{d}{\sqrt{n}\sigma_H(X,\lambda)}} e^{-\frac{t^2}{2}} dt - 2C \frac{1}{\sqrt{n}} \frac{M_H(X,\lambda)}{\sigma_H^3(X,\lambda)} \\
&\ge \frac{1}{\sqrt{n}} \left(\frac{d}{\sqrt{2\pi}\sigma_H(X,\lambda)} e^{-\frac{d^2}{2n\sigma_H^2(X,\lambda)}} - \frac{2M_H(X,\lambda)}{\sigma_H^3(X,\lambda)} \right) \qquad (15)
\end{aligned}$$

which, combined with (14), implies (10).

Graphical Interpretation



$$\Pr\left\{e^{-\lambda \ln p(X^n)} \geq e^{n\lambda(H(X)+\delta)}\right\}$$
$$=$$
$$\sum_{k=1}^{\infty}\left(\Pr\left\{e^{-\lambda \ln p(X^n)} \geq e^{n\lambda(H(X)+\delta+\frac{k-1}{n})}\right\} - \Pr\left\{e^{-\lambda \ln p(X^n)} \geq e^{n\lambda(H(X)+\delta+\frac{k}{n})}\right\}\right)$$

Central Limit Theorem

For any $\delta \le c\sqrt{\frac{\ln n}{n}}$, where $c < \sigma_H(X)$ is a constant,

$$\begin{aligned} Q\left(\frac{\delta\sqrt{n}}{\sigma_H(X)}\right) - \frac{CM_H(X)}{\sqrt{n}\sigma_H^3(X)} &\le \Pr\left\{-\frac{1}{n}\ln p(X^n) \ge H(X)+\delta\right\} \\ &\le Q\left(\frac{\delta\sqrt{n}}{\sigma_H(X)}\right) + \frac{CM_H(X)}{\sqrt{n}\sigma_H^3(X)} \quad (16) \end{aligned}$$

where $Q(t) = \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-u^2/2}du$, and $C < 1$ is the universal constant in the central limit theorem of Berry and Esseen.

Application to Fixed Rate Source Coding

Given a memoryless source X, the performance of optimal fixed rate source coding for any block length n is characterized by,

$$\delta \geq R_n - H(X) \geq \delta - r_X(\delta) - \frac{\ln n}{2n} - O(n^{-1}) \qquad (17)$$

whenever

$$\left| \frac{\ln \epsilon_n}{n} + r_X(\delta) + \frac{\ln n}{2n} + \frac{\ln \lambda}{n} \right| \leq O(n^{-1}) \qquad (18)$$

for $\Omega\left(\frac{1}{\sqrt{n}}\right) = \delta \leq \ln|\mathcal{X}| - H(X)$ and $\lambda = r'_X(\delta)$.

Application to Fixed Rate Source Coding

(a) Let δ be a constant with respect to n. Then

$$\begin{aligned}
& r_X^{(inv)}\left(-\frac{\ln \epsilon_n}{n} - \frac{\ln n}{2n}\right) + O(n^{-1}) \\
\geq \; & R_n - H(X) \\
\geq \; & r_X^{(inv)}\left(-\frac{\ln \epsilon_n}{n} - \frac{\ln n}{2n}\right) + \frac{\ln \epsilon_n}{n} - O(n^{-1})
\end{aligned} \tag{19}$$

whenever ϵ_n decreases exponentially with respect to n, where $r_X^{(inv)}$ is the inverse function of r_X.

Application to Fixed Rate Source Coding

(b) Let $\delta = \sigma_H(X)\sqrt{\frac{2\alpha \ln n}{n}}$ for some $\alpha > 0$. Then

$$\begin{aligned} & \sigma_H(X)\sqrt{\frac{2\alpha \ln n}{n}} \\ \geq \quad & R_n - H(X) \\ \geq \quad & \sigma_H(X)\sqrt{\frac{2\alpha \ln n}{n}} - \left(\frac{1}{2} + \alpha\right)\frac{\ln n}{n} - O(n^{-1}) \end{aligned} \quad (20)$$

whenever

$$\epsilon_n = \Theta\left(\frac{n^{-\alpha}}{\sqrt{\ln n}}\right). \quad (21)$$

Application to Fixed Rate Source Coding

(c) Let $\delta = \frac{c}{\sqrt{n}}$ for a constant c. Then

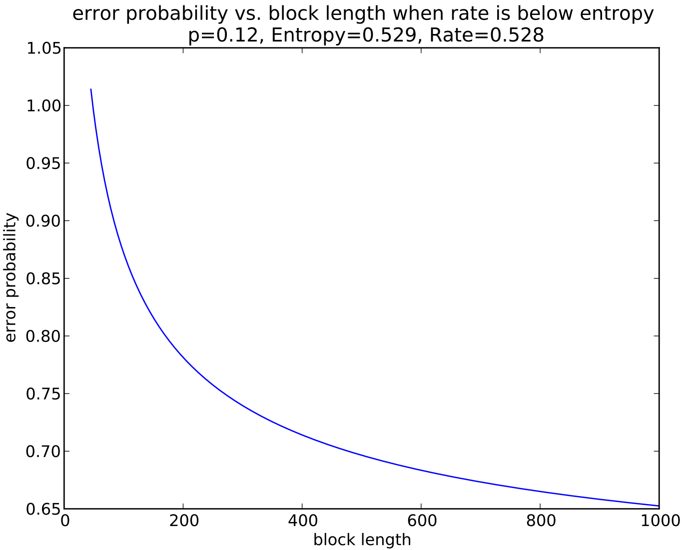
$$\frac{c}{\sqrt{n}} \geq R_n - H(X) \geq \frac{c}{\sqrt{n}} - \frac{\ln n}{2n} - O(n^{-1}) \qquad (22)$$

whenever

$$\left| \epsilon_n - Q\left(\frac{c}{\sigma_H(X)}\right) \right| \leq \frac{CM_H(X)}{\sqrt{n}\sigma_H^3(X)} \qquad (23)$$

where $Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-u^2/2} du$, and $C < 1$ is the universal constant in the central limit theorem of Berry and Esseen.

Working below Entropy!



Outline

Right Strong Conditional Entropy NEP

For any $\delta \in (0, \Delta^*(X|Y))$ and any positive integer n

$$\Pr\left\{-\frac{1}{n}\ln p(X^n|Y^n) \geq H(X|Y)+\delta\right\} \leq \frac{1}{1-e^{-\lambda}}\left[\frac{1}{\sqrt{2\pi}\sigma_H(X|Y,\lambda)} + \frac{2M_H(X|Y,\lambda)}{\sigma_H^3(X|Y,\lambda)}\right] e^{-nr_{X|Y}(\delta)-\frac{1}{2}\ln n}$$

and

$$\Pr\left\{-\frac{1}{n}\ln p(X^n|Y^n) \geq H(X|Y)+\delta\right\} \geq e^{-\lambda d}\left[\frac{de^{-\frac{d^2}{2n\sigma_H^2(X|Y,\lambda)}}}{\sqrt{2\pi}\sigma_H(X|Y,\lambda)} - \frac{2M_H(X|Y,\lambda)}{\sigma_H^3(X|Y,\lambda)}\right] e^{-nr_{X|Y}(\delta)-\frac{1}{2}\ln n}$$

for any $d > 0$, where $\lambda = r'_{X|Y}(\delta) > 0$.

Central Limit Theorem

For any $\delta \leq c\sqrt{\frac{\ln n}{n}}$, where $c < \sigma_H(X|Y)$ is a constant,

$$\begin{aligned} & Q\left(\frac{\delta\sqrt{n}}{\sigma_H(X|Y)}\right) - \frac{CM_H(X|Y)}{\sqrt{n}\sigma_H^3(X|Y)} \\ \leq \quad & \Pr\left\{-\frac{1}{n}\ln p(X^n|Y^n) \geq H(X|Y) + \delta\right\} \\ \leq \quad & Q\left(\frac{\delta\sqrt{n}}{\sigma_H(X|Y)}\right) + \frac{CM_H(X|Y)}{\sqrt{n}\sigma_H^3(X|Y)} \end{aligned} \tag{24}$$

where $Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-u^2/2} du$, and $C < 1$ is the universal constant in the central limit theorem of Berry and Esseen.

Non-Asymptotic Channel Coding Theorem

Given a BIMC with $C_{\mathrm{BIMC}} \in (0,1)$, let $P_e(\mathcal{C}_{n,k}^{(i)})$, $i = 1, 2$, denote the average word error probability of $\mathcal{C}_{n,k}^{(i)}$ with respect to the random message q, the BIMC, and the random linear code $\mathcal{C}_{n,k}^{(i)}$ itself.

- BIMC: Binary Input Memoryless Channel with Uniform Capacity-Achieving Input Distribution.
- C_{BIMC}: Channel Capacity.
- $\mathcal{C}_{n,k}^{(1)}$: Elias' Generator Ensemble.
- $\mathcal{C}_{n,k}^{(2)}$: Gallager's Parity Check Ensemble.

Non-Asymptotic Channel Coding Theorem

(a) For any $\delta \in (0, \Delta^*(X|Y))$

$$P_e(\mathcal{C}_{n,k}^{(i)}) \leq 2C^{(i)}\Psi(X|Y,\lambda)e^{-nr_{X|Y}(\delta)-\frac{1}{2}\ln n} \tag{25}$$

whenever

$$\mathcal{R}(\mathcal{C}_{n,k}) \leq C_{\mathrm{BIMC}} - \delta - r_{X|Y}(\delta) - \frac{\frac{1}{2}\ln n - \ln C^{(i)}\Psi(X|Y,\lambda)}{n} \tag{26}$$

where $\lambda = r'_{X|Y}(\delta)$ and

$$C^{(i)} = \begin{cases} 1 & \text{if } i = 1 \\ \frac{1}{1-2^{-n}} & \text{otherwise.} \end{cases} \tag{27}$$

Non-Asymptotic Channel Coding Theorem

(b) For any $\alpha \geq 0.5$

$$P_e(\mathcal{C}_{n,k}^{(i)}) \leq \frac{2C^{(i)}\sigma_{\mathrm{H}}(X|Y)\Psi(X|Y)}{\sqrt{2\alpha \ln n}} n^{-\alpha} + O\left(n^{-\alpha}\frac{\ln n}{\sqrt{n}}\right) \tag{28}$$

whenever

$$\mathcal{R}(\mathcal{C}_{n,k}) \leq C_{\mathrm{BIMC}} - \sigma_{\mathrm{H}}(X|Y)\sqrt{\frac{2\alpha \ln n}{n}} - \frac{\alpha \ln n}{n} - O\left(\frac{\ln \ln n}{n}\right). \tag{29}$$

Non-Asymptotic Channel Coding Theorem

(c) For any real number c

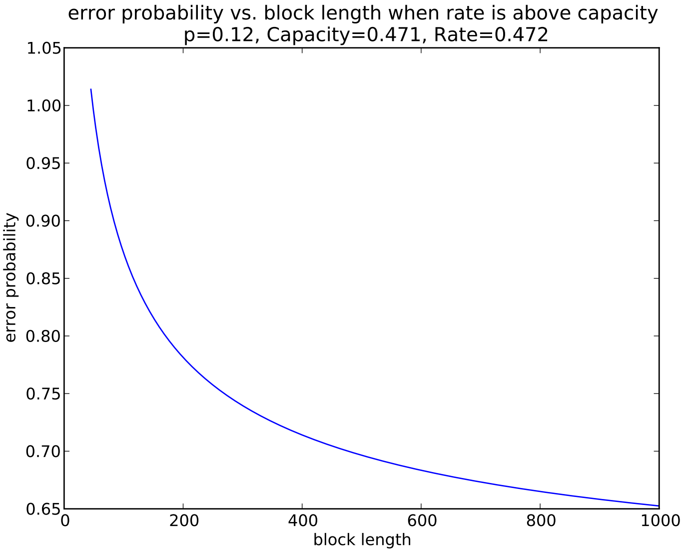
$$P_e(\mathcal{C}_{n,k}^{(i)}) \leq C^{(i)} \left(Q \left(\frac{c}{\sigma_{\mathrm{H}}(X|Y)} \right) + \frac{M_{\mathrm{H}}(X|Y)}{\sigma_{\mathrm{H}}^3(X|Y)} \frac{1}{\sqrt{n}} \right) \quad (30)$$

whenever

$$\mathcal{R}(\mathcal{C}_{n,k}) \leq C_{\mathrm{BIMC}} - \frac{c}{\sqrt{n}} - \frac{\ln n}{2n} + \frac{1}{n} \ln \frac{C^{(i)}(1 - C_{BE}) M_{\mathrm{H}}(X|Y)}{\sigma_{\mathrm{H}}^3(X|Y)} \quad (31)$$

where $0 < C_{BE} < 0.4784$ is the universal constant in the Berry-Esseen central limit theorem.

Working above Capacity!



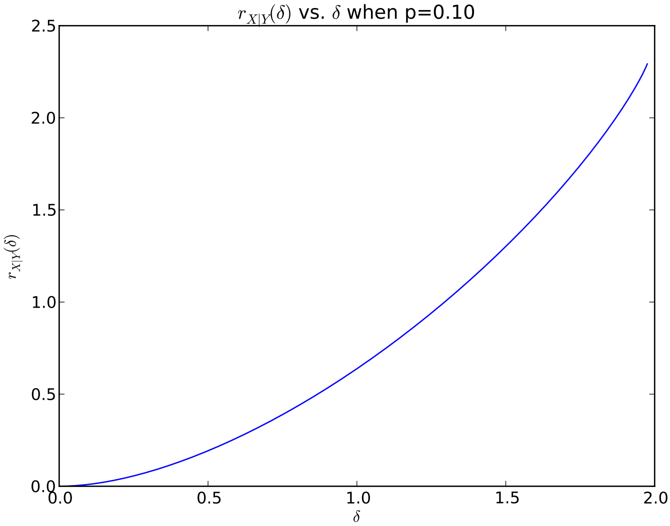
Example of $r_{X|Y}(\delta)$ - Binary Symmetric Channel

Let p be the crossover probability of BSC,

$$r_{X|Y}(\delta) = D\left(p + \frac{\delta}{\ln \frac{1-p}{p}} \middle\| p\right) \tag{32}$$

$$\sigma_H^2(X|Y) = p(1-p)\ln^2 \frac{1-p}{p} \tag{33}$$

Plot of $r_{X|Y}(\delta)$



Example of $r_{X|Y}(\delta)$ - Binary Input Gaussian Channel

Assume that input of channel is modulated to $\{+1,-1\}$, and therefore

$$p(y|x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|y-x|^2}{2\sigma^2}} \tag{34}$$

for $x = \{+1,-1\}$, where σ is the variance of the noise.

Example of $r_{X|Y}(\delta)$ - Binary Input Gaussian Channel

Let U be gaussian random variable with mean 0 and variance 1.

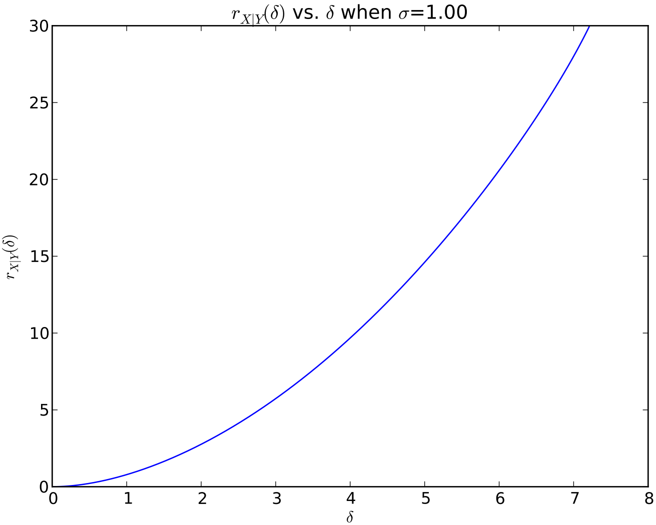
$$\delta(\lambda) = \frac{\mathbf{E}\left[g^{\lambda}\left(\frac{\sigma U+1}{\sigma^2}\right)\ln g\left(\frac{\sigma U+1}{\sigma^2}\right)\right]}{\mathbf{E}\left[g^{\lambda}\left(\frac{\sigma U+1}{\sigma^2}\right)\right]} - \mathbf{E}\left[\ln g\left(\frac{\sigma U+1}{\sigma^2}\right)\right]$$

$$r_{X|Y}(\delta(\lambda)) = \lambda\frac{\mathbf{E}\left[g^{\lambda}\left(\frac{\sigma U+1}{\sigma^2}\right)\ln g\left(\frac{\sigma U+1}{\sigma^2}\right)\right]}{\mathbf{E}\left[g^{\lambda}\left(\frac{\sigma U+1}{\sigma^2}\right)\right]} - \ln\left\{\mathbf{E}\left[g^{\lambda}\left(\frac{\sigma U+1}{\sigma^2}\right)\right]\right\}$$

and

$$\sigma_H^2(X|Y) = \mathbf{E}\left[\ln^2 g\left(\frac{\sigma U+1}{\sigma^2}\right)\right] - \left\{\mathbf{E}\left[-\ln g\left(\frac{\sigma U+1}{\sigma^2}\right)\right]\right\}^2$$

Plot of $r_{X|Y}(\delta)$



Outline

Left NEP on Mutual Information

For any $\delta \in (0, \Delta^*_-(X;Y))$ and any positive integer n

$$\Pr\left\{\frac{1}{n}\ln\frac{p(Y^n|X^n)}{p(Y^n)} \leq I(X;Y)-\delta\right\}$$

$$\leq \quad \frac{1}{1-e^{-\lambda}}\left[\frac{1}{\sqrt{2\pi}\sigma_{I,-}(X;Y,\lambda)}+\frac{2M_{I,-}(X;Y,\lambda)}{\sigma^3_{I,-}(X;Y,\lambda)}\right]e^{-nr_{X;Y,-}(\delta)-\frac{1}{2}\ln n}$$

and

$$\Pr\left\{\frac{1}{n}\ln\frac{p(Y^n|X^n)}{p(Y^n)} \leq I(X;Y)-\delta\right\}$$

$$\geq \quad e^{-\lambda d}\left[\frac{de^{-\frac{d^2}{2n\sigma^2_{I,-}(X;Y,\lambda)}}}{\sqrt{2\pi}\sigma_{I,-}(X;Y,\lambda)}-\frac{2M_{I,-}(X;Y,\lambda)}{\sigma^3_{I,-}(X;Y,\lambda)}\right]e^{-nr_{X;Y,-}(\delta)-\frac{1}{2}\ln n}$$

for any $d > 0$, where $\lambda = r'_{X;Y,-}(\delta) > 0$.

Central Limit Theorem

For any $\delta \leq c\sqrt{\frac{\ln n}{n}}$, where $c < \sigma_I(X;Y)$ is a constant,

$$\begin{aligned} & Q\left(\frac{\delta\sqrt{n}}{\sigma_I(X;Y)}\right) - \frac{CM_I(X;Y)}{\sqrt{n}\sigma_I^3(X;Y)} \\ \leq\ & \Pr\left\{\frac{1}{n}\ln\frac{p(Y^n|X^n)}{p(Y^n)} \leq I(X;Y) - \delta\right\} \\ \leq\ & Q\left(\frac{\delta\sqrt{n}}{\sigma_I(X;Y)}\right) + \frac{CM_I(X;Y)}{\sqrt{n}\sigma_I^3(X;Y)} \end{aligned} \tag{35}$$

where $Q(t) = \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-u^2/2}du$, and $C < 1$ is the universal constant in the central limit theorem of Berry and Esseen.

Some Definition on Relative Entropy

Let t be the type of x^n, i.e. $nt(a)$ is the number of times the symbol a appears in x^n. Define

$$q_t(y) \triangleq \sum_{x\in\mathcal{X}} t(a)p(y|x)$$

$$I(t;P) \triangleq \sum_{x\in\mathcal{X}} t(x)\int p(y|x)\ln\frac{p(y|x)}{q_t(y)}dy$$

Left Relative Entropy NEP

For any $\delta \in (0, \Delta_-^*(X;Y))$

$$\Pr\left\{ \frac{1}{n} \ln \frac{p(Y^n|X^n)}{q_t(Y^n)} \le I(t;P) - \delta \middle| X^n = x^n \right\}$$

$$\le \frac{1}{1-e^{-\lambda}} \left[\frac{1}{\sqrt{2\pi}\sigma_{D,-}(t;P,\lambda)} + \frac{2M_{D,-}(t;P,\lambda)}{\sigma_{D,-}^3(t;P,\lambda)} \right] e^{-nr_-(t,\delta) - \frac{1}{2}\ln n}$$

and

$$\Pr\left\{ \frac{1}{n} \ln \frac{p(Y^n|X^n)}{q_t(Y^n)} \le I(t;P) - \delta \middle| X^n = x^n \right\}$$

$$\ge e^{-\lambda d} \left[\frac{d e^{-\frac{d^2}{2n\sigma_{D,-}^2(t;P,\lambda)}}}{\sqrt{2\pi}\sigma_{D,-}(t;P,\lambda)} - \frac{2M_{D,-}(t;P,\lambda)}{\sigma_{D,-}^3(t;P,\lambda)} \right] e^{-nr_-(t,\delta) - \frac{1}{2}\ln n}$$

for any $d > 0$, where $\lambda = \frac{\partial r_-(t,\delta)}{\partial \delta} > 0$.

Central Limit Theorem

For any $\delta \leq c\sqrt{\frac{\ln n}{n}}$, where $c < \sigma_D(t;P)$ is a constant,

$$\begin{aligned} & Q\left(\frac{\delta\sqrt{n}}{\sigma_D(t;P)}\right) - \frac{CM_D(t;P)}{\sqrt{n}\sigma_D^3(t;P)} \\ \leq \quad & \Pr\left\{ \frac{1}{n}\ln\frac{p(Y^n|X^n)}{q_t(Y^n)} \leq I(t;P) - \delta \,\middle|\, X^n = x^n \right\} \\ \leq \quad & Q\left(\frac{\delta\sqrt{n}}{\sigma_D(t;P)}\right) + \frac{CM_D(t;P)}{\sqrt{n}\sigma_D^3(t;P)} \end{aligned} \tag{36}$$

where $Q(t) = \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-u^2/2}du$, and $C < 1$ is the universal constant in the central limit theorem of Berry and Esseen.

Non-Asymptotic Channel Coding Theorem

Given a DIMC P with $C_{\mathrm{DIMC}} \in (0, |\mathcal{X}|)$, let $P_e(\mathcal{C}_{t,n,k})$ denote the average word error probability (under jar decoding) of $\mathcal{C}_{t,n,k}$ with respect to the DIMC and the random code $\mathcal{C}_{t,n,k}$ itself.

- DIMC: Discrete Input Memoryless Channel with Any Capacity-Achieving Input Distribution.
- C_{DIMC}: Channel Capacity.
- $\mathcal{C}_{t,n,k}$: Random Code Chosen within Type t.
- There always exists t such that

$$||t - p_X||_1 \leq \frac{|\mathcal{X}|}{n} \tag{37}$$

  where p_X is the optimal input distribution.

Non-Asymptotic Channel Coding Theorem

(a) For any $\delta \in (0, \Delta_-^*(t))$

$$P_e(\mathcal{C}_{t,n,k}) \leq 2\Psi(t; P, \lambda) e^{-nr_-(t,\delta) - \frac{1}{2}\ln n} \quad (38)$$

whenever

$$\mathcal{R}(\mathcal{C}_{t,n,k}) \leq \mathrm{I}(t; P) - \delta - r_-(t,\delta) - \frac{\left(\frac{1}{2} + |\mathcal{X}|\right)\ln(n+1) - \ln\Psi(t; P, \lambda)}{n} \quad (39)$$

where $\lambda = \frac{\partial r_-(t,\delta)}{\partial \delta}$ satisfying $\delta_-(t, \lambda) = \delta$.

Non-Asymptotic Channel Coding Theorem

(b) For any $\alpha \geq 0.5$ and any t satisfying (37)

$$P_e(\mathcal{C}_{t,n,k}) \leq \frac{2\sigma_{\mathrm{D}}(X;Y)\Psi(X;Y)}{\sqrt{2\alpha \ln n}} n^{-\alpha} + O\left(n^{-\alpha}\frac{\ln n}{\sqrt{n}}\right) \quad (40)$$

whenever

$$\mathcal{R}(\mathcal{C}_{t,n,k}) \leq C_{\mathrm{DIMC}} - \sigma_{\mathrm{D}}(X;Y)\sqrt{\frac{2\alpha \ln n}{n}} - \frac{(\alpha + |\mathcal{X}|)\ln(n+1)}{n} - O\left(\frac{\ln \ln n}{n}\right) \quad (41)$$

Non-Asymptotic Channel Coding Theorem

(c) For any t satisfying (37)

$$P_e(\mathcal{C}_{t,n,k}) \le Q\left(\frac{c}{\sigma_{\mathrm{D}}(X;Y)}\right) + \frac{M_{\mathrm{D}}(X;Y)}{\sigma^3_{\mathrm{D}}(X;Y)}\frac{1}{\sqrt{n}} + O\left(n^{-1.5}\right)$$

whenever

$$\begin{aligned}\mathcal{R}(\mathcal{C}_{t,n,k}) \le\ & C_{\mathrm{DIMC}} - \frac{c}{\sqrt{n}} - \left(\frac{1}{2} + |\mathcal{X}|\right)\frac{\ln(n+1)}{n} \\ & - \frac{1}{n}\ln\frac{(1-C_{BE})M_{\mathrm{D}}(X;Y)}{\sigma^3_{\mathrm{D}}(X;Y)} - O\left(n^{-1}\right)\end{aligned}$$

for any real number c, where $0 < C_{BE} < 0.56$ is the universal constant in the Berry-Esseen central limit theorem.