Finite Blocklength Analysis of Channel Capacity

K Gautam Shenoy

ECE Dept., Indian Institute of Science, Bangalore, India

Student Seminar Series

What is Finite Blocklength(FB) analysis?

- Traditional channel capacity assumes codeword blocklengths tending to infinity.
- FB analysis adds a blocklength constraint (n is finite but possibly large).
- Study the backoff from capacity due to FB.

Why FB analysis?

- In reality, we are limited by blocklength.
- Capacity obtained as a function of blocklength is a more useful than channel capacity which is asymptotic.
- Analysis doesn't require explicit code construction!!
- Useful also in source coding, JSCC and even information theoretic secrecy.

Related Terminologies

Under FB analysis:

- If probability of error (p.o.e.) $\to 0$ exponentially and we study rates such that this happens, it's called an error exponent analysis.
- If p.o.e. is fixed and we study the maximum rate achievable, it is a second order analysis.
- If p.o.e. $\to 0$ and rate tends to capacity, this study is moderate deviation asymptotics.

We focus on the second perspective.

Goal

- The exact characterization of FB capacity is unknown even for the most basic channels.
- We settle for tight lower and upper bounds.
- Preferably, lower and upper bound should have matching second order terms.
- If the first order term is independent of ε, you get strong converse for free.

History of FB

- DMCs, second order characterization (*Strassen 1964*).
- AWGN channels, second order characterization (*Hayashi 2009*).
- Refinements of above (*Polyanskiy et.al. 2010, Tomamichel and Tan 2013*).
- Channels with state (*Tomamichel and Tan, 2014*).
- Energy harvesting channels (*Fong and Tan 2015, Shenoy and Sharma 2016*).
- Fading channels (*Yang et.al. 2015*).

Maximal and Average p.o.e.

Let U (equiprobable) be the message to be transmitted, $\hat{U}$ the decoded message.

- Maximal p.o.e: $\max_{1 \leq m \leq M} Pr[\hat{U} \neq m | U = m]$.
- Average p.o.e: $Pr[\hat{U} \neq U]$.
- Slightly different theorems under each criteria.
- Affects higher order terms.

Information Density[1]

$$i_Q(x;y) = \log \frac{W(y|x)}{Q(y)} \tag{1}$$

- If $Q = PW$, then expected value of above is $I(X;Y)$.
- In some sense, FB analysis is a detailed study of this.
- Basically a log likelihood ratio.

[1] T.S. Han, *Information spectrum methods in Information Theory*, Springer 2003.

Notation

- An (n, M, ε) code is a code with M codewords having blocklength n and p.o.e. ε.
- The channel will be represented by $W(y|x)$ or $P_{Y|X}$.

Basic Single Shot Achievability Lemmas

Lemma (Shannon)

For any input distribution P_X, $0 < \varepsilon < 1$ *average p.o.e., there exists a* (M, ε) *code such that for any* $\gamma > 0$,

$$\varepsilon \leq Pr[i(X;Y) \leq \log \gamma] + \frac{M-1}{\gamma} \tag{2}$$

Lemma (Feinstein)

For any input distribution P_X, $0 < \varepsilon < 1$ *maximal p.o.e., there exists a* (M, ε) *code such that for any* $\gamma > 0$,

$$\varepsilon \leq Pr[i(X;Y) \leq \log \gamma] + \frac{M}{\gamma} \tag{3}$$

Random Coding Union bound

Lemma (Polyanskiy)

For any input distribution P_X, $0 < \varepsilon < 1$ *average p.o.e., there exists a* (M, ε) *code such that*

$$\varepsilon \leq \mathbb{E}[1 \wedge (M-1)P[i(\hat{X};Y) \geq i(X;Y)|X,Y]] \tag{4}$$

where $P_{XY\hat{X}}(x,y,u) = P_X(x)W(y|x)P_X(u)$.

- Non-parametric.
- Shannon's lemma can be recovered.

Dependence Testing Bound

Lemma (Polyanskiy)

For any input distribution P_X, $0 < \varepsilon < 1$ average p.o.e., there exists a (M, ε) code such that

$$\varepsilon \leq \mathbb{E}\left[exp\left(-\left[i(X;Y) - \log \frac{M-1}{2} \right]^+ \right) \right] \tag{5}$$

where $(x)^+ = max(x, 0)$.

Hypothesis Testing Methods

Given two distributions P and Q on $\mathcal{X}$, define

$$\beta_\alpha(P, Q) \triangleq \inf \int T(1|x)dQ(x) \tag{6}$$

where the infimum is over all test functions T such that $\int T(1|x)dP(x) \geq \alpha$.

Given distributions $P_i,\ i \in \mathcal{I}$ and Q on $\mathcal{X}$, define

$$\kappa_\tau(\mathcal{I}, Q) \triangleq \inf \int T(1|x)dQ(x) \tag{7}$$

where the infimum is over all test functions T such that $\int T(1|x)dP_i(x) \geq \alpha$ for every $i \in \mathcal{I}$.

Lemma (Polyanskiy)

For any $0 < \varepsilon < 1$, maximal p.o.e., there exists an (M, ε) code with codewords from $\mathbb{F}$ such that

$$M \geq \frac{\kappa_\tau(\mathbb{F}, Q_Y)}{\sup_{x \in \mathbb{F}} \beta_{1-\varepsilon-\tau}(W(.|x), Q_Y)} \tag{8}$$

for any output distribution Q_Y and any $0 < \tau < \varepsilon$.

- *Recovers earlier bounds.*
- *$\kappa_\tau(\mathbb{F}, Q_Y)$ is usually hard to bound.*

Useful properties

- $\beta_\alpha(P,Q) \leq \frac{1}{\gamma}$ for $\gamma > 0$ such that $P[\frac{dP}{dQ} \geq \gamma] \geq \alpha$.
- $\beta_\alpha(P,Q) \geq \frac{1}{\gamma}\left(\alpha - P[\frac{dP}{dQ} \geq \gamma]\right)$ for any $\gamma > 0$.
- $\kappa_\tau \leq \tau$.
- $\kappa_\tau \geq \tau Q_X(\mathbb{F})$ if $Q_Y = Q_X W$.

Berry Esseen Theorem

Theorem

If X_i are i.i.d. random variables with zero mean, variance V and third moment $K < \infty$, then $\forall x \in \mathbb{R}$

$$\left| Pr\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{nV}} \leq x \right) - \Phi(x) \right| \leq \frac{K}{\sqrt{n}V^{3/2}} \tag{9}$$

where Φ is the cdf of standard normal.

- Used to bound the probability terms.
- Finite n version of central limit theorem.
- Standard strategy to get tight second order terms.

Achievability: Results

- For AWGN channels with $0 < \varepsilon < 1$, maximal p.o.e., SNR P, capacity C_G and $V = \frac{P(P+2)}{2(P+1)^2}\log_2^2(e)$,

$$\log M \geq nC_G + \sqrt{nV}\Phi^{-1}(\varepsilon) + O(1) \tag{10}$$

- For DMC with $V_D > 0$,

$$\log M \geq nC_D + \sqrt{nV_D}\Phi^{-1}(\varepsilon) + O(1) \tag{11}$$

Converse: Beyond Fano

$$\frac{\log M}{n} \leq \frac{C + h(\varepsilon)}{1 - \varepsilon} \tag{12}$$

- Usually Fano's inequality is the starting point but...
- Not refined for second order analysis.
- Require stronger bounding techniques.

Basic Converses

Lemma (Han-Verdu)

For $0 < \varepsilon < 1$, average p.o.e., every (M, ε) code satisfies the following for any $\gamma > 0$,

$$\varepsilon \geq \inf_{P_X} Pr[i(X; Y) \leq \log \gamma] - \frac{\gamma}{M} \tag{13}$$

Lemma (Wolfowitz)

For $0 < \varepsilon < 1$, maximal p.o.e., every (M, ε) code satisfies the following for any $\gamma > 0$,

$$\varepsilon \geq \inf_{x \in \mathcal{X}} Pr[i(x; Y) \leq \log \gamma] - \frac{\gamma}{M} \tag{14}$$

The Meta-Converses

Theorem (Polyanskiy)

For $0 < \varepsilon < 1$, average p.o.e., every (M, ε) code satisfies the following for any output distribution Q_Y

$$M \leq \sup_{P_X} \frac{1}{\beta_{1-\varepsilon}(P_{XY}, P_X Q_Y)}. \tag{15}$$

Theorem (Polyanskiy)

For $0 < \varepsilon < 1$, maximal p.o.e., every (M, ε) code, with codewords from $\mathbb{F}$, satisfies the following for any output distribution Q_Y

$$M \leq \sup_{x \in \mathbb{F}} \frac{1}{\beta_{1-\varepsilon}(W(.|x), Q_Y)}. \tag{16}$$

Converse: Results

- For AWGN channels with $0 < \varepsilon < 1$, maximal p.o.e., SNR P, capacity C_G and $V = \frac{P(P+2)}{2(P+1)^2} \log_2^2(e)$,

$$\log M \leq nC_G + \sqrt{nV}\Phi^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1) \tag{17}$$

- For DMC with $V_D > 0$,

$$\log M \leq nC_D + \sqrt{nV_D}\Phi^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1) \tag{18}$$

where $V_D = \min_{P \in \Pi} V(P; W)$ for $0 < \varepsilon < 1/2$ and $V_D = \max_{P \in \Pi} V(P; W)$ for $1/2 < \varepsilon < 1$.

A useful example: BSC

For *BSC* with crossover probability $\alpha \neq 0, 0.5, 1$, we have

$$\log M = n(1-h(\alpha)) + \sqrt{n\alpha(1-\alpha)} \log\left(\frac{1-\alpha}{\alpha}\right) \Phi^{-1}(\varepsilon) + \frac{\log n}{2} + O(1) \tag{19}$$

- For $\alpha = 0.11$, $\varepsilon = 10^{-3}$ and $n \geq 20$, the achievability and converse gap in $\log M$ is less than 4 bits.
- Even though we don't know what code achieves that...

Recent Advances: $\beta - \beta$ bounds

Theorem (Yang et. al.)

For $0 < \varepsilon < 1$, average p.o.e., there exists an (M, ε) code that satisfies the following for any input distribution P_X, output distribution Q_Y and $0 < \delta < \varepsilon$

$$M \geq \frac{\beta_{\delta}(P_Y, Q_Y)}{\beta_{1-\varepsilon+\delta}(P_{XY}, P_X Q_Y)}. \tag{20}$$

Theorem (Polyanskiy, Verdu)

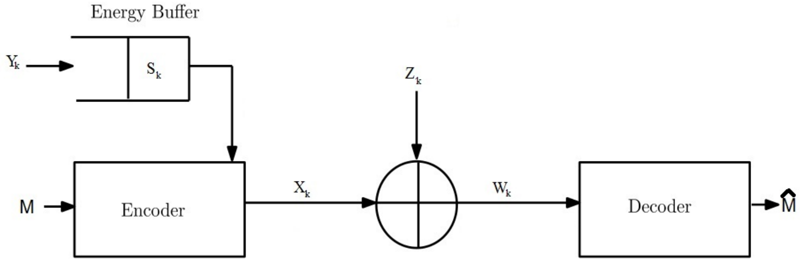
For $0 < \varepsilon < 1$, average p.o.e., every (M, ε) code satisfies the following for any output distribution Q_Y and $0 < \delta < 1 - \varepsilon$

$$M \leq \sup_{P_X} \frac{\beta_{1-\delta}(P_Y, Q_Y)}{\beta_{1-\varepsilon-\delta}(P_{XY}, P_X Q_Y)}. \tag{21}$$

Recent Advances

- The beta-beta bounds share a duality similar to KL divergence.
- These bounds have been proven to be tight, including meta converses.
- General recipe is to start with one of these bounds and use tools like Berry Esseen to refine results.

Energy Harvesting AWGN (EH-AWGN)



- If the energy arrival process has mean μ_Y, then capacity is $\frac{1}{2}\log(1+\frac{\mu_Y}{\sigma^2})$.
- For this channel, it was shown that (Fong et. al., Shenoy et. al.)
$$\log M = nC + \Theta(\sqrt{n}) \tag{22}$$
- No matching second order term as of now.

Summary

- Finite blocklength analysis is more useful practically as opposed to the limiting case.
- Lots of tools to refine second order asymptotics.
- Research in progress for energy harvesting channels and fading channels.
- Major issue is in obtaining matching second order coefficients.

It's a fact
A ratio immutable
Of circle round and width
Produces geometry's deepest conundrum
For as the numerals stay random
No repeat lets out its presence.
Yet it forever stretches forth.
Nothing to eternity.