Asymptotic Capacity of a Random Channel

Tobias Sutter, David Sutter, and John Lygeros

Automatic Control Laboratory & Institute for Theoretical Physics

Allerton Conference, 2014
Discrete memoryless channel (DMC) $W^{(n)} : \mathcal{X} \rightarrow \mathcal{Y}$, where $\mathcal{X} = \{1, 2, \ldots, n\}$, $\mathcal{Y} = \{1, 2, \ldots, n\}$

Capacity of $W^{(n)}$ [Shannon’48]

$$C(W^{(n)}) = \max_{p} I(p, W^{(n)})$$

$p \in \Delta_n$

What if the channel matrix $W^{(n)}$ is chosen at random?
- e.g. Dirichlet distributed row vectors of $W^{(n)}$
Outline

- Random channel construction
- Capacity of a random channel
- Proof (sketch)
- Numerical example
- Rate of convergence
- Summary & Outlook
Random channel construction

- Consider a probability space \((\Omega, \mathcal{A}, \mathbb{P})\)
- Construction
  1. Let \((V_{x,y})_{x,y \in [n]}\) be i.i.d. nonnegative random variables on \(\Omega\) such that \(\mathbb{E}[(V_{x,y} \log V_{x,y})^2] < \infty\)
  2. Channel matrix \(W^{(n)}\) with components \(W_{x,y}^{(n)} := \frac{V_{x,y}}{\sum_{y \in [n]} V_{x,y}}\)
- Special case: i.i.d. Dirichlet distributed row vectors of \(W^{(n)}\)

**Lemma (Measurability)**

The capacity of such a channel \(C(W^{(n)}) := \max_{p \in \Delta_n} I(p, W^{(n)})\) and the optimal input distributions \(p^*(W^{(n)}) \in \arg \max_{p \in \Delta_n} I(p, W^{(n)})\) are random variables.

- Questions: \(C(W^{(n)}) \xrightarrow{n \to \infty} ?\), \(p^*(W^{(n)}) \xrightarrow{n \to \infty} ?\)
Capacity of a random channel

- Define $\mu_1 := \mathbb{E}[V_{x,y}]$ and $\mu_2 := \mathbb{E}[V_{x,y} \log V_{x,y}]$

Theorem (Asymptotic capacity)

For $n \to \infty$ the capacity $C(W^{(n)})$ of the DMC defined above converges to $\frac{\mu_2}{\mu_1} - \log \mu_1$ almost surely and in $L^2$. 
Capacity of a random channel

- Define $\mu_1 := \mathbb{E}[V_{x,y}]$ and $\mu_2 := \mathbb{E}[V_{x,y} \log V_{x,y}]$

Theorem (Asymptotic capacity)

For $n \to \infty$ the capacity $C(W^{(n)})$ of the DMC defined above converges to $\frac{\mu_2}{\mu_1} - \log \mu_1$ almost surely and in $L^2$.

The asymptotic capacity

- is nonnegative (direct consequence of Jensen’s inequality)
- can be zero (e.g., consider $V_{x,y}$ concentrated at a point)
- can be arbitrarily large (e.g., consider random variables $V_{x,y}$ such that $\mathbb{P}[V_{x,y} = 0] = 1 - \varepsilon$ and $\mathbb{P}[V_{x,y} = 1] = \varepsilon$ leading to $\frac{\mu_2}{\mu_1} - \log \mu_1 = \log \frac{1}{\varepsilon}$)
Define $\mu_1 := \mathbb{E}[V_{x,y}]$ and $\mu_2 := \mathbb{E}[V_{x,y} \log V_{x,y}]$.

**Theorem (Asymptotic capacity)**

For $n \to \infty$ the capacity $C(W^{(n)})$ of the DMC defined above converges to $\frac{\mu_2}{\mu_1} - \log \mu_1$ almost surely and in $L^2$.

**Example 1 (Uniform distribution)**

Consider $V_{x,y} \overset{i.i.d.}{\sim} \mathcal{U}([0, A])$ for $A > 0$, then $\lim_{n \to \infty} C(W^{(n)}) = 1 - \frac{1}{2 \ln 2}$.
Define $\mu_1 := E[V_{x,y}]$ and $\mu_2 := E[V_{x,y} \log V_{x,y}]$

**Theorem (Asymptotic capacity)**

For $n \to \infty$ the capacity $C(W^{(n)})$ of the DMC defined above converges to $\frac{\mu_2}{\mu_1} - \log \mu_1$ almost surely and in $L^2$.

**Example 1 (Uniform distribution)**

Consider $V_{x,y} \overset{\text{i.i.d.}}{\sim} U([0, A])$ for $A > 0$, then $\lim_{n \to \infty} C(W^{(n)}) = 1 - \frac{1}{2 \ln 2}$

**Example 2 (Symmetric Dirichlet distribution)**

Consider $V_{x,y} \sim \mathcal{E}(\lambda)$ for $\lambda > 0$, then the channel rows are i.i.d. Dirichlet distributed, i.e., $W^{(n)}_{x,y} \overset{\text{i.i.d.}}{\sim} \text{Dir}(\lambda, \ldots, \lambda)$ and $\lim_{n \to \infty} C(W^{(n)}) = \frac{1-\gamma}{\ln 2}$, where $\gamma \approx 0.5772$ denotes Euler’s constant
Proof sketch
Proof sketch (lower bound)

- Primal program \( C(W^{(n)}) = \left\{ \begin{array}{l}
\max_p \quad I(p, W^{(n)}) \\
p \in \Delta_n
\end{array} \right. \)

- Lower bound \( C_{LB}^{(p \sim \mathcal{U})}(W^{(n)}) := I(p, W^{(n)}) \), where \( p \) is the uniform distribution on \( \Delta_n \)

Lemma (Lower bound)

For \( n \to \infty \), the random variable \( C_{LB}^{(p \sim \mathcal{U})}(W^{(n)}) \) converges to \( \frac{\mu_2}{\mu_1} - \log \mu_1 \) almost surely and in \( L^2 \).
Proof sketch (lower bound)

- Primal program
  \[ C(W^{(n)}) = \max_{p} I(p, W^{(n)}) \quad p \in \Delta_n \]

- Lower bound
  \[ C_{LB}^{(p\sim\mathcal{U})}(W^{(n)}) := I(p, W^{(n)}) \], where \( p \) is the uniform distribution on \( \Delta_n \)

Lemma (Lower bound)

For \( n \to \infty \), the random variable \( C_{LB}^{(p\sim\mathcal{U})}(W^{(n)}) \) converges to \( \frac{\mu_2}{\mu_1} - \log \mu_1 \) almost surely and in \( L^2 \).

- Introduce \( q := W^{(n)^T} p \) \hspace{1cm} [Chiang & Boyd’04]
- Equivalent formulation (prerequisite for the upper bound)

\[ C(W^{(n)}) = \max_{p,q} -r^T p + H(q) \quad s.t. \quad W^{(n)^T} p = q \quad p \in \Delta_n, \quad q \in \Delta_n \]

with \( r_i := -\sum_{j=1}^{N} W_{i,j}^{(n)} \log W_{i,j}^{(n)} \)
Proof sketch (upper bound)

Dual program (with strong duality)

\[
C(W^{(n)}) = \min_{\lambda} \{ G(\lambda) + F(\lambda) : \lambda \in \mathbb{R}^n \},
\]

parametric LP

\[
G(\lambda) = \max_p -r^T p + \lambda^T W^{(n)^T} p \quad \text{s.t.} \quad p \in \Delta_n,
\]

\[
= \max_{i \in [n]} \left( W^{(n)} \lambda - r \right)_i
\]

Entropy maximization

\[
F(\lambda) = \max_q H(q) - \lambda^T q \quad \text{s.t.} \quad q \in \Delta_n
\]

\[
= \log \left( \sum_{i=1}^n 2^{-\lambda_i} \right)
\]

Lemma (Upper bound)

For \( n \to \infty \), the random variable \( C_{UB}^{(\lambda=0)}(W^{(n)}) := G(0) + F(0) \) converges to \( \frac{\mu_2}{\mu_1} - \log \mu_1 \) almost surely and in \( L^2 \).
Example 1 (Uniform distribution)

[TS et al.’14]
Example 2 (Symmetric Dirichlet distribution)

[TS et al.'14]
Rate of convergence

- Assume $V_{x,y} \in [a, b]$ almost surely for $0 \leq a < b < \infty$
- $c := \min_{x \in [a,b]} x \log x$, $d := \max_{x \in [a,b]} x \log x$
- $f_1(t, n) := 2 \exp \left( -\frac{2nt^2}{(b-a)^2} \right)$, $f_2(t, n) := 2 \exp \left( -\frac{2nt^2}{(d-c)^2} \right)$

Theorem (Rate of convergence)

The capacity of the DMC defined above satisfies for any $t \in \mathbb{R}_{>0}$

$$
\begin{align*}
\mathbb{P} \left[ \left| C(W(n)) - \left( \frac{\mu_2}{\mu_1} - \log \mu_1 \right) \right| \geq t \right] \\
\leq \left( f_1(\alpha_{t/2}, n) + f_2(\alpha_{t/2}, n) + f_1(\frac{t}{2L}, n) \right) \vee \\
\left( f_1(\alpha_{t/4}, n) + f_2(\alpha_{t/4}, n) + f_1(\frac{t}{4L}, n) + 2f_1(\beta_{t/(2L)}, n) \right),
\end{align*}
$$

\[\beta_t = \frac{t \mu_1}{2 + t}, \quad L = \frac{1}{a \ln 2}, \quad \alpha_t = \left\{ \begin{array}{ll} 
\frac{t \mu_1^2}{\mu_1(1+t) + \mu_2} & \text{if } \mu_1 + \mu_2 \geq 0 \\
\frac{t \mu_2^2}{\mu_1(1-t) + \mu_2} & \text{otherwise}
\end{array} \right. \]
Proof (sketch)

Key steps

- Main idea is to derive concentration bounds for the random variables $C_{UB}^{(\lambda=0)}(W(n))$ and $C_{LB}^{(p\sim U)}(W(n))$ around $\frac{\mu_2}{\mu_1} - \log \mu_1$
- Main tool is Hoeffding’s inequality

Remarks

- Boundedness assumption of the random variables can be relaxed (e.g., by using a general Bernstein inequality)
- The rate is by no means optimal
Summary & Outlook

Summary

- Asymptotic capacity of DMCs, whose channel entries are chosen at random
- Exponential rate of convergence

Outlook

- Application and interpretation of the limiting capacity of a random channel in the context of Bayesian estimation and optimal experiment design.
- Decay rate of the variance
- Remove the i.i.d. assumption