$\boxed{\text{Theorem 2.1.3}}$ Weak-law of large numbers for i.i.d. random variables

Let $\{X_n\}_{n\geq 1}$ be independent, identically distributed random variables with mean values $\mathbb{E}[X_n] = \mu$ and variances $\mathbb{E}[(X_n-\mu)^2] = \sigma^2 < +\infty$.

Then,
$$\boxed{\lim_n \operatorname{Prob}\left(\left| \frac{1}{n}\sum_{j=1}^{n} X_j - \mu \right| \geq \varepsilon \right) = 0 \quad \forall\ \varepsilon > 0.}$$

Proof: Let X be a random variable, then

$$\sum_i p(x_i)\, x_i^2 \geq \sum_{i:\ x_i \geq \eta} p(x_i)\, x_i^2 \geq \eta^2 \sum_{i:\ x_i\geq \eta} p(x_i) = \eta^2 \operatorname{Prob}(X \geq \eta)$$

$$\boxed{\operatorname{Prob}(X \geq \eta) \leq \frac{\mathbb{E}[X^2]}{\eta^2}}$$ (Chebyshev inequality)

Set $X = \left| \frac{1}{n}\sum_{j=1}^{n} X_j - \mu \right|$, then

$$\operatorname{Prob}\left(\left| \frac{1}{n}\sum_{j=1}^{n} X_j - \mu \right| \geq \eta \right) \leq \frac{1}{\eta^2}\, \mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n} X_j - \mu \right)^2 \right]$$

Now, $$\mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right)^2\right] = \frac{1}{n^2}\sum_{j=1}^{n}\mathbb{E}[X_j^2] + \frac{1}{n^2}\sum_{j\neq k=1}^{n}\mathbb{E}[X_j X_k] - \frac{2\mu}{n}\sum_{j=1}^{n}\mathbb{E}[X_j] + \mu^2$$

$$\boxed{j\neq k:\ \mathbb{E}[X_j X_k] = \mathbb{E}[X_j]\,\mathbb{E}[X_k] = \mu^2 \quad \text{because of i.i.d.}}$$

$$= \frac{1}{n^2}\sum_{j=1}^{n}\mathbb{E}[X_j^2] + \frac{n(n-1)}{n^2}\mu^2 - \mu^2$$

$$= \frac{1}{n^2}\sum_{j=1}^{n}\mathbb{E}[X_j^2] - \frac{\mu^2}{n}$$

$$= \frac{1}{n^2}\sum_{j=1}^{n}\left(\mathbb{E}[X_j^2] - \mu^2\right) = \frac{\sigma^2}{n}$$

Thus, $$\operatorname{Prob}\left(\left|\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right| \geq \eta\right) \leq \frac{\sigma^2}{n\eta} \xrightarrow[n]{} 0$$

and $$\boxed{\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{P} \mu}$$

$\boxed{\text{Corollary } 2.1.1}$ For a Bernoulli source, the information rate equals the Shannon entropy of a single use:

$$\boxed{H = H(\pi) = -\sum_{x_i \in \mathcal{X}} p(x_i) \log p(x_i)}$$

Proof: since $\boxed{p(x_{i(n)}) = \prod_{j=1}^{n} p(x_{i_j})}$ $\boxed{\xi_n(x_{i(n)}) = -\frac{1}{n} \sum_{j=1}^{n} \log p(x_{i_j})}$

The i.i.d random variables are $Y_j = -\log p(x_j)$

$$\mathbb{E}[Y_j] = -\sum_{k:\, x_{j_k} \in \mathcal{X}} p(x_{j_k}) \log p(x_{j_k}) = H(\pi) \quad (= \mu \text{ in the w.l.l.n.})$$

$$\mathbb{E}[Y_j^2] - (H(\pi))^2 = \sum_{k:\, x_{j_k} \in \mathcal{X}} p(x_{j_k}) \left(\log p(x_{j_k}) - H(\pi)\right)^2 \quad (= \sigma^2 \text{ in the w.l.l.n.})$$

Then, $\boxed{\xi_n \xrightarrow{P} H(\pi)}$

Remark: the convergence in probability of the log-likelihood for more general sources than Bernoulli ones is the content of the Shannon - McMillan - Breiman theorem (see Cover & Thomas page 474-475)

The convergence is to the Shannon entropy rate and holds for stationary ergodic sources.

Therefore, for this class of sources the optimum compression rate is given by the Shannon entropy rate.

Remark: a source is stationary if the probabilities $p(x_{i(n)})$ depend only on the words and not on the uses at which they are emitted:

$$H(X_1, X_2, \cdots, X_n) = H(X_{1+p}, X_{2+p}, \cdots, X_{n+p})$$

(Prove this).

<u>Remark</u>: a stationary source is <u>ergodic</u> if the probability measure π defined on the stochastic process $\{X_j\}_{j\geq 1}$ by the probabilities

$\pi^{(n)} = \{ p(x_{i(n)}) \}_{x_{i(n)} \in \mathcal{X}^{(n)}}$ <u>cannot</u> be convexly decomposed as $\pi = \mu \pi_1^* + (1-\mu)\pi_2 \quad 0 \leq \mu \leq 1$ with $\pi_{1,2}$ both stationary.

$\boxed{\text{Theorem } 2.1.4.}$ <u>Shannon - McMillan - Breiman</u>

For stationary ergodic sources with entropy rate h,

$$-\frac{1}{n} \log p(X_1, X_2, \dots, X_n) \xrightarrow{P} h.$$

<u>Question</u>: what is the entropy rate h ?

Proposition 2.1.1 For stationary sources the Shannon entropy rate is

$$h = \lim_n \frac{H(X_1, \dots, X_n)}{n} = \inf \frac{H(X_1, \dots, X_n)}{n}.$$

<u>Proof:</u> set $H_n := H(X_1, X_2, \dots, X_n)$ and let $n = mp+q$ with $m \in \mathbb{N}$, $p \in \mathbb{N}$ fixed and $0 \leq q < p$.

Then, by sub-additivity of the Shannon entropy and by stationarity:

$$H_{mp+q} \leq H_p + H(X_{p+1}, \dots, X_{2p}) + \dots + H(X_{(m-1)p+1}, \dots, X_{mp})$$
$$+ H(X_{mp+1}, \dots, X_{mp+q})$$
$$\leq m H_p + (q-1) H(X)$$

Therefore,

$$\limsup_n \frac{H_n}{n} \leq \limsup_m \frac{H_{mp+q}}{mp+q} \leq \frac{H_p}{p} \leq \liminf_n \frac{H_n}{n}.$$

Exercise 2.1.2. Show that for an ergodic stationary Markov source the optimum compression rate
$$\boxed{H = h = -\sum_{x_1, x_2 \in \mathcal{X}} p(x_2|x_1)\, p(x_1) \log p(x_2|x_1)}$$

Shannon – Mc Millan – Breiman theorem ensures that $\boxed{H = \lim_n \frac{1}{n} H(X_1, \dots, X_n) = h}$

From $p(x_{i(n)}) = \left(\prod_{j=1}^{n-1} p(x_{i_{j+1}} | x_{i_j}) \right) p(x_{i_1})$ it follows that

$$H(X_1, \dots, X_n) = -\sum_{i(n)} p(x_{i(n)}) \log p(x_{i(n)}) = -\sum_{i(n)} \left(\prod_j p(x_{i_{j+1}}|x_{i_j})\, p(x_{i_1}) \right) \left(\sum_{k=1}^{n-1} \left(\log p(x_{i_{k+1}}|x_{i_k}) \right) + \log p(x_{i_1}) \right)$$

$$= -\sum_{k=1}^{n-1} \sum_{i(n)} \left(\prod_{j=1}^{n-1} p(x_{i_{j+1}}|x_{i_j}) \right) p(x_{i_1}) \log p(x_{i_{k+1}}|x_{i_k}) - \sum_{i(n)} \prod_{j=1}^{n-1} p(x_{i_{j+1}}|x_{i_j})\, p(x_{i_1}) \log p(x_{i_1})$$

Using compatibility, $\sum_{x_2 \in \mathcal{X}} p(x_2|x_1) = 1$, and stationarity, $\sum_{x_1 \in \mathcal{X}} p(x_2|x_1)\, p(x_1) = p(x_2)$,

$$\frac{1}{n} H(X_1, \dots, X_n) = -\frac{1}{n} \sum_{k=1}^{n-1} \sum_{x_{i_{k+1}}, x_{i_k} \in \mathcal{X}} p(x_{i_k})\, p(x_{i_{k+1}}|x_{i_k}) \log p(x_{i_{k+1}}|x_{i_k}) - \frac{1}{n} \sum_{x \in \mathcal{X}} p(x) \log p(x)$$

$$= -\sum_{x_1, x_2 \in \mathcal{X}} p(x_1)\, p(x_2|x_1) \log p(x_2|x_1) - \frac{1}{n} \sum_{x \in \mathcal{X}} p(x) \log p(x)$$

2.2. Quantum Bernoulli Sources: Schumacher's Theorem

Quantum sources are emitters of quantum states.

— After n uses, a quantum source will have emitted state vectors $\boxed{|\psi_\alpha^{(n)}\rangle \in (\mathbb{C}^m)^{\otimes n} = \mathbb{C}^{m^n}}$ with weights $\boxed{\lambda_\alpha^{(n)}, \ \lambda_\alpha^{(n)} \geq 0, \ \sum_\alpha \lambda_\alpha^{(n)} = 1,}$ the state vectors being not necessarily orthogonal.

— After n uses, the statistics of the quantum source is determined by the density matrix

$$\rho^{(n)} = \sum_\alpha \lambda_\alpha^{(n)} |\psi_\alpha^{(n)}\rangle\langle\psi_\alpha^{(n)}| \in \left(M_m(\mathbb{C})\right)^{\otimes n} = M_{m^n}(\mathbb{C})$$

Definition 2.2.1

A Bernoulli (i.i.d) quantum sources is such that

$$\rho^{(n)} = \rho \otimes \rho \otimes \cdots \otimes \rho = \rho^{\otimes n}$$

Remark: if $\rho = \sum_{i=1}^{m} r_i |r_i\rangle\langle r_i|$ is the spectralization of ρ,

then $\rho^{(n)} = \sum_{i(n)} r_{i(n)} |r_{i(n)}\rangle\langle r_{i(n)}|$

$r_{i(n)} = r_{i_1} r_{i_2} \cdots r_{i_n}$, $r_{i(n)} \geq 0$, $\sum_{i(n)} r_{i(n)} = 1$

$|r_{i(n)}\rangle = |r_{i_1}\rangle \otimes |r_{i_2}\rangle \otimes \cdots \otimes |r_{i_n}\rangle \in \mathbb{C}^{m^n}$

Remark: the eigenvalues $r_{i(n)}$ of $\rho^{(n)}$ are words of length n with letters from the alphabet consisting of the eigenvalues of ρ.

Remark : • the log-likelihoods $\boxed{-\frac{1}{n}\log r_{i^{(n)}} \xrightarrow{P} h = S(\rho) = -\sum_{i=1}^{m} r_i \log r_i}$

• Because of the asymptotic equipartition theorem (see Example 2.1.3)

out of words $r_{i^{(n)}} \in \underbrace{Sp(\rho)\times Sp(\rho)\times\cdots\times Sp(\rho)}_{n\text{ times}} = Sp(\rho)^{\times n}$

we can extract a typical subset

$$T^{(n)}_{\varepsilon,\eta} \equiv \left\{ r_{i^{(n)}} : \left|-\frac{1}{n}\log r_{i^{(n)}} - S(\rho)\right| \leq \varepsilon\right\}$$

such that $\boxed{\mathrm{Prob}\left(T^{(n)}_{\varepsilon,\eta}\right) = \sum_{i^{(n)}:\ r_{i^{(n)}}\in T^{(n)}_{\varepsilon,\eta}} r_{i^{(n)}} \geq 1-\eta}$

and $\boxed{(1-\eta)\, 2^{n(S(\rho)-\varepsilon)} \leq \#\left(T^{(n)}_{\varepsilon,\eta}\right) \leq 2^{n\left(S(\rho)+\varepsilon\right)}}$

• Let $\tilde{\mathcal{H}}^{(n)}_{\varepsilon,\eta}$ be the linear span of the eigenvectors $\boxed{|r_{i^{(n)}}\rangle,\ r_{i^{(n)}}\in T^{(n)}_{\varepsilon,\eta}}$

and $\tilde{P}^{(n)}_{\varepsilon,\eta} : \mathbb{C}^{m^n} \to \tilde{\mathcal{H}}^{(n)}_{\varepsilon,\eta}$ the corresponding orthogonal projections

$$\tilde{P}^{(n)}_{\varepsilon,\eta} = \sum_{i^{(n)}:\ r_{i^{(n)}}\in T^{(n)}_{\varepsilon,\eta}} |r_{i^{(n)}}\rangle\langle r_{i^{(n)}}|$$

$\boxed{\text{Lemma } 2.2.1.}$ $\quad \mathrm{Prob}\left(T^{(n)}_{\varepsilon,\eta}\right) = \mathrm{Tr}\left(\rho^{(n)}\, \tilde{P}^{(n)}_{\varepsilon,\eta}\right)$

Proof:
$$\mathrm{Tr}\left(\rho^{(n)}\, \tilde{P}^{(n)}_{\varepsilon,\eta}\right) = \sum_{i^{(n)}:\ r_{i^{(n)}} \in T^{(n)}_{\varepsilon,\eta}} \langle r_{i^{(n)}} | \rho^{(n)} | r_{i^{(n)}} \rangle$$
$$= \sum_{i^{(n)}:\ r_{i^{(n)}} \in T^{(n)}_{\varepsilon,\eta}} r_{i^{(n)}} = \mathrm{Prob}\left(T^{(n)}_{\varepsilon,\eta}\right)$$

Remark: compression in the quantum setting means reducing to a Hilbert (sub)space of dimension $\boxed{d_c(n)}$ smaller than $\boxed{m^n}$, in such a way that the reduction gets closer and closer to $\boxed{\rho^{(n)}}$ when n increases.

$\boxed{\text{Definition 2.2.2}}$ A reliable compression-decompression scheme for a quantum source described by the density matrices $\boxed{\rho^{(n)} = \sum_\alpha \lambda_\alpha^{(n)} |\psi_\alpha^{(n)}\rangle\langle\psi_\alpha^{(n)}|}$

- consists of encoding CPT maps $\mathcal{E}^{(n)} : M_{m^n}(\mathbb{C}) \longrightarrow M_{d_c(n)}(\mathbb{C})$ such that $\boxed{\mathcal{E}^{(n)}[|\psi_\alpha^{(n)}\rangle\langle\psi_\alpha^{(n)}|] = \tilde{\rho}_\alpha^{(n)} \in M_{d_c(n)}(\mathbb{C})}$, $\boxed{d_c(n) \leq m^n}$,
- and decoding CPT maps $\mathcal{D}^{(n)} : M_{d_c(n)}(\mathbb{C}) \longrightarrow M_{m^n}(\mathbb{C})$
- such that the $\boxed{\text{fidelities}}$ $F_n := \sum_\alpha \lambda_\alpha^{(n)} \langle\psi_\alpha^{(n)}|\mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}]|\psi_\alpha^{(n)}\rangle$ tend to 1 when n tends to infinity. $\boxed{F_n \xrightarrow[n]{} 1}$.
- The associated compression rates are $R_n := \frac{1}{n}\frac{\log d_c(n)}{\log m}$ and one looks for the optimum $\boxed{R := \lim_n R_n}$ such that $\boxed{F_n \xrightarrow[n]{} 1}$

<u>Remark</u> : the fidelities $F_n = \sum_\alpha d_\alpha^{(n)} \langle \psi_\alpha^{(n)} | \mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] | \psi_\alpha^{(n)} \rangle \leq 1$.

Indeed, $\mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}]$ is a density matrix and then $\boxed{\mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] \leq \mathbb{1}}$

whence $F_n \leq \sum_\alpha d_\alpha^{(n)} = 1$.

Also, $F_n = 1$ iff $\mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] = |\psi_\alpha^{(n)}\rangle\langle\psi_\alpha^{(n)}|$.

Indeed, if $\langle \psi_\alpha^{(n)} | \mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] | \psi_\alpha^{(n)} \rangle < 1$ for some α then $F_n < 1$.

Then, $\langle \psi | \rho \psi \rangle = 1 \iff \rho = |\psi\rangle\langle\psi|$. (Prove it)

- $\boxed{\text{Compression and decompression scheme}}$

- $\mathcal{E}^{(n)}[|\psi_\alpha^{(n)}\rangle\langle\psi_\alpha^{(n)}|] = \tilde{\rho}_\alpha^{(n)} = \mu_\alpha^2 \, |\tilde{\varphi}_\alpha^{(n)}\rangle\langle\tilde{\varphi}_\alpha^{(n)}| + \gamma_\alpha^2 \, |\tilde{\varphi}_0\rangle\langle\tilde{\varphi}_0|$

$$|\tilde{\varphi}_\alpha^{(n)}\rangle = \frac{\tilde{P}_{\varepsilon,\eta}^{(n)} |\psi_\alpha^{(n)}\rangle}{\| \tilde{P}_{\varepsilon,\eta}^{(n)} |\psi_\alpha^{(n)}\rangle \|} \in \mathcal{H}_{\varepsilon,\eta}^{(n)} \quad , \quad |\tilde{\varphi}_0\rangle \in \mathcal{H}_{\varepsilon,\eta}^{(n)}$$

$$\mu_\alpha^2 = \| \tilde{P}_{\varepsilon,\eta}^{(n)} |\psi_\alpha^{(n)}\rangle \|^2 \quad , \quad \gamma_\alpha^2 = \| (1 - \tilde{P}_{\varepsilon,\eta}^{(n)}) |\psi_\alpha^{(n)}\rangle \|^2$$

$\boxed{\text{Exercise } 2.2.1}$ Express $\mathcal{E}^{(n)}$ in Kraus-Stinespring form.

$$\mathcal{E}^{(n)} : M_{m^n}(\mathbb{C}) \longrightarrow M_{d_c(n)}(\mathbb{C}) \; ; \quad (1-\eta)\, 2^{n(S(\rho)-\varepsilon)} \leq d_c(n) \leq 2^{n(S(\rho)+\varepsilon)}$$

$$\mathcal{E}^{(n)}[X] = \tilde{P}^{(n)}_{\varepsilon,\eta}\, X\, \tilde{P}^{(n)}_{\varepsilon,\eta} + \sum_{i^{(n)} :\, r_{i^{(n)}} \notin T^{(n)}_{\varepsilon,\eta}} |\tilde{\psi}_0\rangle\langle r_{i^{(n)}}|\, X\, |r_{i^{(n)}}\rangle\langle\tilde{\psi}_0|$$

$$\mathcal{E}^{(n)}\left[|\psi^{(n)}_\alpha\rangle\langle\psi^{(n)}_\alpha|\right] = \tilde{P}^{(n)}_{\varepsilon,\eta}|\psi^{(n)}_\alpha\rangle\langle\psi^{(n)}_\alpha|\tilde{P}^{(n)}_{\varepsilon,\eta} + |\tilde{\psi}_0\rangle\langle\tilde{\psi}_0| \sum_{i^{(n)} :\, r_{i^{(n)}} \notin T^{(n)}_{\varepsilon,\eta}} |\langle\psi^{(n)}_\alpha|r_{i^{(n)}}\rangle|^2$$

$$= \underbrace{\|\tilde{P}^{(n)}_{\varepsilon,\eta}|\psi^{(n)}_\alpha\rangle\|^2}_{\mu_\alpha^2} |\tilde{\psi}^{(n)}_\alpha\rangle\langle\tilde{\psi}^{(n)}_\alpha| + \underbrace{\|(1-\tilde{P}^{(n)}_{\varepsilon,\eta})|\psi_\alpha\rangle\|^2}_{\nu_\alpha^2} |\tilde{\psi}_0\rangle\langle\tilde{\psi}_0|$$

- Decompression : $\mathcal{D}^{(n)}[\tilde{\rho}^{(n)}_\alpha] = \tilde{\rho}^{(n)}_\alpha \oplus 0$

$\tilde{\rho}^{(n)}_\alpha$ is a $d_c(n) \times d_c(n)$ matrix which can be turned into a $m^n \times m^n$ matrix by adding suitable rows and columns consisting of zeroes.

• Fidelities: $F_n = \sum_\alpha d_\alpha^{(n)} \langle \psi_\alpha^{(n)} | \mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] \psi_\alpha^{(n)} \rangle$

$$= \sum_\alpha d_\alpha^{(n)} \left(\mu_\alpha^2 \left| \langle \psi_\alpha^{(n)} | \tilde{\psi}_\alpha^{(n)} \rangle \right|^2 + \nu_\alpha^2 \left| \langle \psi_\alpha^{(n)} | \tilde{\psi}_0 \rangle \right|^2 \right)$$

$$\geq \sum_\alpha d_\alpha^{(n)} \mu_\alpha^2 \left| \langle \psi_\alpha^{(n)} | \tilde{\psi}_\alpha^{(n)} \rangle \right|^2 = \sum_\alpha d_\alpha^{(n)} \mu_\alpha^4$$

$$\geq \sum_\alpha d_\alpha^{(n)} \left(2\mu_\alpha^2 - 1 \right) = 2 \sum_\alpha d_\alpha^{(n)} \mu_\alpha^2 - 1$$

Indeed, $\mu_\alpha^2 = \| \tilde{P}_{\varepsilon,\eta}^{(n)} | \psi_\alpha^{(n)} \rangle \|^2 = \langle \psi_\alpha^{(n)} | \tilde{P}_{\varepsilon,\eta}^{(n)} | \psi_\alpha^{(n)} \rangle = \langle \psi_\alpha^{(n)} | \tilde{\psi}_\alpha^{(n)} \rangle \mu_\alpha$

and $(x-1)^2 = x^2 - 2x + 1 \geq 0 \Rightarrow x^2 \geq 2x - 1$

Theorem 2.2.1 <u>Schumacher</u>

Let $\{\rho^{(n)} = \rho^{\otimes n}\}_{n \geq 1}$ describe an i.i.d Bernoulli quantum source.

If $\boxed{R > S(\rho)}$ then there exists a reliable compression-decompression scheme. If $\boxed{R < S(\rho)}$ then any compression-decompression scheme with that rate is <u>unreliable</u>.