ACHILLE
CANNAVALE

APPUNTI

INFORMATION THEORY

2023

CIAO! QUESTI APPUNTI SONO FRUTTO DEL MIO STUDIO E DELLA MIA INTERPRETAZIONE, QUINDI POTREBBERO CONTENERE ERRORI, SVISTE O COSE MIGLIORABILI. BUONO STUDIO📚✍️

ENTROPY IS NOT NEGATIVE

$$H(x) = \mathbb{E}\left[\log_2\left(\frac{1}{P(x)}\right)\right] = \sum_{x \in A_x} P(x) \cdot \log_2\left(\frac{1}{P(x)}\right)$$

≥ 0 overall; $P(x) \geq 0$; $\log_2\left(\frac{1}{P(x)}\right) \geq 0$

IMPORTANTE

Ciao! Questi appunti sono frutto del mio studio e della mia interpretazione, quindi potrebbero contenere errori, sviste o cose migliorabili. Buono studio📚✍️

ENTROPY CHAIN RULE

$$H(x,y) = H(x|y) + H(y)$$

PROOF

$$H(x,y) = \mathbb{E}\left[\log_2\left(\frac{1}{P(x,y)}\right)\right] = \mathbb{E}\left[\log_2\left(\frac{1}{P(x|y)\cdot P(y)}\right)\right] = \mathbb{E}\left[\log_2\left(\frac{1}{P(x|y)}\right) + \log_2\left(\frac{1}{P(y)}\right)\right] =$$

$$= H(x|y) + H(y) \quad \square$$

FANO'S INEQUALITY

$$H(x|y) \leq H(P_e) + P_e \log_2(M-1), \qquad M = \text{CARD}(A_x)$$

LET BE:

$$E = \begin{cases} 0 & \text{IF } \hat{x} = x \\ 1 & \text{, OTHERWISE} \end{cases} \qquad \text{SO } E \sim \mathcal{B}(P_e)$$

PROOF

$$H(E, x|y) = \begin{cases} H(E|x,y) + H(x|y) \\ H(x|E,y) + H(E|y) \end{cases}$$

$H(E|x,y)$ IT'S 0 SINCE x KNOWN, y KNOWN $\Rightarrow$ E KNOWN

$$\Rightarrow H(x|y) = H(x|E,y) + H(E|y) \leq H(E) = H(P_e)$$

$$= P_e\, H(x|y, E=1) + (1-P_e)\, H(x|y, E=0)$$

$H(x|y, E=0)$ IT'S 0 SINCE $x = \hat{x} = y$

$H(x|y, E=1) \leq \log_2(M-1)$ IT'S THE LARGEST ENTROPY

$$\Rightarrow H(x|y) \leq H(P_e) + P_e \log_2(M-1) \quad \square$$

ACHILLE CANNAVALE

DATA PROCESSING INEQUALITY

LET $X \to Y \to Z$ BE A MARKOV'S CHAIN, X, Y, Z R.V.s

$$I(X;Z) \leq I(X;Y)$$

PROOF

$$I(X;Y,Z) = \begin{cases} I(X;Y|Z) + I(X;Z) \\ I(X;Z|Y) + I(X;Y) \end{cases}$$

$I(X;Z|Y) = 0$ SINCE X AND $Z|Y$ ARE INDEPENDENT

$$\Rightarrow I(X;Y) = I(X;Y|Z) + I(X|Z) \Rightarrow I(X;Z) \leq I(X;Y) \quad \square$$

$I(X;Y|Z) \geq 0$

STATIONARY PROCESS

$\{X_i\}_{i \in \mathbb{Z}}$ IS A STATIONARY PROCESS IFF:

$$\mathbb{P}((X_m, \dots, X_n) \in A) = \mathbb{P}((X_{m-\Delta}, \dots, X_{n-\Delta}) \in A)$$

$\forall m \in \mathbb{Z}, \forall n \geq m, n \geq \mathbb{Z}, \forall A, \forall \Delta \in \mathbb{N}$

ENTROPY RATE

IF $\exists \lim_{n \to \infty} \frac{1}{n} H(X_1, \dots, X_n)$,

THEN $H_\infty = \lim_{n \to \infty} \frac{1}{n} H(X_1, \dots, X_n)$ $\left[\text{bit/symbol}\right]$

KRAFT'S INEQUALITY

1. C IS AN INSTANTANEOUS D-ARY SOURCE CODE WITH CODEWORD LENGTH $\{l_i\}_{i=1}^m$, $m < \infty$ $\Rightarrow \sum_{i=1}^m D^{-l_i} \leq 1$

2. IF $\{l_i\}_{i=1}^m$, $m < \infty$ ARE SUCH THAT $l_i \in \mathbb{N}$, $\forall i$ AND $\sum_{i=1}^m D^{-l_i} \leq 1$ $\Rightarrow$ $\exists$ A D-ARY INSTANTANEOUS CODEWORD WITH CODEWORD LENGTH: $\{l_i\}_{i=1}^m$

ACHILLE CANNAVALE

PROOF · 1

Every D-ary code with maximum codeword l_{MAX} can be represented with a d-ary tree.

If the code is prefix and a node is used for a codeword, then all the descending nodes cannot be used, otherwise the corresponding codewords would violate the prefix condition.

$\Longrightarrow$ A codeword of length l_i eliminates $D^{l_{MAX}-l_i}$ nodes at depth l_{MAX}

$$\sum_{i=1}^{m} D^{l_{MAX}-l_i} \leq D^{l_{MAX}} \Rightarrow \sum_{i=1}^{m} D^{-l_i} \leq 1 \quad \square$$

$\sum_{i=1}^{m} D^{l_{MAX}-l_i}$: * ELIMINATED TERMINAL NODES

$D^{l_{MAX}}$: * TERMINAL NODES

PROOF · 2

IF $l_i \in \mathbb{N}$ · AND $\sum_{i=1}^{m} D^{-l_i} \Rightarrow$

1) Build a d-ary tree of depth $l_{MAX} = MAX\{l_1, \dots, l_M\}$
2) Place a codeword on a node of depth l_i
3) Remove all descending nodes
4) Iterate until there is no l_i available
5) Assign a label to every branch of the tree
6) Assign a codeword to every node by reading the labels from the root to the node

$\square$

ACHILLE CANNAVALE

STRONG LAW LARGE NUMBER

$\{X_i\}_{i\in\mathbb{N}}$ i.i.d. r.v.

$\mathbb{E}[|X_i|] < \infty \quad \forall i \in \mathbb{N}$

LET $\bar{X}_m = \frac{1}{m}\sum_{i=1}^{m} X_i$, THEN $\bar{X}_m \xrightarrow[m\to\infty]{} \mu = \mathbb{E}[X_i]$ ALMOST SURELY AND IN MEAN SQUARE.

PROOF (WEAK LAW): $\mathrm{VAR}(X_i) = \sigma^2 < \infty$, CONV. IN $\mathbb{P}$

$\forall \varepsilon > 0 \quad \mathbb{P}(|\bar{X}_m - \mu| > \varepsilon) \xrightarrow[m\to\infty]{} 0$

$$\mathbb{P}(|\bar{X}_m - \mu| > \varepsilon) = \mathbb{P}(|\bar{X}_m - \mu|^2 > \varepsilon^2) \leq \frac{\mathbb{E}[(\bar{X}_m - \mu)^2]}{\varepsilon} = \frac{\mathrm{VAR}(\bar{X}_m)}{\varepsilon} = *$$

MARKOV'S INEQUALITY

$$\mathbb{E}[\bar{X}_m] = \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m} X_i\right] = \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}[X_i] = \mu$$

$$\mathrm{VAR}(\bar{X}_m) = \mathrm{VAR}\left(\frac{1}{m}\sum_{i=1}^{m} X_i\right) = \frac{1}{m}\left(\sum_{i=1}^{m}\underbrace{\mathrm{VAR}(X_i)}_{\sigma^2} + \sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq i}}^{m}\underbrace{\mathrm{VAR}(X_i, X_j)}_{0 \Leftarrow \text{i.i.d}}\right)$$

$$* = \frac{\sigma^2}{m\varepsilon} \xrightarrow[m\to\infty]{} 0 \quad \square$$

ASYMPTOTIC EQUIPARTITION PROPERTY

$\{X_i\}_{i\in\mathbb{N}}$ i.i.d

$H(X_i) = H(X) < \infty$

$$\frac{1}{m}\log_2\left(\frac{1}{P(X_1,\dots,X_m)}\right) \xrightarrow[m\to\infty]{} H(X)$$ ALMOST SURE AND IN MEAN SQUARE.

PROOF

$$\frac{1}{m}\log_2\left(\frac{1}{P(X_1,\dots,X_m)}\right) = \frac{1}{m}\sum_{i=1}^{m}\underbrace{\log_2\left(\frac{1}{P(X_i)}\right)}_{Y_i}$$

IF IT SATISFIES STRONG LAW LARGE NUMBER $\Rightarrow \xrightarrow[m\to\infty]{} \mathbb{E}[Y_i] = \mathbb{E}\left[\log_2\left(\frac{1}{P(X_i)}\right)\right] = H(X)$

- i.i.d? YES, CAUSE Y_i IS A FUNCTION OF X_i THAT IS i.i.d.
- $\mathbb{E}[|Y_i|] < \infty$? YES, CAUSE $\log_2(\cdot)$ IS POSITIVE. $\square$

ACHILLE CANNAVALE

TYPICAL SET

LET $\{X_i\}_{i\in\mathbb{N}}$ i.i.d, $H(X)<\infty$

$$A_\varepsilon^{(n)} = \left\{ \underline{x} = (x_1,\dots,x_n) \in \mathcal{A}_x^n : \left| \frac{1}{n} \log_2\left(\frac{1}{P(\underline{x})}\right) - H(X) \right| < \varepsilon \right\}$$

PROP. 1

$$\mathbb{P}\left((x_1,\dots,x_n) \in A_\varepsilon^{(n)} \right) > 1-\varepsilon \quad \forall \varepsilon>0, \text{ FOR. } n \text{ SUFF. LARGE.}$$

PROOF. 1

SINCE $\mathbb{P}\left((x_1,\dots,x_n) \in A_\varepsilon^n \right) \xrightarrow[n\to\infty]{} 1 \Rightarrow \mathbb{P}\left((x_1,\dots,x_n) \in A_\varepsilon^n \right) > 1-\varepsilon$

$\forall \varepsilon>0$, FOR. n SUFF. LARGE

PROP. 2

$$\forall \varepsilon>0 \quad \text{CARD}\left(A_\varepsilon^{(n)}\right) \leq 2^{n(H(X)+\varepsilon)}$$

PROOF 2

FROM THE DEF. OF TYPICAL SET:

$$-\frac{1}{n}\log_2\left(P(\underline{x})\right) - H(\underline{x}) < \varepsilon \rightarrow P(\underline{x}) \geq 2^{-n(H(X)+\varepsilon)}$$

$$-\left(-\frac{1}{n}\log_2\left(P(\underline{x})\right) - H(\underline{x})\right) < \varepsilon \rightarrow P(\underline{x}) \leq 2^{-n(H(X)-\varepsilon)}$$

NOW WE CAN WRITE: $1 = \sum_{x\in\mathcal{A}_x^n} P(\underline{x}) \geq \sum_{x\in A_\varepsilon^{(n)}} P(\underline{x}) \geq \sum_{x\in A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} =$

$$= \text{CARD}\left(A_\varepsilon^{(n)}\right)\cdot 2^{-n(H(X)+\varepsilon)} \Rightarrow 1 \geq \text{CARD}\left(A_\varepsilon^{(n)}\right)\cdot 2^{-n(H(X)+\varepsilon)} \Rightarrow$$

$$\Rightarrow 2^{n(H(X)+\varepsilon)} \geq \text{CARD}\left(A_\varepsilon^{(n)}\right)$$

PROP. 3

$$\forall \varepsilon>0 \quad \text{CARD}\left(A_\varepsilon^{(n)}\right) > (1-\varepsilon)\cdot 2^{n(H(X)-\varepsilon)}, \text{ FOR. } n \text{ SUFF. LARGE}$$

PROOF. 3

FOR. n SUFF. LARGE $\mathbb{P}(\underline{x}\in A_\varepsilon^{(n)}) > 1-\varepsilon$, SO:

$$1-\varepsilon < P(\underline{x}\in A_\varepsilon^{(n)}) \overset{(2)}{\leq} \sum_{x\in A_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)} = \text{CARD}\left(A_\varepsilon^{(n)}\right)\cdot 2^{-n(H(X)-\varepsilon)} \Rightarrow$$

$$\Rightarrow \text{CARD}\left(A_\varepsilon^n\right) > (1-\varepsilon)\,2^{n(H(X)-\varepsilon)}, \; n \longrightarrow \infty$$

ACHILLE CANNAVALE

SOURCE CODE

A SOURCE CODE FOR A R.V. X IS AN APPLICATION:

$$C : \mathcal{A}_X \longrightarrow \mathcal{D}^*$$

$\mathcal{A}_X$: ALPHABET OF X

$\mathcal{D}^*$: SET OF ALL FINITE LENGTH STRING OF SYMBOLS TAKEN FROM THE D-ARY ALPHABET:

$$\mathcal{D} = \{1, \dots, D-1\}$$

NON-SINGULAR

A CODE IS NON-SINGULAR IF C IS INJECTIVE:

$$\text{IF } \forall x_1 \neq x_2 \Rightarrow C(x_1) \neq C(x_2)$$

EXTENDED CODE

THE EXTENSION C^* OF A CODE C IS THE APPLICATION:

$$C^* : \mathcal{A}_X^* \longrightarrow \mathcal{D}^* = C^*(x_1, \dots x_m) = C(x_1)\, C(x_2) \dots C(x_m)$$

$\mathcal{A}_X^*$: SET OF ALL FINITE-LENGTH SEQUENCES OF SYMBOLS TAKEN FROM $\mathcal{A}_X$.

UNIQUELY DECODABLE

A CODE C IS UNIQUELY DECODABLE IF C^* IS NON-SINGULAR.

PREFIX CODE

C IS A PREFIX CODE IF IT SATISFIES THE PREFIX CONDITION:

NO CODEWORD IS PREFIX OF ANY CODEWORD

ACHILLE CANNAVALE

PROVE THAT EXPECTED LENGTH OF ANY PREFIX CODE IS ALWAYS GREATER THAN ENTROPY

C IS A D-ARY PREFIX CODE FOR X R.V.

$$L = \mathbb{E}[\ell(x)] \geq H(x)$$

$= \iff P(x)$ IS D-ADIC

PROOF

$$L - H(x) = \mathbb{E}[\ell(x)] - \mathbb{E}\left[\log_D\left(\frac{1}{P(x)}\right)\right] = \mathbb{E}\left[\ell(x) - \log_D\left(\frac{1}{P(x)}\right)\right] =$$

$$= \mathbb{E}\left[-\log_D\left(D^{-\ell(x)}\right) - \log_D\left(\frac{1}{P(x)}\right)\right] = \mathbb{E}\left[\log_D\left(\frac{P(x)}{D^{-\ell(x)}}\right)\right] =$$

$$= \mathbb{E}\left[\log_D\left(\frac{P(x)}{\frac{D^{-\ell(x)}}{\sum_{y\in A_x} D^{-\ell(y)}}}\right)\right] + \mathbb{E}\left[\log_D\left(\frac{1}{\sum_{y\in A_x} D^{-\ell(y)}}\right)\right] =$$

$\cdot\, q(x)$ — IT'S JUST A NUMBER

$$= \mathbb{E}_P\left[\log_D\left(\frac{P(x)}{q(x)}\right)\right] + \log_D\left(\frac{1}{\sum_{y\in A_x} D^{-\ell(y)}}\right) =$$

$$= D_P(P(x)\,//\,q(x)) + \log_D\left(\frac{1}{\sum_{y\in A_x} D^{-\ell(y)}}\right)$$

≥ 0

PREFIX $\Rightarrow$ KRAFT $\Rightarrow \leq 1$

≥ 0

$$\Rightarrow L - H(x) \geq 0 \quad \square$$

BOUNDS TO THE AVERAGE LENGTH

C IS AN OPTIMAL CODE $\Rightarrow \quad H(x) \leq L^* \leq H(x) + 1$

PROOF

LET C BE THE SHANNON CODE:

$$\Rightarrow \ell(x) = \left\lceil \log_D\left(\frac{1}{P(x)}\right) \right\rceil \quad \text{SINCE } a \leq \lceil a \rceil \leq a+1$$

$$\Rightarrow \log_D\left(\frac{1}{P(x)}\right) \leq \ell(x) \leq \log_D\left(\frac{1}{P(x)}\right) + 1$$

$\mathbb{E}[\cdot] \Rightarrow \quad H(x) \leq L \leq H(x) + 1 \quad \square$

ACHILLE CANNAVALE

CHANNEL CODE

AN (M, n) CHANNEL CODE IS THE MAPPING:

$$\underline{x}_n : \{1, \dots, M\} \to \mathcal{A}_X^n$$

$$\{\underline{x}_n(1), \dots, \underline{x}_n(M)\} = \text{CODEBOOK}$$

$\underline{x}_n(M) \to$ CODEWORD

DECODING RULE

A DECODING RULE IS THE MAPPING:

$$g : \mathcal{A}_Y^n \longrightarrow \{1, \dots, M\}$$

$$\hat{W} = g(\underline{Y}_n) = \text{ESTIMATE OF } W$$

RATE

THE RATE R OF A (M, n) CHANNEL CODE IS:

$$R = \frac{\log_2(M)}{n} \left[\text{bits} / \text{CHANNEL USES} \right]$$

ACHIEVABLE RATE

THE RATE R IS ACHIEVABLE IF $\exists$ A SEQUENCE OF $(\lceil 2^{nR} \rceil, n)$ CHANNEL CODES:

$$P_n^{MAX}(e) \xrightarrow[n \to \infty]{} 0$$

ACHILLE CANNAVALE

CHANNEL CODING THEOREM

1) $R < C \Rightarrow \exists$ A SEQUENCE OF $\left(\lceil 2^{mR} \rceil, m\right)$ CHANNEL CODES: $P_m^{MAX}(e) \xrightarrow[m \to \infty]{} 0$

2) IF FOR A SEQUENCE OF $\left(\lceil 2^{mR} \rceil, m\right)$ CHANNEL CODES: $P_m^{MAX}(e) \xrightarrow[m \to \infty]{} 0 \Rightarrow R \leq C$

PROOF

LET $M = \lceil 2^{mR} \rceil \approx 2^{mR}$, THEN

1) RANDOM CODING

THE (M, m) CHANNEL CODE IS GENERATED AS FOLLOWS:

$$\begin{pmatrix} x_1(1) & \dots & x_m(1) \\ x_1(2) & \dots & x_m(2) \\ \vdots & & \\ x_1(M) & \dots & x_m(M) \end{pmatrix} = \begin{pmatrix} \underline{x}_m(1) \\ \underline{x}_m(2) \\ \vdots \\ \underline{x}_m(M) \end{pmatrix} = \text{CODEBOOK}$$

↳ THE ENTRIES OF THIS MATRIX ARE i.i.d. R.V. DRAWN FROM THE DISTRIBUTION $P^*(x) = \underset{P(x)}{\text{ARGMAX}}\ I(X;Y)$

THE CODE IS SHARED WITH THE DESTINATION.

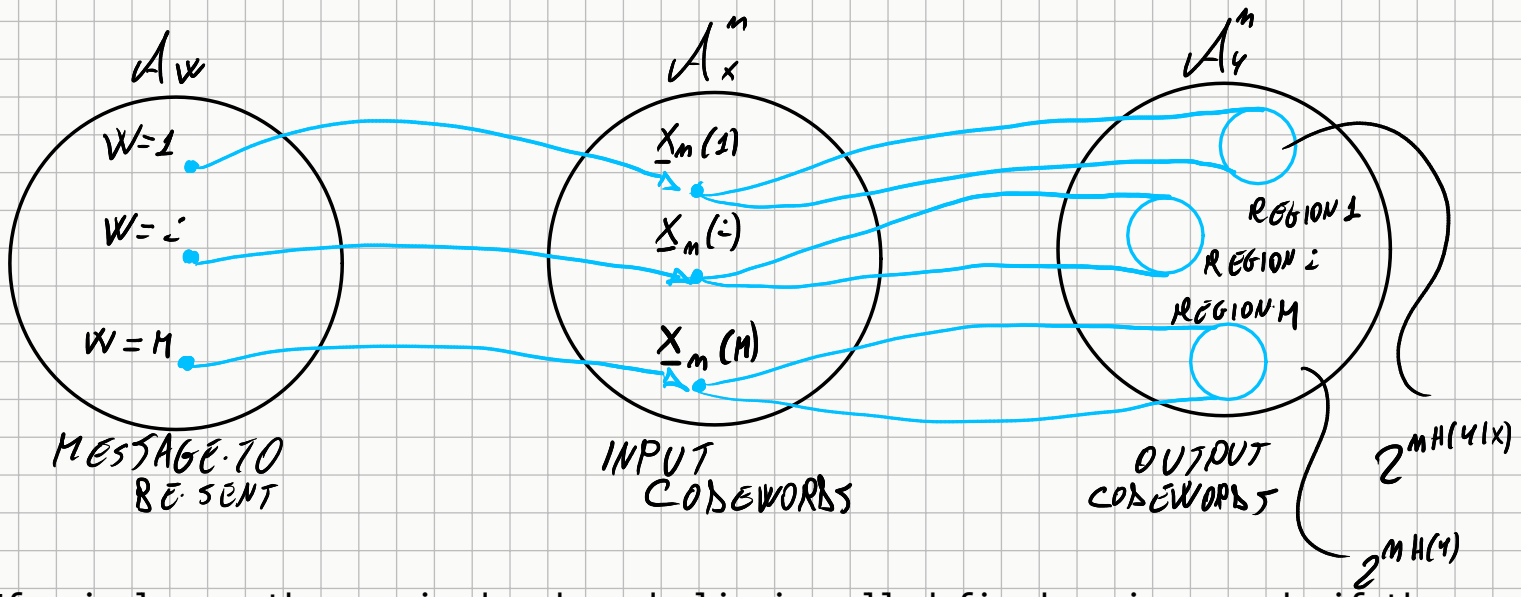
2) JOINTLY TYPICAL DECODING RULE

$\hat{W} = g(\underline{Y}_m) = i \Leftrightarrow \underline{X}_m(i)$ (THE CODEWORD ASSOCIATED TO $W = i$) IS THE ONLY SEQUENCE JOINTLY TYPICAL WITH $\underline{Y}_m$, i.e.

$$\begin{cases} (\underline{X}_m(i), \underline{Y}_m) \in A_\varepsilon^{(m)} \\ (\underline{X}_m(j), \underline{Y}_m) \notin A_\varepsilon^{(m)} \quad \forall j \neq i \end{cases}$$

ACHILLE CANNAVALE

3) SPHERE



If n is large, the received codewords lie in well defined regions, and, if these regions do not overlap, there is no error in decoding; this is possible only if the rate is sufficiently small.

$$\left(R \cdot \text{SMALL} \Rightarrow M \cdot \text{SMALL} \Rightarrow \text{FEW} \cdot \text{REGIONS}\right)$$

BUT:

- THE TOTAL NUMBER OF TYPICAL (RECEIVED) SEQUENCES IS $2^{mH(Y)}$
- $\forall\, \underline{X}_m$ TYPICAL, THERE ARE $2^{mH(Y|X)}$ $\underline{Y}_m$ TYPICAL.
- THE NUMBER OF REGIONS IS M.

IN ORDER TO HAVE NON-OVERLAPPED REGIONS, WE MUST HAVE:

$$M\, 2^{mH(Y|X)} \leq 2^{mH(Y)}$$

- M: # OF REGIONS
- $2^{mH(Y|X)}$: # OF CODEWORDS IN EACH REGION
- $2^{mH(Y)}$: TOTAL # OF CODEWORDS.

SINCE $M = 2^{mR}$

$$2^{mR}\, 2^{mH(Y|X)} \leq 2^{mH(Y)}$$

$$2^{mR} \leq 2^{m(H(Y) - H(Y|X))} = 2^{mI(X;Y)} = 2^{mC}$$

SINCE $P^*(X) = \underset{P(X)}{\text{ARGMAX}}\ I(X;Y)$

$$\Rightarrow R \leq C \quad \square$$

ACHILLE CANNAVALE

CAPACITY OF A GAUSSIAN CHANNEL WITH POWER CONSTRAINT

$$C = \frac{1}{2}\log_2\left(1+\frac{P}{N}\right)$$

ANS. CAN BE ACHEIVED WITH: $x \sim \mathcal{N}(0,P)$

PROOF

$$I(x;y) = h(y) - h(y|x) =$$

$h(y|x) = h(x+z|x) = h(z|x) = h(z) = \frac{1}{2}\log_2(2\pi e N)$ — z AND x ARE INDEP.; $z \sim \mathcal{N}(0,N)$

$$= h(y) - \frac{1}{2}\log_2(2\pi e N)$$

x, z INDEP

$$VAR(y) = VAR(x+z) = VAR(x) + VAR(z) = \underbrace{\mathbb{E}[x^2]}_{\leq P} - \underbrace{\left(\mathbb{E}[x]\right)^2}_{\geq 0} + N \leq P+N$$

$$\Rightarrow h(y) \leq \frac{1}{2}\log_2\left(2\pi e(P+N)\right)$$

"=" IF $y \sim \mathcal{N}(0,(P+N))$

IF $x \sim \mathcal{N}(0,P)$

$$\Rightarrow I(x;y) = h(y) - \frac{1}{2}\log_2(2\pi e N) \leq \frac{1}{2}\log_2(2\pi e(P+N)) - \frac{1}{2}\log_2(2\pi e N)$$

$$= \frac{1}{2}\log_2\left(\frac{P+N}{N}\right) = \frac{1}{2}\log_2\left(1+\frac{P}{N}\right)$$

$$\Rightarrow C = \max_{P_x \leq P}\{I(x;y)\} = \frac{1}{2}\log_2\left(1+\frac{P}{N}\right)$$

FOR $x \sim \mathcal{N}(0,P)$ □

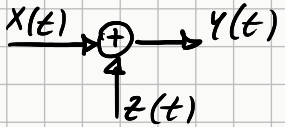
ACHILLE CANNAVALE

BAND-LIMITED GAUSSIAN CHANELL

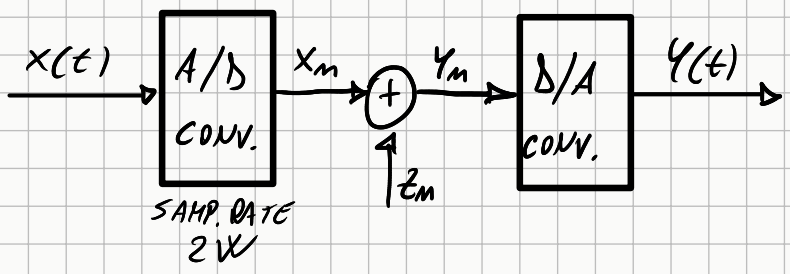
$$C = W \log_2\left(1 + \frac{P}{N_0 W}\right) \left[\text{bit/s}\right]$$

PROOF

THE CHANNEL:



HAS THE SAME CAPACITY AS



$$C = \frac{1}{2} \log_2\left(1 + \frac{P}{N_0 W}\right) \left[\text{bit/CH. US.}\right]$$

SINCE IT IS USED $2W$ TIMES PER SECOND WE HAVE:

$$C = 2W \cdot \underbrace{\frac{1}{2} \log_2\left(1 + \frac{P}{N_0 W}\right)}_{\left[\text{bit/CH. USE}\right]} \Longrightarrow$$

$\left[\text{CH USE/s}\right]$

$$\Longrightarrow W \cdot \log_2\left(1 + \frac{P}{N_0 W}\right) \left[\text{bit/s}\right] \quad \square$$

ACHILLE CANNAVALE