Frame

Information Theory and Coding

Basics

Sterling's Approximation

$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$

Approx. of Binomial Coeff

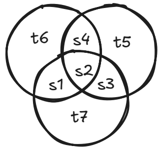
$\binom{N}{r} \approx 2^{NH_2(r/N)}$

Probability

Bayes' Law $P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$

Product Rule $P(A \cap B) = P(A)P(B \mid A)$

Hamming Coding



Source Coding Theorem

Statement $\frac{1}{N}H_\delta(X^N) < H(X) + \epsilon$

Typical Seqeunce

$T_\delta^{(N)} = \left\{x^N \in \mathcal{X}^N : \left| -\frac{1}{N}\log_2 P(x^N) - H(X)\right| < \beta\right\}$

Asymptotic Eqipartition Theorem

Probability $2^{-N(H(X)+\delta)} \leq \Pr(x^N) \leq 2^{-N(H(X)-\delta)}$

Count $|T_\delta^{(N)}| \leq 2^{N(H(X)+\delta)}$

Markov Inequality $\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$

Tschebyshev Inequality $\Pr(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$

Symbol Codes

Kraft's Inequality $\sum_{i=1}^{I} 2^{-l_i} \leq 1$

Expected Length $\bar{L} = \sum_{i=1}^{I} p_i l_i \geq H(X)$

Equality $l_i = -\log_2 p_i \quad \forall i$

Eg: Huffman Code

Stream Codes

1. Arithmetic codes — Compression close to entropy
2. Lempel Ziv — Doesn't require a model
3. Burrow's Wheeler — Improves Lempel Ziv

Measure of Information

Entropy $H(X) = -\sum_{i=1}^{n} P(A_i)\log_2 P(A_i)$

Redundancy $R_{\text{norm}} = 1 - \frac{H(X)}{H_{\max}}$

Joint Entropy $H(X,Y) = -\sum_x \sum_y P(x,y)\log_2 P(x,y)$

Conditional Entropy $H(Y \mid X) = -\sum_x P(x) \sum_y P(y \mid x)\log_2 P(y \mid x)$

Differential Entropy $h(X) = -\int f(x)\log f(x)\,dx$

Mutual Information $I(X;Y) = \sum_x \sum_y p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$

$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$

Conditional Mutual Information $I(X;Y \mid Z) = \sum_z p(z) \sum_x \sum_y p(x,y \mid z)\log\frac{p(x,y \mid z)}{p(x \mid z)p(y \mid z)}$

$I(X;Y \mid Z) = H(X \mid Z) - H(X \mid Y,Z) = H(Y \mid Z) - H(Y \mid X,Z)$

Chain Rule $H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y) = H(X) + H(Y) - H(X,Y)$

Chain rule $I(X;Y,Z) = I(X;Y) + I(X;Z \mid Y)$

Decomposing Entropy $H(p_1, p_2, \ldots, p_I) = H_b(p_1) + (1-p_1)H\left(\frac{p_2}{1-p_1}, \ldots, \frac{p_I}{1-p_1}\right)$

Entropy Gaussian PDF $h(X) = \frac{1}{2}\log_2(2\pi e\sigma^2)$

Some other points

Symmetric Channel

- Each input row in the transition matrix is a rearrangement of the same probabilities.
-All inputs are treated equally.

Shannon Information

1. Ideal codeword length = Information Content
2. Average information content = Entropy

Channel Capacity

1. Entropy is maximised for uniform distribution discrete random variable.

Continuous Distribution

1. Mutual Information is maximised for Gaussian distribution under $\mathbb{E}[x^2] \leq \sigma_x^2$

$h(X) \leq \frac{1}{2}\log(2\pi e\,\sigma_x), \quad \text{equality iff } X \sim \mathcal{N}(\mu, \sigma_x^2)$

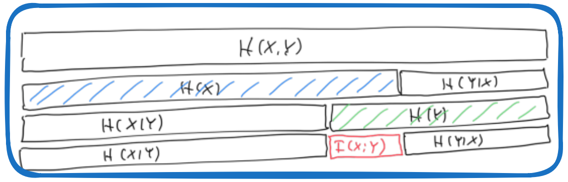
Your own points

Jensen's Inequality $\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$

Gibb's Inequality $D_{\text{KL}}(P \parallel Q) = \sum_x P(x)\log_2\frac{P(x)}{Q(x)} \geq 0$

Lempel-Ziv

s(n) (Index)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Bits Needed	1	1	2	2	3	3	3	3	4	4	4	4	4	4	4	4
(Pointer, New Bit)	(,1)	(0,0)	(01,1)	(10,1)	(100,0)	(101,0)	(110,0)	(111,0)	(1000,1)	(1001,0)	(1010,1)	(1011,0)	(1100,1)	(1101,0)	(1110,1)	(1111,0)
Pointer Binary Addr	0	0	01	10	100	101	110	111	1000	1001	1010	1011	1100	1101	1110	1111



Data Processing

If X -> Y -> Z is Markov Chain

$p(z \mid x, y) = p(z \mid y)$

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

Fano's Inequality

$H(W \mid \hat{W}) \leq h_2(P_e) + P_e\log(|\mathcal{W}| - 1)$

Shannon Capacity Theorem

Statement $C = \max_{P_X(x)} I(X;Y)$

Proof

1. Achievablilty: Typical sets don't collide if R < I(X;Y)
2. Converse: Fano + data processing = no rate beyond capacity

BSC	$1 - H_2(\epsilon)$	Rate(BER) $R \approx \frac{C}{1 - H_2(\text{BER})}$
BEC	$1 - \epsilon$	
Typewriter	$\log_2\frac{\lvert A_X\rvert}{L}$	

Message passing

1. Distributed & efficient
2. Sum Product makes marginalisation efficient

Perfect code

1. No free space in between codes
2. Not a good solution as it wastes lots of code words

LDPC

1. Achieves data rates close to capacity
2. Sum product can be used to decode effeciently