MA4261 Information and Coding Theory

Owen Leong
owenleong@u.nus.edu

November 26, 2024

1 Probability

$$E[Y] = E[E[Y|X]]$$
$$Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$$

Union bound $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$
Markov chain $X - Y - Z$ has

$$P_{XYZ}(x, y, z) = P_X(x) P_{Y|X}(y|x) P_{Z|Y}(z|y)$$
$$P_{XZ|Y}(x, z|y) = P_{X|Y}(x|y) P_{Z|Y}(z|y)$$

Markov's inequality $\Pr(X > a) \leq \frac{\mathbb{E}[X]}{a}$
Chebyshev's $\Pr(|X - \mu| > a\sigma) \leq \frac{1}{a^2}$
WLLN $\lim_{n\to\infty} \Pr\left(\left|\frac{1}{n}\sum_{i=1}^n X_i\right| > \epsilon\right) = 0$

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$$

CDF $\Phi(y) = \int_{-\infty}^{y} \mathcal{N}(x; 0, 1) \mathrm{d}x$
CLT $\lim_{n\to\infty} \Pr(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n X_i < a) = \Phi(a)$
Jensen's inequality $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$

2 Information Quantities

$$\mathbf{H(X)} = \sum_{x\in\mathscr{X}} p(x) \log \frac{1}{p(x)} \text{ bits for R.V. } X$$
$$0 \leq H(X) = \mathbb{E}\left[1/\log p_X(X)\right] \leq \log|\mathscr{X}|$$
$$H(X, Y) = \mathbb{E}_{P_{X,Y}}\left[\log \frac{1}{p_{X,Y}(X, Y)}\right]$$
$$H(Y|X) = \sum p(x) H(Y|X = x)$$
$$H(Y|X) = \mathbb{E}_{P_{X,Y}}\left[\log \frac{1}{p_{Y|X}(Y|X)}\right]$$
$$H(Y|X = x) = -\sum_y p(y|x) \log p(y|x)$$
$$H(X, Y) = H(X) + H(Y|X)$$
$$H(X, Y|Z) = H(X|Z) + H(Y|X, Z)$$
$$\mathbf{D(p \parallel q)} = \sum p(x) \log[p(x)/q(x)]$$

with $0 \log \frac{0}{0} = 0 \log \frac{0}{q} = 0, p \log \frac{p}{0} = +\infty$

$\mathbf{I(X;Y)} = D(p_{X,Y} \parallel p_X p_Y) = H(X) - H(X|Y)$

2.1 Chain Rule(s)

$$H(X_1, ... X_n) = \sum_{i=1}^n H(X_i|X_1, ... X_{i-1})$$
$$I(X_1, ..., X_n; Y) = \sum_{i=1}^n I(X_i; Y|X_1, ..., X_{i-1})$$
$$D(p_{X,Y} \parallel q_{X,Y}) = D(p_X \parallel q_X) + D(p_{X|Y} \parallel q_{X|Y}|p_X)$$
$$D(p_{Y|X} \parallel q_{Y|X}|p_X) = \sum p(x) D(p_{Y|x} \parallel q_{Y|x})$$

2.2 Information Inequalities

$$D(p \parallel q) \geq 0, I(X; Y) \geq 0$$

$H(X_1, ..., X_n) \leq \sum H(X_i)$ with equality when X_i's are mutually independent
Log-sum inequality for non-negative a_i, b_i

$$\sum \left(a_i \log \frac{a_i}{b_i}\right) \geq a \log \frac{a}{b} = (\sum a_i) \log \frac{\sum a_i}{\sum b_i}$$
$$D(\lambda p_1 + (1 - \lambda)p_2 \parallel \lambda q_1 + (1 - \lambda)q_2)$$
$$\leq \lambda D(p_1 \parallel q_1) + (1 - \lambda) D(p_2 \parallel q_2) \text{ convex}$$
$$H(\lambda p + (1-\lambda) q) \geq \lambda H(p) + (1-\lambda) H(q) \text{ concave}$$

Fix $p_{Y|X}$, $p_X \to I(p_X, p_{Y|X})$ is concave
Fix p_X, $p_{Y|X} \to I(p_X, p_{Y|X})$ is convex

2.2.1 Data Processing Ineq. for M.I.

If $X - Y - Z$, then $I(X; Y) \geq I(X; Z)$ with $=$ iff $X - Z - Y$, i.e. $I(X; Y) \geq I(X; g(Y))$

2.2.2 Fano's Inequality

$$H_b(P_e) + P_e \log|\mathscr{X}| \geq H(X|\hat{X}) \geq H(X|Y)$$
$$P_e = \Pr(\hat{X} \neq X) \geq \frac{H(X|Y) - 1}{\log|\mathscr{X}|}$$

Han's: $H(X^n) \leq \frac{1}{n-1} \sum_{i=1}^n H(X^{n\setminus i})$
Shearer's: If $S \subseteq [n]$ is random following a distribution $\forall i \in [n], \Pr(i \in S) \geq \mu$, then

$$E_S[H(X_S)] \geq \mu H(X^n), X_S = \{X_i : i \in S\}$$

3 Asymptotic Equipartition Prop

ϵ-weakly typical set of $X \sim p(x)$ is

$$A_\epsilon^{(n)}(X) = \left\{x^n : \left|\frac{1}{n} \log \frac{1}{p(x^n)} - H(X)\right| < \epsilon\right\}$$
$$2^{-n(H(X)+\epsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\epsilon)}$$

$\Pr(X^n \in A_\epsilon^{(n)}(X)) \geq 1 - \epsilon$ for suff large n

$$(1 - \epsilon) 2^{n(H(X)-\epsilon)} \leq |A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$$

4 Source Coding

If $R^*(x) = \inf\{R \geq 0 : R \text{ is achievable}\}$, then $R^*(X) = H(X)$. Prove that $R^*(X) \leq H(X)$ using AEP, prove that $R^*(X) \geq H(X)$ using Fano's inequality.

4.1 Han Verdu Lemma

Fix $(n, 2^{nR})$-code, then $P_e = \Pr(X^n \neq \hat{X}^n)$

$$P_e \geq \sup_{\gamma > 0} \left\{\Pr\left(\frac{1}{n} \log \frac{1}{p(X^n)} \geq R + \gamma\right) - 2^{-n\gamma}\right\}$$

5 Stochastic Processes

Stationary means $\Pr(X_1^n) = \Pr(X_{1+l}^{n+l})$
Markov $\Pr(X_{n+1}|X_1^n) = \Pr(X_{n+1}|X_n)$
Time invariant $p(x_{n+1}|x_n)$ indep of n.
Irreducible: possible to go from any state to any other statein a finite number of steps
Aperiodic: GCD of the lengths of the paths from a state to itself is 1
Entropy rate of stochastic process is

$$H(X) = \lim_{n\to\infty} \frac{1}{n} H(X_1, ..., X_n)$$
$$H'(X) = \lim_{n\to\infty} H(X_n|X_1^{n-1})$$

For stationary process, $H(X) = H'(X)$
For stationary ergodic process,

$$\lim_{n\to\infty} -\frac{1}{n} \log p(X_1, ..., X_n) = H(X)$$

For Markov chain,

$$H(X) = H(X_2|X_1) = \sum_{i,j} -\mu_i P_{ij} \log P_{ij}$$

Hidden Markov Model is has Y_i fn of X_i with X_i's forming a Markov chain.

$$H(Y_n|Y_1^{n-1}, X_1) \leq H'(Y) \leq H(Y_n|Y_1^{n-1})$$

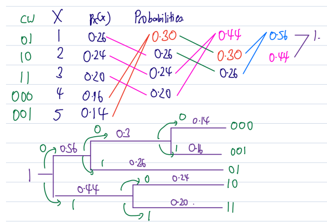
with convergence as $n \to \infty$.

6 Fixed-to-variable

Non-singular if each x mapped to a diff CW
Extension C^* of a code C is the map $C^*(x_1 \cdots x_n) = C(x_1) \cdots C(x_n)$
Code is uniquely decodable if its extension is non-singular
Code is prefix-free if no codeword is a prefix of any other codeword
Kraft's inequality $\sum 2^{-l_i} \leq 1$
Expected codeword length $L^* \geq H(X)$
Shannon code sets $l_i = \lceil \log \frac{1}{p_i} \rceil$, obtains $H(X) \leq L^* < H(X) + 1$

6.1 Huffman Codes

Given $p_1 \geq p_2 \geq p_3 \geq \cdots p_M$, $p_i > p_j \Rightarrow l_i \leq l_j$, and exists some optimal code where $C(M-1)$ and $C(M)$ are siblings, same length and differ only in last bit



7 Channels

$M = 2^{nR}$ is max number of distinguishable messages reliably sent through the channel rate R, the max rate C is channel capacity
Discrete channel has $(\mathscr{X}, \mathscr{Y}, p_{Y|X})$, (finite) input & output alphabet, and transition probabilities
Memoryless chan $\Pr(y^n|x^n) = \prod p(y_i|x_i)$
$C = C(p_{Y|X}) = \max_{p_x} I(X; Y)$
Noiseless BC $\mathscr{X} = \mathscr{Y} = \{0, 1\}$, $p = I_2$
Noisy Typewriter $C = \log 13$ use alternate $p(i|i) = 1/2, p(i + 1(mod\ 26)|i) = 1/2$
BSC flips bits with prob p, $C = 1 - H_b(p)$
$p_{Y|X}(y|x) = p$ if $y \neq x$ else $1 - p$
Binary erasure channel, $\{0, 1\} \to \{0, e, 1\}$

$$p_{Y|X} = \begin{bmatrix} 1 - \alpha & \alpha & 0 \\ 0 & \alpha & 1 - \alpha \end{bmatrix}, C = 1 - \alpha$$

Symmetric channels if rows and columns are permutations of each other
Weakly symmetric channels if column sums are same and rows are permutations
$C = \log|\mathscr{Y}| - H(\underline{r})$

7.1 Channel Coding

(M,n)-code for DMC $(\mathscr{X}, p_{Y|X}, \mathscr{Y})$ consists of message set $[m]$, encoder set $f : [m] \to \mathscr{X}^n$ and decoder $\varphi : \mathscr{Y}^n \to [m]$. M is size, n is block length.

$$\lambda_w = \Pr(\varphi(Y^n) \neq w | X^n = x^n(w)) = \sum_{y^n} p(y^n|x^n(w)) \mathbb{1}[\varphi(y^n) \neq w]$$

$$P_e^{(n)} = \lambda_{ave}^{(n)} = \frac{\sum_{w\in[M]} \lambda_w}{M}, \lambda_{max}^{(n)} = \max_{w\in[M]} \lambda_w$$

7.2 Jointly Typical Sequences

$A_\epsilon^{(n)} \subseteq \{(x^n, y^n) \in \mathscr{X}^n \times \mathscr{Y}^n\}$ such that

$$\left|-\frac{\log p(x^n)}{n} - H(X)\right| < \epsilon, \left|-\frac{\log p(y^n)}{n} - H(Y)\right| < \epsilon$$

$$|-\log p(x^n, y^n)/n - H(X,Y)| < \epsilon$$

where $p(x^n, y^n) = \prod p(x_i, y_i)$

$$\exists N : \forall n > N, \Pr((X^n, Y^n) \notin A_\epsilon^{(n)}) < \epsilon$$

$$|A_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+\epsilon)}$$

If $\tilde{X}^n \perp\!\!\!\perp \tilde{Y}^n$,

$$\Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)) \leq 2^{-n(I(X;Y)-3\epsilon)}$$

$$\Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)) \geq (1-\epsilon)2^{-n(I(X;Y)+3\epsilon)}$$

7.3 Proof of Achievability

Fix $p(x)$, generate codebook

$$C = \begin{bmatrix} x_1(1) & \cdots & x_n(1) \\ x_1(2) & \cdots & x_n(2) \\ & \vdots & \\ x_1(2^{nR}) & \cdots & x_n(2^{nR}) \end{bmatrix} \sim p_X(x)$$

Encoder: given w, send $x^n(w)$

Decoder: given y^n, declare $\hat{w}$ is sent if $(x^n(\hat{w}), y^n) \in A_\epsilon^{(n)}$ and no other $w' \in [2^{nR}]$ satisfies $(x^n(w'), y^n) \in A_\epsilon^{(n)}$
Given $w = 1$ was sent, the error scenarios are either the received $(x^n(1), y^n) \notin A_\epsilon^{(n)}$, or that other $(x^n(i), y^n) \in A_\epsilon^{(n)}$ for $i \geq 2$.
$E_w = \{(X^n(w), Y^n) \in A_\epsilon^{(n)}\}$

$$\Pr(E|W_1) \leq \Pr(E_1^c|W_1) + \sum_{w=2}^{2^{nR}} \Pr(E_w|W_1)$$

$$\Pr(E_1^c|W_1) = \Pr((X^n(1), Y^n) \notin A_\epsilon^{(n)}) < \frac{\epsilon}{4}$$

$$\Pr(E_w|W_1) = \Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}) \leq 2^{-n(I-3\epsilon)}$$

$$\Pr(E|W_1) \leq \frac{\epsilon}{4} + 2^{nR} 2^{-n(I-3\epsilon)} = \frac{\epsilon}{4} + 2^{-n(I-R-3\epsilon)}$$

Take p_X that maximizes $I(X;Y)$, and take $R < C - 3\epsilon$ so that $2^{-n(I(X;Y)-R-3\epsilon)} < \frac{\epsilon}{4}$, so $\Pr(E|W=1) < \frac{\epsilon}{2}$.
$\implies \exists$ a code C^* with rate R and average error prob $< \frac{\epsilon}{2}$

$$E_C[\lambda_{ave}^{(n)}(C)] < \frac{\epsilon}{2} \implies \exists C^* : \lambda_{ave}^{(n)}(C^*) < \frac{\epsilon}{2}$$

To get bound on λ_{max}, take only the better half of the codebook, new size of $\tilde{C}^* = \frac{2^{nR}}{2} = 2^{n(R-\frac{1}{n})}$ and $\max_w \lambda_w(\tilde{C}^*) < \epsilon$
$\implies \exists$ a code $\tilde{C}^*$ of rate $R - \frac{1}{n}$ with max error prob $< \epsilon$.

7.4 Proof of Converse

$$\Pr(W \neq \hat{W}) \geq \frac{H(W|\hat{W}) - 1}{\log|W|}$$

$$\implies H(W|\hat{W}) \leq P_e^{(n)} \cdot nR + 1$$

$$\begin{aligned} nR &= H(W) = I(W;\hat{W}) + H(W|\hat{W}) \\ &\leq I(X^n;Y^n) + P_e^{(n)} nR + 1 \\ &= nC + nRP_e^{(n)} + 1 \\ R &\leq \frac{1}{1-P_e^{(n)}} C + \frac{1}{n} \end{aligned}$$

As $n \to \infty$, $P_e^{(n)} \to 0$, $R \leq C$.

8 Differential Entropy

$h(X) = -\int_S f(x) \log f(x)\, dx$
For $X \sim \text{Uniform}(0,a)$, $h(X) = \log a$
For $X \sim \phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2}{2\sigma^2})$, $h(\phi) = \frac{1}{2}\log(2\pi e \sigma^2)$ bits
$\max_{EX^2=\alpha} h(X) = \frac{1}{2}\log(2\pi e \alpha)$

9 Gaussian Channels

$$Y_i = X_i + Z_i,\ Z_i \sim \mathcal{N}(0,N)$$

Require $\frac{1}{n}\sum_{i=1}^n x_i^2 \leq P$.

$$C = \max_{E[X^2]\leq P} I(X;Y) = \frac{1}{2}\log(1 + \frac{P}{N})$$

achieving max when $X \sim \mathcal{N}(0,P)$
For parallel guassian channels each with N_j, Water filling theorem, $P_i = (v - N_i)^+, \sum(v-N_i)^+ = P$

10 Finite Fields

Group has elements G and op $\oplus$ satisfying closure, associativity, identity and inverse.
Alternatively, satisfy associativity, identity and permutation property $a \oplus G$ is a permutation of G
Fields is a set $\mathbb{F}$ of ≥ 2 elements with operations $\oplus$ and $*$ such that $\mathbb{F}$ forms an abelian group under $\oplus$ and $\mathbb{F}\setminus\{0\}$ forms an abelian group under $*$ and $(a\oplus b)*c = (a*c)\oplus(b*c)$
$\forall$ prime p, $R_p = \{0, \cdots, p-1\}$ forms a field under mod-p addition and multiplication
Polynomial $g(x)$ divides $f(x)$ if $f(x) = q(x)g(x)$ for some polynomial $q(x)$. $g(x)$ is a factor of $f(x)$ if $g(x)$ is monic and a nontrivial divisor of $f(x)$
Irreducible polynomials have no factors. A monic irreducible polynomial is a prime polynomial.
To construct field with p^m elements, take $\mathbb{F}_{g(x)} = \{r_0 + r_1 x + \cdots + r_{m-1}x^{m-1} : r_i \in \mathbb{F}_p\}$ with polynomial addition and multiplication mod-$g(x)$ where $\deg(g(x)) = m$ and $g(x)$ is a prime polynomial.

11 Codes

(n,M,d)-code, codewords $C \subset \mathbb{F}_q^n$, $|C| = M$, $d = \min_{c\neq c'\in C} d(c,c')$, $R = \frac{\log_q M}{n}$
$[n,k,d]$-linear code has codewords forming a vector space with dim k, $M = q^k$, $R = \frac{k}{N}$
Dual code of C is $C^\perp = \{x : \langle x, c\rangle = 0\}$
Hamming weight $wt(x) = d(x,0)$, $wt(C) = \min_{c\in C, c\neq 0} wt(c)$
Generator Matrix $G \in \mathbb{F}^{k\times n}$ for linear code has rows formed by basis for C, standard form $[I_k|\mathbb{F}_q^{k\times(n-k)}]$, every codeword expressed as some vG
Parity-check Matrix $H \in \mathbb{F}^{(n-k)\times n}$ is generator matrix for $C^\perp$, standard form $[\mathbb{F}_q^{(n-k)\times k}|I_k]$, $HG^T = 0$
$d(C) \geq d$ iff $\forall$ subsets $d-1$ cols of H are LI
$d(C) \leq d$ iff $\exists$ a subset of d cols that is L.D.
Relative dist $\delta(C) = (d-1)/n$

11.1 Performance Bounds

$A_q(n,d) = \max\{M : \exists(n,M,d)\text{-code}\}$
$B_q(n,d) = \max\{q^k : \exists[n,k,d]\text{-LC over } \mathbb{F}_q\}$
Sphere volume $V_q^n(r) = \sum_{i=0}^n \binom{n}{i}(q-1)^i$ if $0 \leq r \leq n$ else q^n if $r > n$
Gilbert-Varshamov sphere-covering lower bound $A_q(n,d) \geq \frac{q^n}{V_q^n(d-1)}$
Sphere-packing upper bound Hamming bound $A_q(n,d) \leq \frac{q^n}{V_q^n(\lfloor\frac{d-1}{2}\rfloor)}$
Perfect code achieves hamming bound
Singleton bound $A_q(n,d) \leq q^{n-d+1}$ or $k \leq n-d+1$
MDS codes achieve singleton bound

11.2 Reed-Solomon Codes

Choose n eval points $\alpha_1, \cdots, \alpha_n \in \mathbb{F}_q$
Message $m = (m_0, \cdots, m_{k-1}) \in \mathbb{F}_q^k$
Define polynomial $C_m(x) = \sum_{i=0}^{k-1} m_i x^i$
Encode $RS(m) = (C_m(\alpha_1), \cdots, C_m(\alpha_n))$
For $m \neq m'$, at most $k-1$ evaluation points where $C_m(\alpha_i) = C_{m'}(\alpha_i)$
Thus $d(RS(m), RS(m')) \leq n - (k-1)$
Decoding using Berlekamp-Welch algorithm $\lfloor\frac{n-k}{2}\rfloor$ errors