ENEE626, CMSC858B, AMSC698B

Error Correcting Codes

Part III. Random coding; Coding theorems

ENEE626 Lecture 21: Bounds on codes

Plan:

1. Volume bounds: GV bound, Hamming bound
2. Bassalygo-Elias bound

Let $S_t(\mathbf{y})=S_{n,t}(\mathbf{y})\triangleq\{\mathbf{x}\in\{0,1\}^n: d(\mathbf{x},\mathbf{y})\leq t\}$, $S_t=|S_t|$

Theorem 21.1 (Gilbert bound): Let M, d be such that

$$M\, S_{d-1} < 2^n.$$

Then there exists an (n,M+1,d) binary code.

Proof: Greedy algorithm. Take any point $x_1\in\{0,1\}^n$.
Suppose that $C_i=(\mathbf{x}_1,\ldots,\mathbf{x}_i)$, $i\leq M$. By assumption, there exists a point

$$\mathbf{x}_{i+1}\notin \cup_{j=1}^{i} S_t(\mathbf{x}_j)$$ ▲

A somewhat stronger statement is given in

Theorem 21.2 (Varshamov bound): Let n,k,d be such that

$$S_{n-1,d-2}<2^{n-k}$$

Then there exists a linear [n,k,$\geq$d] binary code C.

Proof: Construct a parity-check matrix of C recursively. In the i^{th} step, $i\leq$ n-1 assume that we have an (n-k) x (i-1) matrix

no d-1 or fewer columns of which are linearly dependent (A)

We can add one more column so that the new matrix satisfies (A) if there is a column that is not spanned by any d-2 or fewer of the existing (i-1) columns. This is possible as long as $S_{i-1,d-2}<2^{n-k}$ holds true ▲

Let A(n,d)=max |C| : C is a binary code of length n and distance d

By the Gilbert bound, $A(n,d) \geq 2^n/S_{d-1}$

Hamming bound:

$$A(n,d) \leq 2^n/S_{[(d-1)/2]}$$

Plotkin bound: $A(n,d) \leq 2d/(2d-n)$, $d>n/2$ (proved in Lect. 6)

Theorem 21.3 (Bassalygo-Elias bound):

$$A(n,d) \leq 2^n \min_w \frac{1}{\binom{n}{w}} \frac{dn}{dn-2wn+2w^2}$$ for d>2w(1-w/n)

Proof: Let $A(n,d,w)=\max C \in \{0,1\}^n$, $d(C)=d$, $\forall_{\mathbf{x}\in C}\ w(\mathbf{x})=w$

Lemma 21.4 (Johnson bound): $A(n,d,w) \leq dn/(dn-2wn+2w^2)$

Proof: Let C be a constant-weight code of size M.

Let λ_i = # 0's in the ith column of the n x M matrix of its codewords.

$$\sum_{i=1}^n \lambda_i = M(n-w);\ \sum \lambda_i^2 \geq (1/n)\left(\sum_{i=1}^n \lambda_i\right)^2 \geq M^2(n-w)^2/n$$

Let $S \triangleq \sum_{c_1,c_2 \in C} d(c_1,c_2)$. We have

$M(M-1)d \leq S = 2\sum\lambda_i(M-\lambda_i) \leq 2M^2(n-w) - 2M^2(n-w)^2/n$; solve for M. ▲

Lemma 21.5: Let $C, Y \subset \mathbb{F}_2^n$. Then

$$|Y||C| = \sum_{\mathbf{x} \in \mathbb{F}_2^n} |(\mathbf{x}+C) \cap Y|.$$

Indeed,

$$\sum_{\mathbf{x}} \sum_{\mathbf{c} \in C} \sum_{\mathbf{y} \in Y} 1(\mathbf{x}+\mathbf{c}=\mathbf{y}) = \sum_{\mathbf{c}} \sum_{\mathbf{y}} 1 = |C||Y|,$$ proving the claim.

Now let $Y = S_w(0)$, then $(x+C) \cap Y$ is a code of constant weight w

$$| (x+C) \cap Y | \leq A(n,d,w)$$

Next replace the l.-h.side by the above lemma:

$$\binom{n}{w} A(n,d) \leq 2^n A(n,d,w)$$

To complete the proof of the B.-E. bound, use the Johnson bound in this inequality. ▲

ENEE626 Lecture 22: Bounds on codes

Plan:
Asymptotics of binomial coefficients
Asymptotic bounds on codes

Asymptotics of binomial coefficients

We would like to compute the asymptotic volume of the sphere $S_{n,l}$={x in $\mathbb{F}_2^n$: wt(x) $\leq$ 0} as n$\to\infty$, l=λ n, λ<1/2.

$$S_{n,l} = \sum_{i=0}^{l} \binom{n}{i}$$

Let $h(\lambda)=-\lambda \log_2\lambda-(1-\lambda)\log_2(1-\lambda)$ $0\leq\lambda\leq 1$

$$2^{-n\,h(\lambda)}=\lambda^{n\lambda}(1-\lambda)^{n(1-\lambda)}$$

Note that

$$\binom{n}{i}\lambda^i(1-\lambda)^{n-i}$$

is maximum for i=λ n. Then

$$1 = \sum_{i=0}^{n} \binom{n}{i}\lambda^i(1-\lambda)^{n-i} \leq (n+1)\binom{n}{l}\lambda^l(1-\lambda)^{n-l} = (n+1)\binom{n}{l}2^{-nh(\lambda)}$$

so

$$S_{n,l} \geq \binom{n}{l} \geq \frac{1}{n+1}2^{nh(\lambda)}$$

$$1 = (\lambda + (1-\lambda))^n \geq \sum_{i=0}^{\lambda n} \binom{n}{i} \lambda^i (1-\lambda)^{n-i}$$

$$\geq \sum_{i=0}^{\lambda n} \binom{n}{i} (1-\lambda)^n \left(\frac{\lambda}{1-\lambda}\right)^{\lambda n} = 2^{-nh(\lambda)} S_{n,l}$$

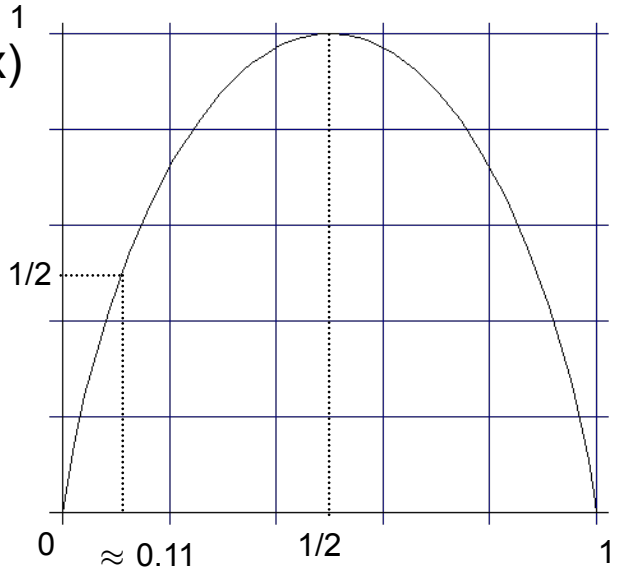
Theorem 22.1:

$$\frac{1}{n+1} 2^{nh(\lambda)} \leq \sum_{i=0}^{l} \binom{n}{i} \leq 2^{nh(\lambda)}$$

$$\lim_{n\to\infty} \frac{1}{n} \log \sum_{i=0}^{l} \binom{n}{i} = \lim_{n\to\infty} \frac{1}{n} \log \binom{n}{l} = h(\lambda) \quad (n \to \infty, l/n \to \lambda)$$

Properties of the binary entropy function h(x)

- h(x) defined for x∈[0,1]
- h(0)=0, h(0.11)≈ ½, h(0.215)≈ ¾, h(½)=1
- h(x)=h(1-x)
- h(x) ∼ -x log x (x→ 0)
- h(½-x) = 1-(2/ln 2)x^2 +O(x^4) (x→ 0)



There are about as many vectors of weight w=n/2 as *all* vectors

$$\binom{n}{n/2} \sim \frac{1}{\sqrt{n\pi/2}} 2^n$$

$$\sum_{i=0}^{\lambda n} \binom{n}{i} \cong 2^n \quad (\lambda > 1/2)$$

Example: n=1000

The number of vectors of weight w=500 is $10^{299.4}$ vs. the total $2^{1000}=10^{301}$

#x of weight $499 \leq w \leq 501$: 10^{300}
#x of weight $495 \leq w \leq 505$: $10^{300.4}$

Remark: in general, $\#\{x \in \mathbb{F}_2^n, |wt(x)-(1/2)n| \leq t\}/2^n \sim 2\Phi(t\sqrt{n/4})-1$,
where Φ is cdf $\mathcal{N}(0,1)$
for instance, t=5: $2\Phi(5/\sqrt{250})-1=2\Phi(0.316)-1 \approx 0.25$, very close to $10^{-0.6}$

Asymptotic bounds on codes

Let $R(\delta) = \limsup_{n \to \infty} n^{-1} \log A(n, \delta n)$

where $A(n,d) = \max_{C \in \mathbb{F}_2^n,\ C \text{ is a } d\text{-code}} |C|$

$R = (1/n) \log |C|$ code rate

$\delta = d(C)/n$ relative distance

Gilbert-Varshamov bound : $M S_{n,d-1} \geq 2^n$ (or $S_{n-1,d-2} \geq 2^{n-\kappa}$)

$$Rn + \log S_{n,d-1} \geq n$$

$$R + h(\delta) \geq 1 \qquad (0 \leq R \leq 1,\ 0 \leq \delta \leq \tfrac{1}{2})$$

$R(\delta) \geq 1 - h(\delta)$ $(0 \leq \delta \leq \tfrac{1}{2})$

valid even for the lim inf definition of $R(\delta)$

Hamming bound: $R(\delta) \leq 1 - h(\delta/2)$ $(0 \leq \delta \leq 1)$

Asymptotic version of the Plotkin bound (see next page)

$$R(\delta) \leq 1 - 2\delta$$

Plotkin bound implies that $R(\delta)=0$ for $\delta \geq ½$.

Lemma 22.2: $A(n,d) \leq 2^t A(n-t,d), \quad t \leq n-d$

Corollary: $R(\delta) \leq \tau + (1-\tau)R(\delta/(1-\tau))$

Proof: Let C be a d-code of size M such that M=A(n,d). Consider the Mxn matrix of codewords. The last coordinate contains at least M/2 1's or 0's, say 1's. Take only those vectors, puncture them on the nth coordinate. This gives a $(n-1, \geq M/2, \geq d)$ code. We can repeat this $t=\tau n$ times as long as $t \leq n-d$, obtain a $(n-t, \geq M/2^t, \geq d)$ code C'. We have

$$A(n,d)/2^t \leq |C'| \leq A(n-t,d)$$ ▲

Proof of Corollary: $(1/n) \log (M/2^t) = R(\delta)-\tau \leq (1/n) \log A(n-t,d) = (1-\tau)R(\delta/(1-\tau))$

The Plotkin bound: for any (n',M',d') code,

$$M' \leq 2d'/(2d'-n').$$

Let C be a d-code such that |C|=A(n,d). Perform the procedure described in the proof t=n-2d+1 times:

$$A(n,d) \leq 2^t A(n-t,d) \leq 2^t \ 2d/(2d-(n-t)) = 2^t.2d$$

$$R(\delta) \leq \lim_{n\to\infty} (1/n)[t + \log (2d)] = 1-2\delta$$

Asymptotic Bassalygo-Elias bound:

$$A(n,d) \le 2^n \min_w \frac{1}{\binom{n}{w}} \frac{dn}{dn - 2wn + 2w^2}$$

Smallest when the binomial is max, i.e., when w is largest possible such that $dn-2wn+2w^2 > 0$. For large n this value of w approaches the root of the quadratic (does not matter which root since h(½-x)=h(½+x))

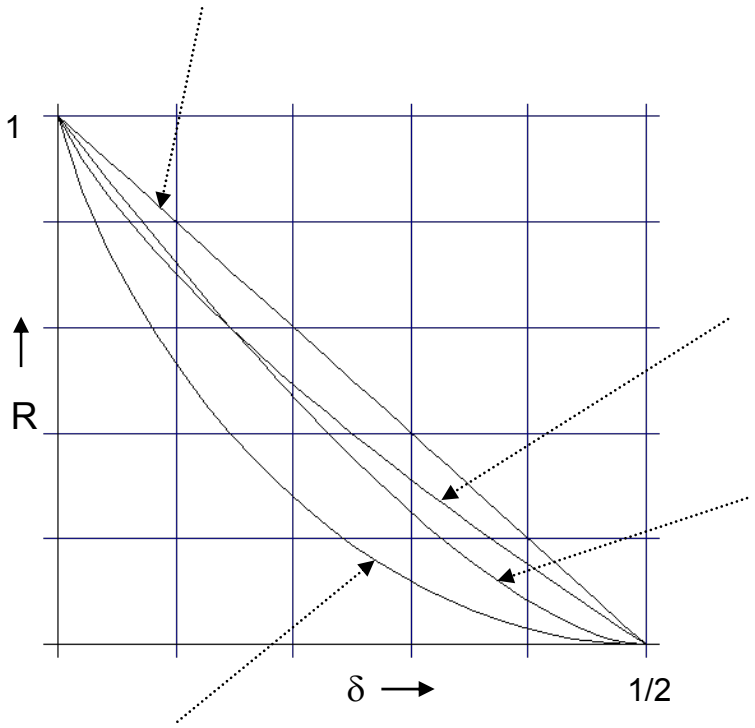
$$R = 1 - h[\tfrac{1}{2}(1 - \sqrt{1-2\delta})]$$

Exercise: Prove that the BE bound is always better that the Hamming bound.

Theorem 22.3 (McEliece, Rodemich, Rumsey, Welch 1977)

$$R = h(\tfrac{1}{2} - \sqrt{\delta(1-\delta)})$$

Asymptotic bounds on codes



BE bound $R = 1 - h[\frac{1}{2}(1 - \sqrt{1-2\delta})]$

LP bound $R = h(\frac{1}{2} - \sqrt{\delta(1-\delta)})$

ENEE626 Lectures 23-24: Ensembles of random codes

Plan:
Why random codes: Shannon's theorem for the BSC
Average and typical properties of random linear codes
Shannon's theorem for the erasure channel

The best known parameters of codes in terms of minimum distance and transmission reliability are attained by random choice. We will study from this point of view the two simplest ensembles of codes: random codes and random linear codes.

Definition 23.1: A random (n,M) code is obtained by choosing randomly with uniform distribution and independently M codewords out of the 2^n vectors
A random linear code of length n is obtained by choosing the entries of the parity-check matrix independently with (½ , ½) probability.

We will show that

- the capacity of the BSC(p) equals 1-h(p) and can be achieved by random codes;
- random linear codes achieve capacity of the BSC
- random linear codes achieve capacity of the erasure channel
- typical random linear codes achieve the GV asymptotic bound on the minimum distance

Notation:

h(x)=-x log x – (1-x) log (1-x) binary entropy

$$S_{n,t} = \sum_{w=0}^{t} \binom{n}{w}$$ volume of the ball of radius t

$S_{n,n\delta} \cong 2^{nh(\delta)}$

Markov inequality: Let X be a nonnegative r.v., then for a>0,

$$\Pr[X \geq a] \leq EX/a$$

Proof:

$$\Pr\{X \geq a\} \leq \frac{1}{a}\int_{a}^{\infty} x d\mu(x) \leq \frac{1}{a}\int_{0}^{\infty} x d\mu(x) = \frac{1}{a} EX$$

Chebyshev inequality:

$$\Pr\{|X - EX| \geq a\} \leq \frac{Var(X)}{a^2}$$

Take $Y=(X-EX)^2$. By Markov,

$$\Pr\{Y \geq a^2\} \leq \frac{EY}{a^2} = \frac{Var(X)}{a^2}$$

Shannon's Theorem 23.1: Given $\varepsilon>0$ and $R\leq 1-h(p)-\gamma$, $\gamma>0$, there is n_0 such that for any $n\geq n_0$ there exists a code $C\subset \mathbb{F}_2^n$ whose error probability of decoding on a BSC(p) satisfies $P_e \leq \varepsilon$.

Proof: Let $M=2^{Rn}$, $C=\{\mathbf{c}_1,\ldots\mathbf{c}_M\}$ be a random code of length n. Transmit $\mathbf{c}_1$ over the channel. Suppose that $\mathbf{y}$ is the received vector. Let $r=n(p+\alpha)$

Decode as follows:

If the sphere $B_r(\mathbf{y})$ contains exactly one codeword $\mathbf{c}$, output it

In all other cases declare an error

Let $X_i=1(\mathbf{c}_i\in B_r(y))$ be an r.v.

$$P_e \leq \Pr\{X_1=0\} + \Pr\{\textstyle\sum_{i=2}^{M} X_i\geq 1\}$$

$$\Pr\{X_1=0\}= \Pr\{ wt(\mathbf{y}-\mathbf{c}_1)>np+n\alpha \}$$
$$\leq \Pr\{|wt(\mathbf{y}-\mathbf{c}_1)-np|\geq n\alpha\}$$
$$\leq np(1-p)/(n\alpha)^2 = p(1-p)n^{-\varepsilon} \rightarrow 0 \qquad \text{by taking } \alpha= n^{-(1-\varepsilon)/2}$$

$$\Pr\{X_i=1\}=2^{-n} S_{n,r} \qquad \text{where } S_{n,r}=|B_r(y)|\cong 2^{nh(r/n)}$$
$$=\exp(-n(1-h(r/n))\}$$

$$\Pr\{\textstyle\sum_{i=2}^{M} X_i\geq 1\} \leq M \Pr\{X_i=1\}=\exp(Rn-n+nh(p+\alpha))$$
$$\leq \exp(n(\alpha'-\gamma)) \qquad \text{where we write } h(p+\alpha)=h(p)+\alpha'$$

Since $\alpha'\rightarrow 0$, it is possible to choose n so that $\alpha'<\gamma$ ▲

Random linear codes

Consider the ensemble of codes defined by random parity-check (n-k)xn matrices H.

For a random code C, let $\xi_i = 1(\mathbf{x}_i \in C)$, $i=1,\ldots,\binom{n}{w}$

$$\Pr\{\xi_i=1\}=E\xi_i=2^{k-n}$$

$$E\xi_i^2=E\xi_i;\ \mathrm{Var}(\xi_i)=E\xi_i-(E\xi_i)^2=2^{k-n}(1-2^{k-n})<E\xi_i$$

Let $A_w=|\{\mathbf{c}\in C: \mathrm{wt}(\mathbf{c})=w\}|$ be the random # of codewords of weight $w=\omega n$

$$EA_w = \binom{n}{w}2^{k-n} \cong 2^{n(R-1+h(\omega))}, 0<\omega<1$$

$$\mathrm{Var}\,(A_w) = \sum_{i=1}^{\binom{n}{w}} \mathrm{Var}\,\xi_i \le EA_w$$

As above, let δ satisfy $R=1-h(\delta)-\varepsilon$, then

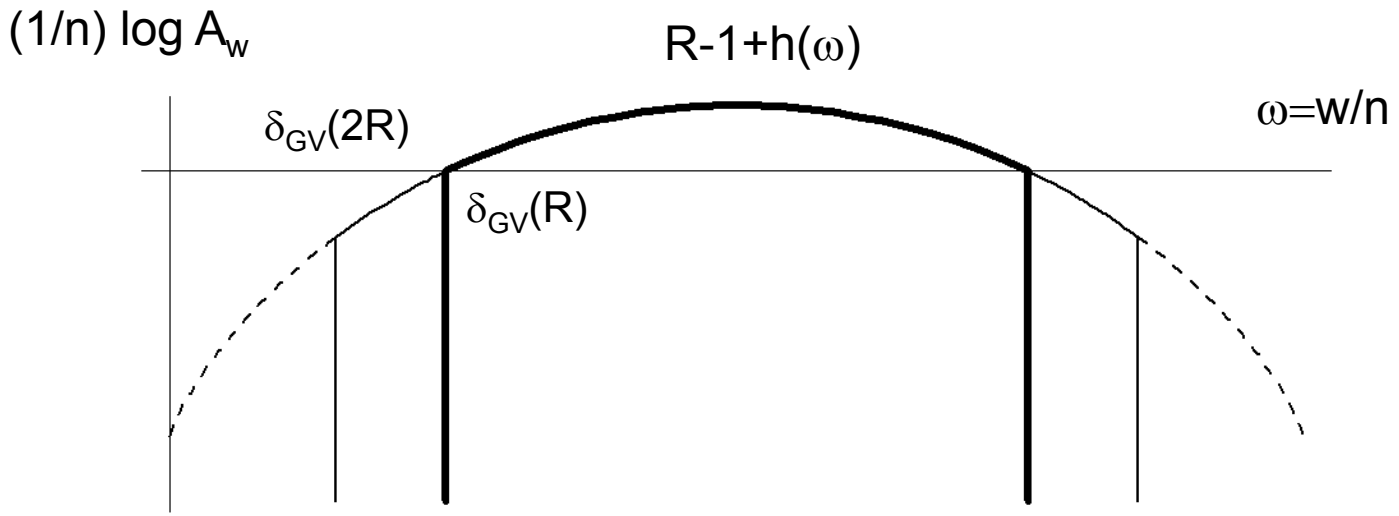
$$\Pr[A_{\delta n} > 0] \le EA_{\delta n} \lesssim 2^{-\varepsilon n} \to 0$$

so the distance $d(C) > n\delta$ with prob. ≈ 1.

For δ chosen from $R=1-h(\delta)+\varepsilon$, $\Pr[A_{\delta n}>0] \to 1$. Together this proves

Theorem 20.5: For all linear codes of rate R except for an exponentially small fraction of them, the relative distance approaches $\delta_{GV}(R)$

Let R be fixed



Bold curve: distance distribution of typical linear codes
Solid curve: distance distribution of typical unrestricted code
Dashed curve: average distance distribution of codes, linear or not

See also notes for Lect. 25 of the 2005 course, online;
A. Barg and G.D.Forney, Random Codes…, IEEE-IT 48, no. 9, 2002

Second part:
Random linear codes achieve capacity of the BSC
Same for the erasure channel

Theorem 23.6: Linear codes achieve capacity of the BSC.

Proof: Let C be an [n,Rn] code whose weight distribution is given by

$$A_w \begin{cases} \leq n\binom{n}{w}2^{k-n}, & \delta_{\mathrm{GV}}(R)n \leq w \leq n(1-\delta_{\mathrm{GV}}(R)) \\ 0 & \text{otherwise} \end{cases}$$

Suppose that **0** was transmitted and **y** is the received vector.
Decode as follows: Find

$$\mathbf{c}=\operatorname{argmin}_{\mathbf{x}\in C} \operatorname{dist}(\mathbf{x},\mathbf{y}). \qquad (1)$$

If $\operatorname{dist}(\mathbf{c},\mathbf{y}) \in [n(p-\alpha),n(p+\alpha)]$ (2)

output **c**, otherwise declare an error

Error event = {**0** does not satisfy (2)} $\cup$ {**0** sat. (2) $\cap$ $\exists$ $\mathbf{c}\neq\mathbf{0}$ sat. (1),(2) }

P_1=Pr{first of the above} = Pr{|wt(y)-np|$\geq n\alpha$} $\rightarrow$ 0 as before

Let $\mathcal{E}$={$\exists$ $\mathbf{c}\neq$ 0 that satisfies (1),(2)}, $\mathcal{E}_w(\mathbf{c})$={(1),(2) holds for $\mathbf{c}\in C\backslash\mathbf{0}$,wt(**c**)=w}

P_2=Pr{$\mathcal{E}$, wt(**y**)$\approx$pn} $\leq$ Pr{$\mathcal{E}$|wt(**y**)$\approx$pn} = $\sum_{w=d}^{n}$ Pr{$\cup_{\mathrm{wt}(\mathbf{c})=w}\mathcal{E}_w(\mathbf{c})$|wt(**y**)$\approx$pn}

$$\Pr\{\mathcal{E}_w(\mathbf{c})|\,\mathrm{wt}(\mathbf{y}) = pn\} = \binom{n}{pn}^{-1} \sum_{i=w/2}^{pn} \binom{w}{i}\binom{n-w}{pn-i} \le n\binom{n}{pn}^{-1}\binom{w}{w/2}\binom{n-w}{pn-w/2}$$

$$P_2 \le \sum_{w=d}^{2pn} A_w \Pr\{\mathcal{E}_w(\mathbf{c})|wt(y) = pn\}$$

In bounding the products of binomial coefficients we rely on the law of large numbers

$$\le n\binom{n}{pn}^{-1} \sum_{w=d}^{2pn} A_w \binom{w}{w/2}\binom{n-w}{pn-w/2}$$

$$\le n^2\binom{n}{pn}^{-1} 2^{k-n} \sum_{w} \binom{n}{w}\binom{w}{w/2}\binom{n-w}{pn-w/2}$$

$$= n^2 2^{k-n} \sum_{w} \binom{n-pn}{w/2}\binom{pn}{w/2}$$

$$\lesssim n^3 2^{k-n} \binom{n-pn}{np(1-p)}\binom{pn}{np(1-p)} \cong n^3 2^{k-n}\binom{n}{pn}$$

$$\binom{n}{w}\binom{w}{w/2}\binom{n-w}{pn-w/2} = \binom{n}{w/2, w/2, pn-w/2, n-pn-w/2} = \binom{n}{pn}\binom{n-pn}{w/2}\binom{pn}{w/2}$$

$$\log P_2 \le n(R - 1 + h(p) + o(1))$$

If R<1-h(p), then $P_2 \to 0$

▲

Definition 23.2: Let $C_i[n_i, R\, n_i]$, i=1,2,… be a sequence of codes. We say that p* is a threshold of the codes on a BSC channel (say) under some decoding $\mathcal{D}$

$p^* = \sup \{p \mid P_e(C_i) < \varepsilon$ starting with some i, under $\mathcal{D}\}$

The threshold of the best possible codes under typical-pairs decoding (or ML decoding) on a BSC was earlier called capacity.

Use the framed equation in the previous proof to extract the following.

Lemma 23.7: Let $a(\omega) = (1/n) \log A_{\omega n}$, where (A_w) is the (asymptotic) weight distribution of some code family. The threshold of these codes on a BSC satisfies $p^* \geq p$, where

$$a(\omega) < h(p) - \omega - (1-\omega) h\left(\frac{p - \omega/2}{1-\omega}\right), \qquad 0 < \omega < 1$$

Proof: By applying (1/n)log to the framed equation, $P_2 \to 0$ if p satisfies the condition of the lemma.

Remark: In the proof of the theorem we took codes with weight profile $a(\omega)=h(\omega)-1+R$, recovering a lower bound $p^* \geq h^{-1}(1-R)$ on their threshold.

Remarks.

1. We could have tried the Bhattacharyya bound for ML decoding error prob.:

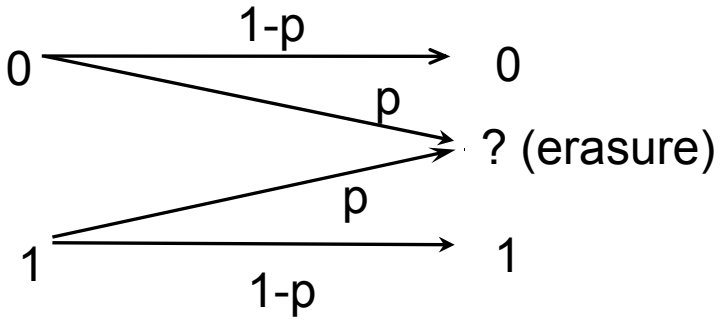
$$P_e \leq \sum_{w=d}^{n} A_w (\sqrt{4p(1-p)})^w$$

It turns out that this is insufficient to prove that the codes achieve capacity

2. It is possible to prove that the error probability of *max-likelihood decoding* of the code used to prove the theorem decreases exponentially as a function of code's length n for all values of the code rate $0 \leq R < 1-h(p)$
   The term error exponent is used to refer to $(1/n)\log 1/P_e$.

[1] A. Barg and G.D.Forney, Random Codes…, IEEE-IT 48, no. 9, 2002
[2] Notes for Lecture 5, class ENEE739C (2003), online.

Erasure channel



Theorem 23.8: The *capacity* of the erasure channel equals 1-p, attained by linear codes (For any $R=1-p-\varepsilon$ there exists a sequence of linear codes of length $n\to\infty$ for which the error prob. of decoding $P_e \to 0$).

Converse: If $R=1-p+\varepsilon$, then with probability bounded away from 0, $pn>n-k$ coordinates are erased. The remaining $<k$ coordinates match more than one codeword, so it is not possible to choose the right decision with high probability.

Proof of Direct part: Decoding: Let X$\subset$[1,…,n] be the set of erased positions in **y**.

If |X| $\notin$ [n(p-α),n(p+α)], discard **y**
Otherwise, if there is unique **c**$\in$ C such that $\mathbf{y}_i=\mathbf{c}_i$, i$\notin$ X, decode to **c**
If nonunique, discard **y**.

W.l.o.g., transmit 0, receive **y** which has 0's or erased coordinates (no errors)
Let $\mathcal{E}$, $\mathcal{E}_i(\mathbf{c})$ be the error event, resp. the error event such that **y** is decoded to **c**, wt(**c**)=i .We will only analyze the case of typical **y**, i.e., |X|$\approx$np.

$$\Pr\{\mathcal{E}||X|=np\}=\sum_{i=d}^{pn}\sum_{\mathbf{c}\in C,wt(\mathbf{c})=i}\Pr\{\mathcal{E}_i(\mathbf{c})||X|=np\}$$

$$\leq\binom{n}{np}^{-1}\sum_{i=d}^{np}A_i\binom{n-i}{pn-i}$$

$$\leq n\binom{n}{np}^{-1}2^{k-n}\sum_{i=d}^{pn}\binom{n}{i}\binom{n-i}{pn-i}=n2^{k-n}\sum_{i=d}^{pn}\binom{pn}{i}$$

$$\leq n2^{k-n+pn}$$

Thus, Pr{$\mathcal{E}$ | |X|=np}$\rightarrow$ 0 if (k/n-1+p)<0, which proves the direct part ▲