An introduction to Information Theory

Adrish Banerjee

Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, Uttar Pradesh
India

Aug. 8, 2016

Lecture #6C: Problem solving session-II

Huffman Coding

- **Problem # 1:** There can exist several fundamentally different optimal prefix-free codes for the same random variable, where fundamentally different means that their ordered list of codeword lengths are different. How many fundamentally different optimal binary prefix-free codes exist for a random variable with symbol probabilities 0.3, 0.2, 0.2, 0.1, 0.1, 0.05 and 0.05?

Huffman Coding

- **Problem # 1:** There can exist several fundamentally different optimal prefix-free codes for the same random variable, where fundamentally different means that their ordered list of codeword lengths are different. How many fundamentally different optimal binary prefix-free codes exist for a random variable with symbol probabilities 0.3, 0.2, 0.2, 0.1, 0.1, 0.05 and 0.05?
- **Solutions:** There are four possible code tree for Huffman coding

Prob. 1 (contd.)

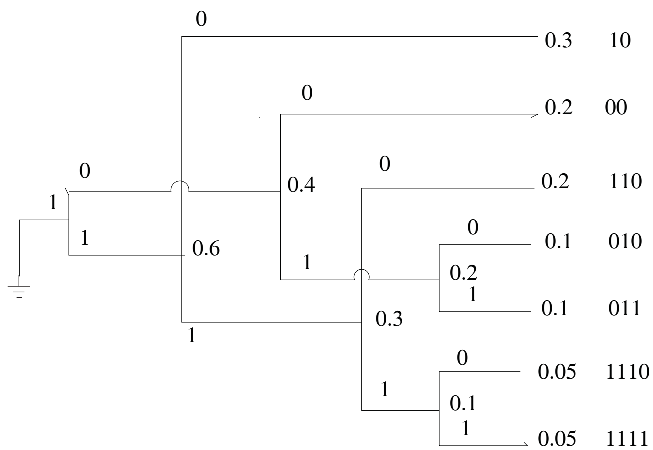


Figure: Code Tree 1

Ordered list of codeword length for Code Tree 1 is $\{2, 2, 3, 3, 3, 4, 4\}$

Prob. 1 (contd.)

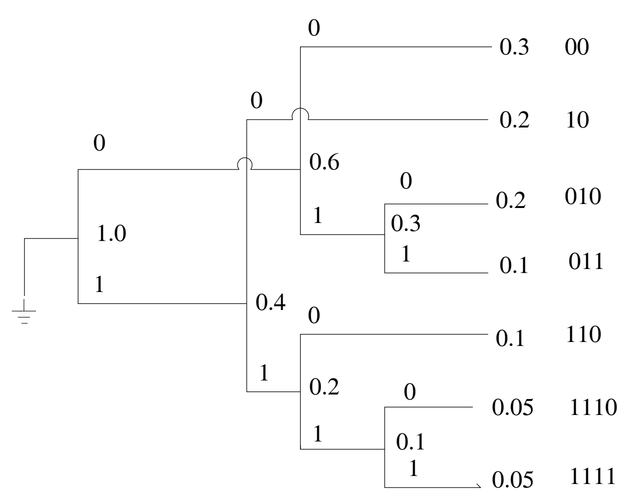


Figure: Code Tree 2

Ordered list of codeword length for Code Tree 2 is $\{2, 2, 3, 3, 3, 4, 4\}$

Prob. 1 (contd.)

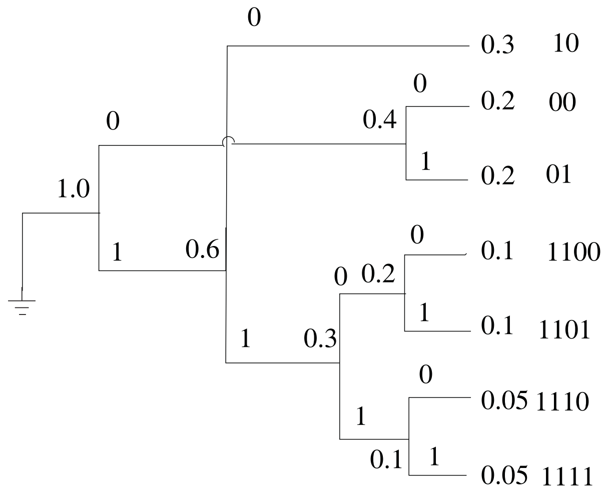
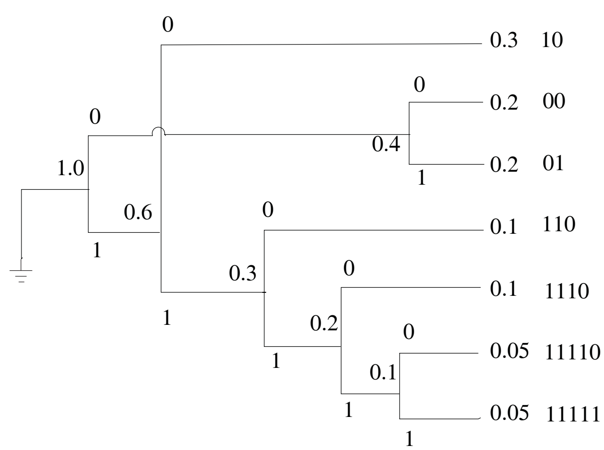


Figure: Code Tree 3

Ordered list of codeword length for Code Tree 3 is $\{2, 2, 2, 4, 4, 4, 4\}$

Prob. 1 (contd.)



Ordered list of codeword length is $\{2, 2, 2, 3, 4, 5, 5\}$.

- Thus there are three different optimal prefix free code exist with codewords lengths given by $\{2, 2, 3, 3, 3, 4, 4\}, \{2, 2, 2, 4, 4, 4, 4\}, \{2, 2, 2, 3, 4, 5, 5\}$

Huffman Coding

- **Problem # 2:** Y16 batch is going to select a team to represent IITK at inter-IIT chess meet. After initially screening, and preliminary matches, the non playing captain Vipul found that following students have the best chance of representing Y16 batch: Saurabh, Arjun, Amrita, Alankrita, Ritwik, Prasham, and Aditya.
- Their probability of winning if they play among each other is respectively, $1/3, 1/3, 1/9, 1/9, 1/27, 1/27, 1/27$. Captain Vipul wants to use an optimal block to variable length coding to convey these probabilities to coach Adrish who is vacationing in Kullu during Dussehra holidays.
- There are two telegraph services available in the campus namely, "*Ajit Speedy Post*" and "*Ketan Super Services*". *Ajit Speedy Post* transmits binary digits at Rs. 40 per digit, while *Ketan Super Services* transmits ternary digits at Rs. 65 per digit. Vipul has to select a service and design a code that will minimize the expected cost.
  i) What service should be selected and what is the expected cost?
  ii) If *Ketan Super Services* decides to increase their charges, at what value of new cost should Vipul change his mind?

Prob. 2 (contd.)

- **Solutions:** If Ajit Speedy Post is used, following codes can be assigned to the probabilities of the respective players

Name	Code
Saurabh	00
Arjun	01
Amrita	10
Alankrita	110
Ritwik	1110
Prasham	11110
Aditya	11111

Expected codeword length in this case is 2.4074 bits, and the expected cost is Rs. 96.296.

Prob. 2 (contd.)

- If Ketan Super Services is used, following codes can be assigned to the probabilities of the respective players

Name	Code
Saurabh	0
Arjun	1
Amrita	20
Alankrita	21
Ritwik	220
Prasham	221
Aditya	222

Expected codeword length in this case is 1.4444 bits, and the expected cost is Rs. 93.88.

i) Ketan Super Services should be used as it is a cheaper service and its cost is Rs. 93.88

ii) If the price exceeds Rs 66.67, Captain Vipul should change his mind.

Huffman Coding

- **Problem # 3:** Prove that for any binary Huffman code, if the most probable message symbol has probability $p_1 < 1/3$, then that symbol must be assigned a codeword of length ≥ 2

Huffman Coding

- **Problem # 3:** Prove that for any binary Huffman code, if the most probable message symbol has probability $p_1 < 1/3$, then that symbol must be assigned a codeword of length ≥ 2
- **Solutions:** Suppose for the sake of contradiction that $l_1 = 1$. Without loss of generality, assume that $c_1 = 0$.

Huffman Coding

- **Problem # 3:** Prove that for any binary Huffman code, if the most probable message symbol has probability $p_1 < 1/3$, then that symbol must be assigned a codeword of length ≥ 2
- **Solutions:** Suppose for the sake of contradiction that $l_1 = 1$. Without loss of generality, assume that $c_1 = 0$.
- The total probability of C_{10} and C_{11} is $1 - p_1 > 2/3$, so at least one of these sets (without loss of generality, C_{10}) has probability greater than $1/3$.

Huffman Coding

- **Problem # 3:** Prove that for any binary Huffman code, if the most probable message symbol has probability $p_1 < 1/3$, then that symbol must be assigned a codeword of length ≥ 2
- **Solutions:** Suppose for the sake of contradiction that $l_1 = 1$. Without loss of generality, assume that $c_1 = 0$.
- The total probability of C_{10} and C_{11} is $1 - p_1 > 2/3$, so at least one of these sets (without loss of generality, C_{10}) has probability greater than $1/3$.
- We can now obtain a better code by interchanging the subtree of the decoding tree beginning with 0 with the subtree beginning with 10; that is we replace codewords of the form 10x... by 0x... and let $c_1 = 10$.

Huffman Coding

- **Problem # 3:** Prove that for any binary Huffman code, if the most probable message symbol has probability $p_1 < 1/3$, then that symbol must be assigned a codeword of length ≥ 2
- **Solutions:** Suppose for the sake of contradiction that $l_1 = 1$. Without loss of generality, assume that $c_1 = 0$.
- The total probability of C_{10} and C_{11} is $1 - p_1 > 2/3$, so at least one of these sets (without loss of generality, C_{10}) has probability greater than $1/3$.
- We can now obtain a better code by interchanging the subtree of the decoding tree beginning with 0 with the subtree beginning with 10; that is we replace codewords of the form 10x... by 0x... and let $c_1 = 10$.
- This improvement contradicts the assumption that $l_1 = 1$, and so $l_1 \geq 2$.

Variable to Block Length Coding

- **Problem # 4:** "Run-length coding" is a popularly used variable-length-to-block coding scheme for binary information sources. Let 0^n1 denote a sequence of n zeros followed by one, i.e. a "run of zeros" of length n. For the run-length coding of blocklength N, the message $v = 0^n1$ would be encoded into $[b_1, b_2, \ldots, b_N]$, the binary representation of the integer n for $0 \leq n < 2^N - 1$, while the message 0^{2^N-1} would be encoded as $[1, 1, \ldots, 1]$. Consider the binary memoryless source with $P_U(0) = 0.9$ and $P_U(1) = 0.1$.

Variable to Block Length Coding

- **Problem # 4:** "Run-length coding" is a popularly used variable-length-to-block coding scheme for binary information sources. Let 0^n1 denote a sequence of n zeros followed by one, i.e. a "run of zeros" of length n. For the run-length coding of blocklength N, the message $v = 0^n1$ would be encoded into $[b_1, b_2, \ldots, b_N]$, the binary representation of the integer n for $0 \leq n < 2^N - 1$, while the message 0^{2^N-1} would be encoded as $[1, 1, \ldots, 1]$. Consider the binary memoryless source with $P_U(0) = 0.9$ and $P_U(1) = 0.1$.
  - i) Find the smallest N such that the run-length coding scheme is not a Tunstall message set.

Variable to Block Length Coding

- **Problem # 4:** "Run-length coding" is a popularly used variable-length-to-block coding scheme for binary information sources. Let 0^n1 denote a sequence of n zeros followed by one, i.e. a "run of zeros" of length n. For the run-length coding of blocklength N, the message $v = 0^n1$ would be encoded into $[b_1, b_2, \ldots, b_N]$, the binary representation of the integer n for $0 \le n < 2^N - 1$, while the message 0^{2^N-1} would be encoded as $[1, 1, \ldots, 1]$. Consider the binary memoryless source with $P_U(0) = 0.9$ and $P_U(1) = 0.1$.
  i) Find the smallest N such that the run-length coding scheme is not a Tunstall message set.
  ii) As a function of N, find $N/E[Y]$ for the run-length coding scheme, where E[Y] is the average message length.

Variable to Block Length Coding

- **Problem # 4:** "Run-length coding" is a popularly used variable-length-to-block coding scheme for binary information sources. Let 0^n1 denote a sequence of n zeros followed by one, i.e. a "run of zeros" of length n. For the run-length coding of blocklength N, the message $v = 0^n1$ would be encoded into $[b_1, b_2, \ldots, b_N]$, the binary representation of the integer n for $0 \le n < 2^N - 1$, while the message 0^{2^N-1} would be encoded as $[1, 1, \ldots, 1]$. Consider the binary memoryless source with $P_U(0) = 0.9$ and $P_U(1) = 0.1$.
  i) Find the smallest N such that the run-length coding scheme is not a Tunstall message set.
  ii) As a function of N, find $N/E[Y]$ for the run-length coding scheme, where E[Y] is the average message length.
  iii) Find the N which maximizes the efficiency ($\eta = H(U)E[Y]/(N \log D)$) of run-length coding for this BMS and compute this maximum efficiency.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Prob. 4 (contd.)

- **Solutions:**

Prob. 4 (contd.)

- **Solutions:**
  i) After Tunstall's algorithm, the number of extensions from the extended root is

$$q = \lfloor \frac{D^N - K}{K - 1} \rfloor$$

Prob. 4 (contd.)

- **Solutions:**
  i) After Tunstall's algorithm, the number of extensions from the extended root is

$$q = \lfloor \frac{D^N - K}{K - 1} \rfloor$$

- For the case of $D = 2$ and $K = 2$, we have the number of extensions, $q = 2^N - 2$. In run length coding the least probable intermediate root has probability

$$p' = [P_u(0)]^q = 0.9^{2^N - 2}$$

Prob. 4 (contd.)

- **Solutions:**
  i) After Tunstall's algorithm, the number of extensions from the extended root is

$$q = \lfloor \frac{D^N - K}{K - 1} \rfloor$$

- For the case of $D = 2$ and $K = 2$, we have the number of extensions, $q = 2^N - 2$. In run length coding the least probable intermediate root has probability

$$p' = [P_u(0)]^q = 0.9^{2^N - 2}$$

- If p' is smaller than $P_u(1)$, then according to Tunstall lemma, $P_u(1)$ should have been extended, but which is against the run-length coding rule.

Prob. 4 (contd.)

- Hence the runlength message set is no longer Tunstall message set if and only if

$$\begin{aligned} P' &< P_u(1) \\ 0.9^{2^N-2} &< 0.1 \\ 2^N - 2 &< \frac{\log 0.1}{\log 0.9} \\ N &> \log\left(\frac{\log 0.1}{\log 0.9} + 2\right) = 4.58 \end{aligned}$$

Thus for $N = 5$, the run-length coding message set is no longer a Tunstall message set.

Prob. 4 (contd.)

ii) E[Y] can be calculated using path length lemma as

$$E[Y] = 1.0 + 0.9^1 + 0.9^2 + \cdots + 0.9^{2^N-2} = \frac{1-(0.9)^{2^N-1}}{0.1}$$

Thus we have

$$\frac{N}{E[Y]} = \frac{(0.1)N}{1-(0.9)^{2^N-1}}$$

Prob. 4 (contd.)

ii) E[Y] can be calculated using path length lemma as

$$E[Y] = 1.0 + 0.9^1 + 0.9^2 + \cdots + 0.9^{2^N-2} = \frac{1-(0.9)^{2^N-1}}{0.1}$$

Thus we have

$$\frac{N}{E[Y]} = \frac{(0.1)N}{1-(0.9)^{2^N-1}}$$

iii) The efficiency η is maximized when $\frac{N}{E[Y]}$ is minimized.

N	N/E[Y]
1	1
2	0.7380
3	0.5750
4	0.5037
5	0.5198
6	0.6008

Prob. 4 (contd.)

ii) E[Y] can be calculated using path length lemma as

$$E[Y] = 1.0 + 0.9^1 + 0.9^2 + \cdots + 0.9^{2^N-2} = \frac{1-(0.9)^{2^N-1}}{0.1}$$

Thus we have

$$\frac{N}{E[Y]} = \frac{(0.1)N}{1-(0.9)^{2^N-1}}$$

iii) The efficiency η is maximized when $\frac{N}{E[Y]}$ is minimized.

N	N/E[Y]
1	1
2	0.7380
3	0.5750
4	0.5037
5	0.5198
6	0.6008

- Hence for N=4, $\frac{N}{E[Y]}$ is minimized.

Typical Sequence

- **Problem # 5:** A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $p(1) = 0.005$ and $p(0) = 0.995$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing three or fewer ones.

Typical Sequence

- **Problem # 5:** A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $p(1) = 0.005$ and $p(0) = 0.995$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing three or fewer ones.
  i) Assuming all the codewords are the same length, find the minimum length required to provide codewords for all sequences with three or fewer ones.

Typical Sequence

- **Problem # 5:** A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $p(1) = 0.005$ and $p(0) = 0.995$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing three or fewer ones.
  - i) Assuming all the codewords are the same length, find the minimum length required to provide codewords for all sequences with three or fewer ones.
  - ii) Calculate the probability of observing a source sequence for which no codeword has been assigned.

Typical Sequence

- **Problem # 5:** A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $p(1) = 0.005$ and $p(0) = 0.995$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing three or fewer ones.
  - i) Assuming all the codewords are the same length, find the minimum length required to provide codewords for all sequences with three or fewer ones.
  - ii) Calculate the probability of observing a source sequence for which no codeword has been assigned.
  - iii) Use Chebyshev's inequality to bound the probability of observing a source sequence for which no codeword has been assigned. Compare this bound with the actual probability computed in part ii).

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Prob. 5 (contd.)

- **Solutions:**

Prob. 5 (contd.)

- **Solutions:**
  a) The number of 100-bit binary sequences with three or fewer ones is given by

$$\begin{aligned} & \binom{100}{0} + \binom{100}{1} + \binom{100}{2} + \binom{100}{3} \\ = \quad & 1 + 100 + 4950 + 161700 = 166751 \end{aligned}$$

  The required codeword length is $\log_2 166751 = 18$

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Prob. 5 (contd.)

- **Solutions:**
  a) The number of 100-bit binary sequences with three or fewer ones is given by

$$\begin{aligned} & \binom{100}{0} + \binom{100}{1} + \binom{100}{2} + \binom{100}{3} \\ = \; & 1 + 100 + 4950 + 161700 = 166751 \end{aligned}$$

  The required codeword length is $\log_2 166751 = 18$

  b) The probability that a 100-bit sequence has three or fewer ones is given by

$$\begin{aligned} & \sum_{i=0}^{3} \binom{100}{i} (0.005)^i (0.995)^{100-i} \\ = \; & 0.605577 + 0.30441 + 0.7572 + 0.01243 = 0.99833 \end{aligned}$$

  Thus the probability that the sequence which is generated cannot be encoded is given by $1 - 0.99833 = 0.00167$.

Prob. 5 (contd.)

c) In the case of a random variable S_N that is the sum of n i.i.d. random variables $X_1, X_2, \cdots, X_n$, Chebyshev's inequality states that

$$P(|S_N - n\mu| \geq \epsilon) \leq \frac{n\sigma^2}{\epsilon^2}$$

where μ and σ^2 are the mean and variance of X_i.

Prob. 5 (contd.)

c) In the case of a random variable S_N that is the sum of n i.i.d. random variables $X_1, X_2, \cdots, X_n$, Chebyshev's inequality states that

$$P(|S_N - n\mu| \geq \epsilon) \leq \frac{n\sigma^2}{\epsilon^2}$$

where μ and σ^2 are the mean and variance of X_i.

- In this problem, $n = 100, \mu = 0.005$ and $\sigma^2 = (0..05)0.995$.

Prob. 5 (contd.)

c) In the case of a random variable S_N that is the sum of n i.i.d. random variables $X_1, X_2, \cdots, X_n$, Chebyshev's inequality states that

$$P(|S_N - n\mu| \geq \epsilon) \leq \frac{n\sigma^2}{\epsilon^2}$$

where μ and σ^2 are the mean and variance of X_i.

- In this problem, $n = 100, \mu = 0.005$ and $\sigma^2 = (0..05)0.995$.
- Also, $S_{100} \geq 4$ if and only if $|S_{100} - 100(0.005)| \geq 3.5$, so we should choose $\epsilon = 3.5$. Then

$$P(|S_{100} \geq 4) \leq \frac{100(0.005)(0.995)}{3.5^2} \approx 0.04061$$

Typical Set

- **Problem # 6:** We have a memoryless source U, i.e. $U_1, U_2, \cdots$, are i.i.d. $\sim$ U takes values in the finite alphabet $\mathcal{U}$. Let u^n denote the n-tuple $(u_1, u_2, \cdots, u_n)$ and $p(u^n)$ be its probability, i.e.,

$$p(u^n) = \prod_{i=1}^{n} P_U(u_i)$$

Let $\delta > 0$ and for every n let $B^{(n)} \subseteq \mathcal{U}^n$ be an arbitrary set of source sequences satisfying $|B^{(n)}| \leq 2^{n(H(U)-\delta)}$. Prove that

$$\lim_{n\to\infty} P(U^n \in B^{(n)}) = 0$$

In other words, the typical set $A_\epsilon^{(n)}$ is essentially smallest (on an exponential scale) among the sets that have non-negligible probability.

Prob. 6 (contd.)

- **Solutions:** Consider the typical set , $A_\epsilon^{(n)}$ with $\epsilon = \frac{\delta}{2}$.

Prob. 6 (contd.)

- **Solutions:** Consider the typical set , $A_\epsilon^{(n)}$ with $\epsilon = \frac{\delta}{2}$.
- As, $B^{(n)} = (B^{(n)} \cap A_\epsilon^{(n)}) \cup (B^{(n)} \backslash A_\epsilon^{(n)})$

$$\begin{aligned} P(U^n \in B^{(n)}) &\leq P(U^n \in B^{(n)} \cap A_\epsilon^{(n)}) + P(U^n \in B^{(n)} \backslash A_\epsilon^{(n)}) \\ &\leq P(U^n \in B^{(n)} \cap A_\epsilon^{(n)}) + P(A_\epsilon^{(n),c}) \end{aligned}$$

where superscript c stands for complement.

Prob. 6 (contd.)

- **Solutions:** Consider the typical set , $A_\epsilon^{(n)}$ with $\epsilon = \frac{\delta}{2}$.
- As, $B^{(n)} = (B^{(n)} \cap A_\epsilon^{(n)}) \cup (B^{(n)} \backslash A_\epsilon^{(n)})$

$$\begin{aligned} P(U^n \in B^{(n)}) &\leq P(U^n \in B^{(n)} \cap A_\epsilon^{(n)}) + P(U^n \in B^{(n)} \backslash A_\epsilon^{(n)}) \\ &\leq P(U^n \in B^{(n)} \cap A_\epsilon^{(n)}) + P(A_\epsilon^{(n),c}) \end{aligned}$$

where superscript c stands for complement.

- Note that $P(A_\epsilon^{(n),c}) \to 0$ as $n \to \infty$ by AEP.

Prob. 6 (contd.)

- Now we have

$$
\begin{aligned}
P(U^n \in B^{(n)} \cap A_\epsilon^{(n)}) &\leq \sum_{u^n \in B^{(n)} \cap A_\epsilon^{(n)}} p(u^n) \\
&\leq \sum_{u^n \in B^{(n)} \cap A_\epsilon^{(n)}} 2^{-n(H(U)-\epsilon)} \\
&\leq |B^{(n)}| 2^{-n(H(U)-\epsilon)} \\
&\leq 2^{-\frac{n\delta}{2}} \rightarrow 0
\end{aligned}
$$

which proves, $P(U^n \in B^{(n)}) \rightarrow 0$.