CT111 Intro to Communication Systems
Lecture 4: Weak Law of Large Numbers and Typical Sets

Yash M. Vasavada

Associate Professor, DA-IICT, Gandhinagar

5th Feb 2020

Overview of Today's Talk

Overview of Today's Talk

Overview of Today's Talk

1. WLLN
2. Typical Sets
3. Compression of Digital Data

Average of N Data Points

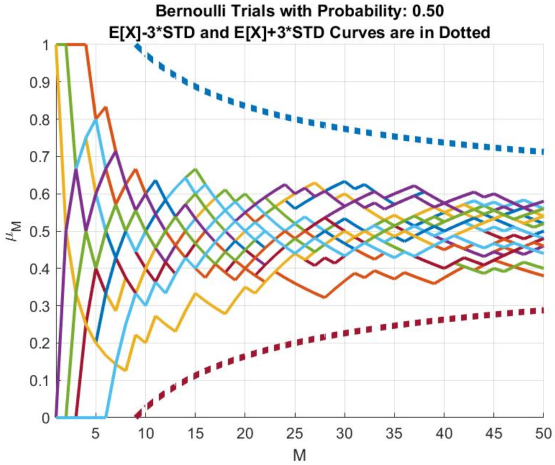
- *Weak Law of Large Numbers (WLLN)* states that the average $\mu_M = \frac{1}{M}\sum_{m=1}^{M} X_m$ of N samples $\{X_1, X_2, \ldots, X_M\}$ of a random variable X with a finite variance $\mathrm{Var}[X]$ converges to its expected value $E[X]$ as $M \to \infty$.
- Specifically, for an arbitrarily small positive scalar ϵ, $\lim_{M\to\infty} P(|\mu_M - E[X]| \geq \epsilon) = 0$.
- Use the Chebyshev's inequality

$$P(|\mu_M - E[\mu_M]| \geq \epsilon) \leq \frac{\mathrm{Var}[\mu_M]}{\epsilon^2}$$

to prove the WLLN.

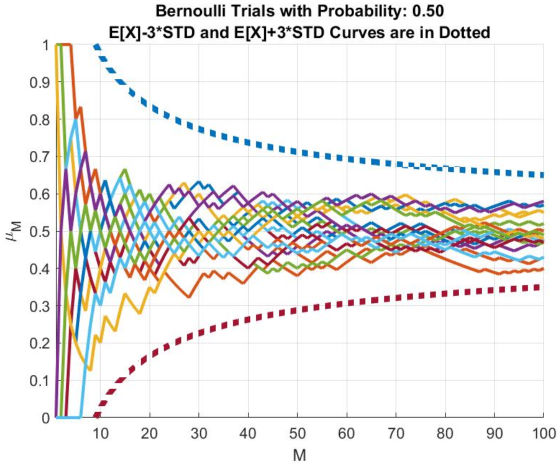
Average of M Data Points

$p = 0.5$, up to $M = 50$



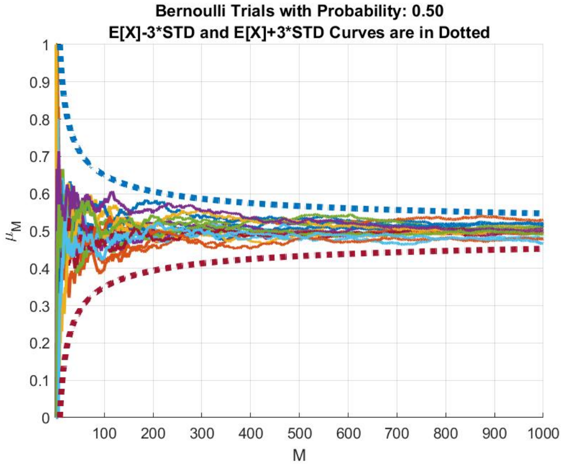
Average of M Data Points

$p = 0.5$, up to $M = 100$



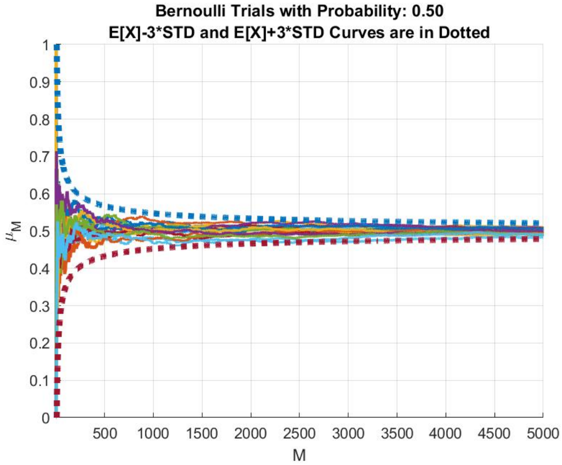
Average of M Data Points

$p = 0.5$, up to $M = 1000$



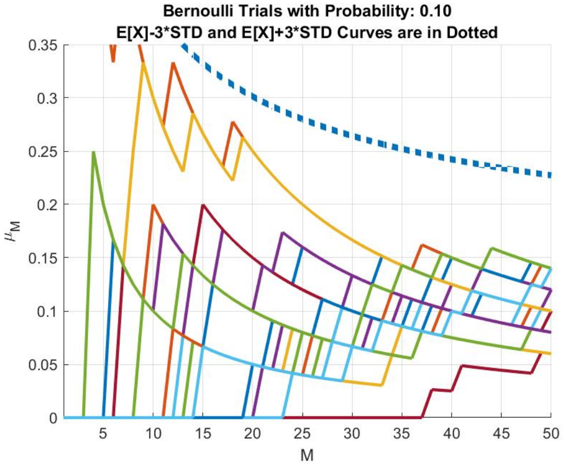
Average of M Data Points

$p = 0.5$, up to $M = 5000$



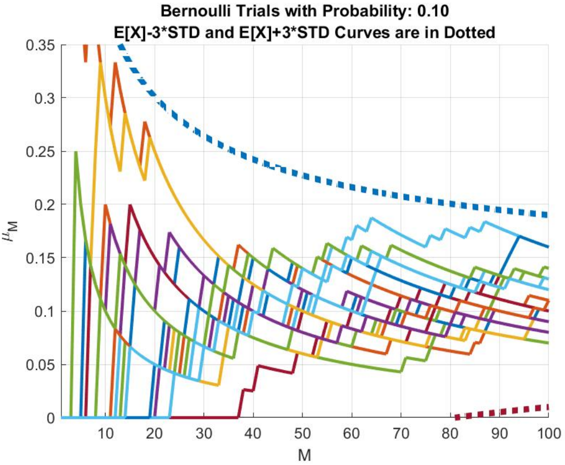
Average of M Data Points

$p = 0.1$, up to $M = 50$



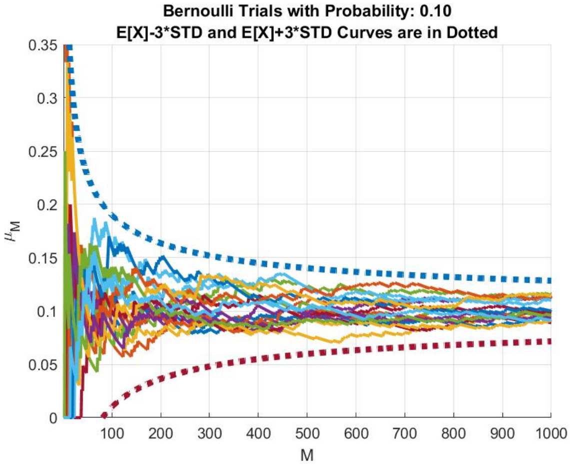
Average of M Data Points

$p = 0.1$, up to $M = 100$



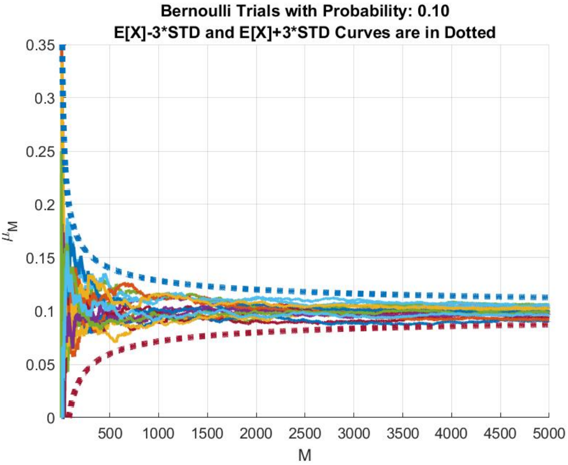
Average of M Data Points

$p = 0.1$, up to $M = 1000$



Average of M Data Points

$p = 0.1$, up to $M = 5000$



A Sequence of M bits

$M = 2, p = 0.1$

- Let us consider a sequence of $M = 2$ consecutive bits generated by a random information source

A Sequence of M bits

$M = 2, p = 0.1$

- Let us consider a sequence of $M = 2$ consecutive bits generated by a random information source
- We will see $2^M = 4$ possible sequences

A Sequence of M bits

$M = 2, p = 0.1$

- Let us consider a sequence of $M = 2$ consecutive bits generated by a random information source
- We will see $2^M = 4$ possible sequences
- These can be divided into $M + 1 = 3$ different, non-overlapping, subsets of binary sequences
  1. Subset 1 is all-zero sequence (no ones)
     - → Has one sequence $[0, 0]$, which occurs with a probability of $(1 - p)^2 = 0.9^2 = 0.81$

A Sequence of M bits

$M = 2, p = 0.1$

- Let us consider a sequence of $M = 2$ consecutive bits generated by a random information source
- We will see $2^M = 4$ possible sequences
- These can be divided into $M + 1 = 3$ different, non-overlapping, subsets of binary sequences
  1. Subset 1 is all-zero sequence (no ones)
     - → Has one sequence $[0, 0]$, which occurs with a probability of $(1 - p)^2 = 0.9^2 = 0.81$
  2. Subset 2 is a set of all sequences with exactly one 1
     - → Has $\binom{M=2}{1} = 2$ sequences ($[0, 1]$ and $[1, 0]$), each of which occurs with probability of $0.9 \times 0.1 = 0.09$. Therefore, total probability of this subset is $2 \times 0.09 = 0.18$.

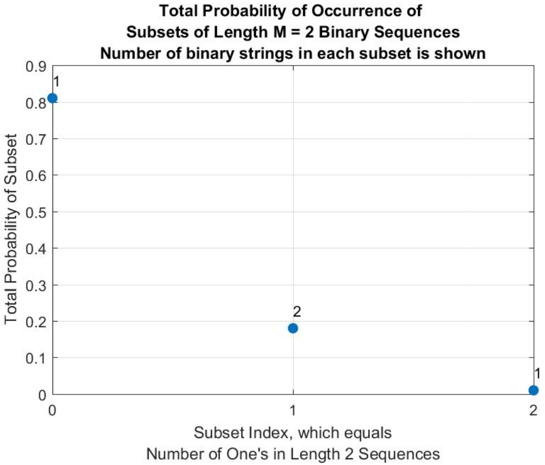
A Sequence of M bits

$M = 2, p = 0.1$

- Let us consider a sequence of $M = 2$ consecutive bits generated by a random information source
- We will see $2^M = 4$ possible sequences
- These can be divided into $M + 1 = 3$ different, non-overlapping, subsets of binary sequences
  1. Subset 1 is all-zero sequence (no ones)
     - $\rightarrow$ Has one sequence $[0, 0]$, which occurs with a probability of $(1-p)^2 = 0.9^2 = 0.81$
  2. Subset 2 is a set of all sequences with exactly one 1
     - $\rightarrow$ Has $\binom{M=2}{1} = 2$ sequences ($[0, 1]$ and $[1, 0]$), each of which occurs with probability of $0.9 \times 0.1 = 0.09$. Therefore, total probability of this subset is $2 \times 0.09 = 0.18$.
  3. Finally, subset 3 is a set of all-ones sequences
     - $\rightarrow$ Has one sequence $[1, 1]$, which occurs with a probability of $(1-p)^2 = 0.1^2 = 0.01$

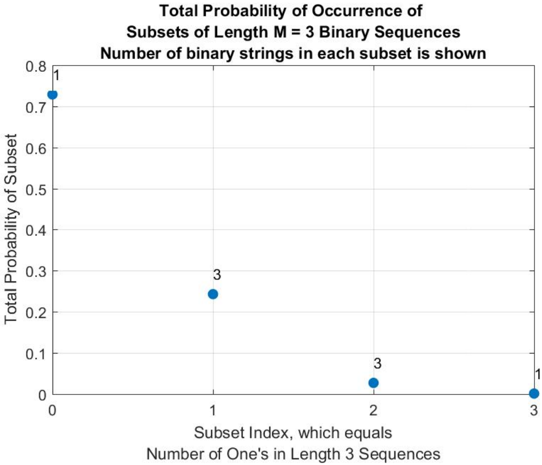
A Sequence of M DMS symbols

$M = 2, p = 0.1$



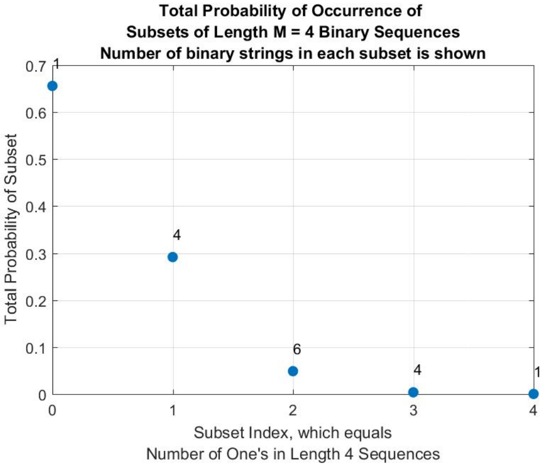
A Sequence of M DMS symbols

$M = 3, p = 0.1$



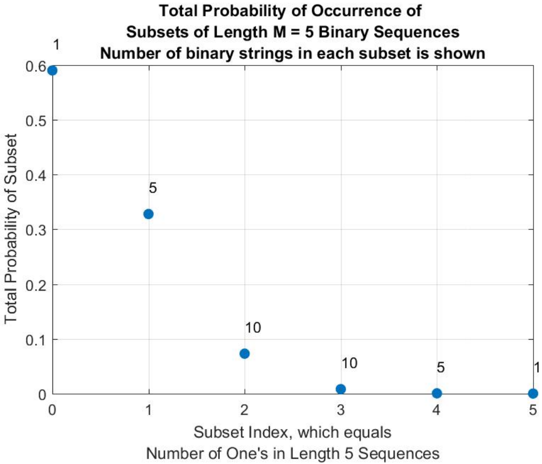
A Sequence of M DMS symbols

$M = 4, p = 0.1$



A Sequence of M DMS symbols

$M = 5, p = 0.1$



A Few More Questions

- What do we expect to see as we keep on increasing M?

A Few More Questions

- What do we expect to see as we keep on increasing M?
- A claim: only one subset will survive!

A Few More Questions

- What do we expect to see as we keep on increasing M?
- A claim: only one subset will survive!
- Which one?

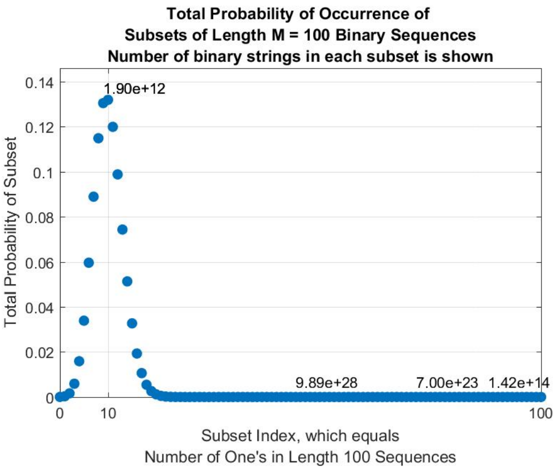
A Note

- As we keep on increasing M, the set of binary sequences, which has total 2^M members, becomes *huge*
  - $\rightarrow$ $M = 58$, 2^M is the age of universe in seconds
  - $\rightarrow$ $M = 171$, 2^M is the number of electrons in the Earth
  - $\rightarrow$ $M = 190$, 2^M is the number of electrons in the solar system
  - $\rightarrow$ $M = 266$, 2^M is the number of electrons in the universe

  (REF: David MacKay book: "Information Theory, Inference, and Learning Algorithms," available on the web for free download)

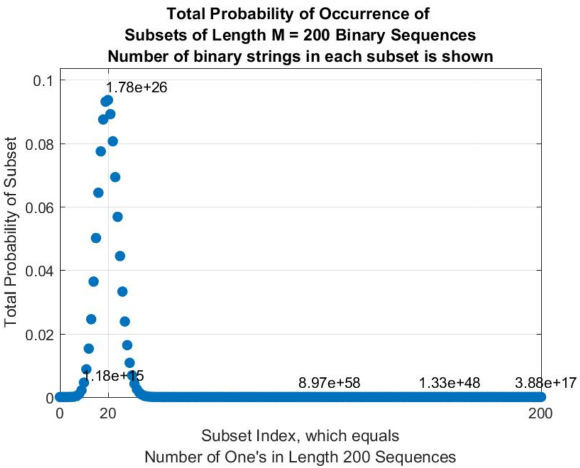
A Sequence of M DMS symbols

$M = 100, p = 0.1$



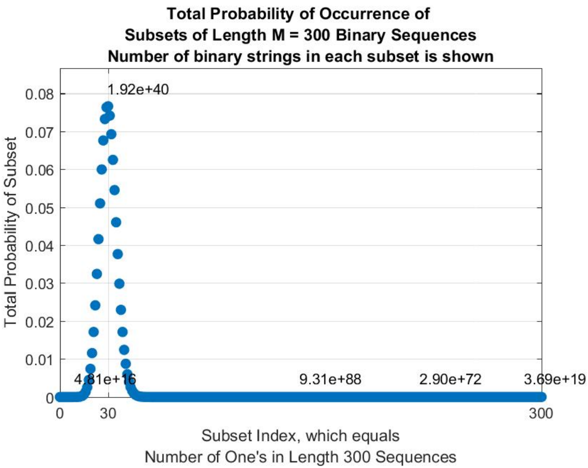
A Sequence of M DMS symbols

$M = 200, p = 0.1$



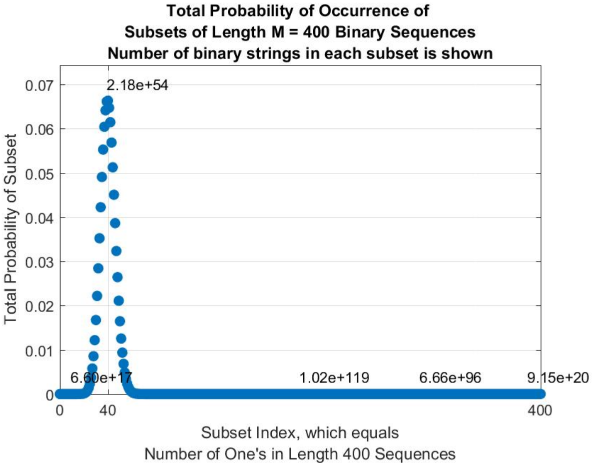
A Sequence of M DMS symbols

$M = 300, p = 0.1$



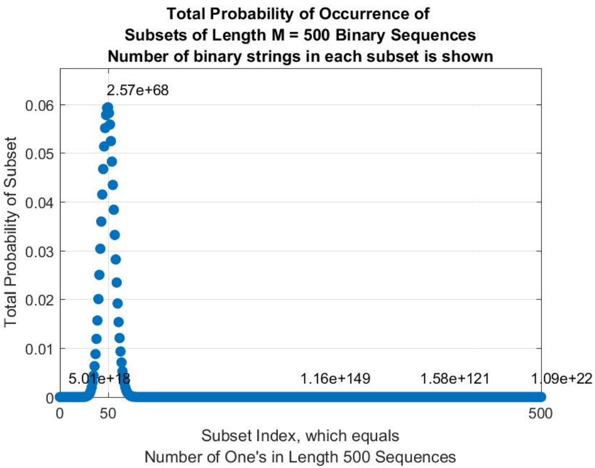
A Sequence of M DMS symbols

$M = 400, p = 0.1$



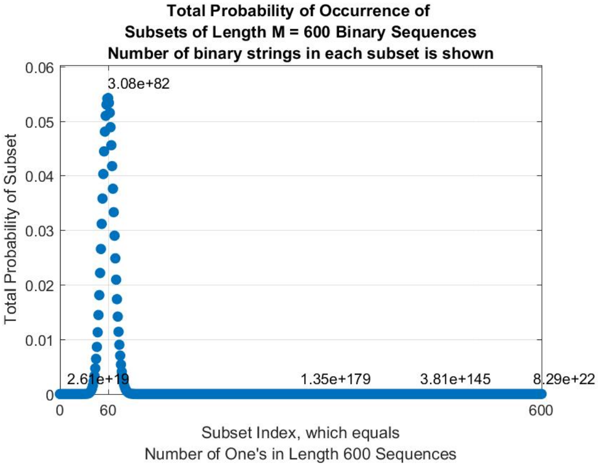
A Sequence of M DMS symbols

$M = 500, p = 0.1$



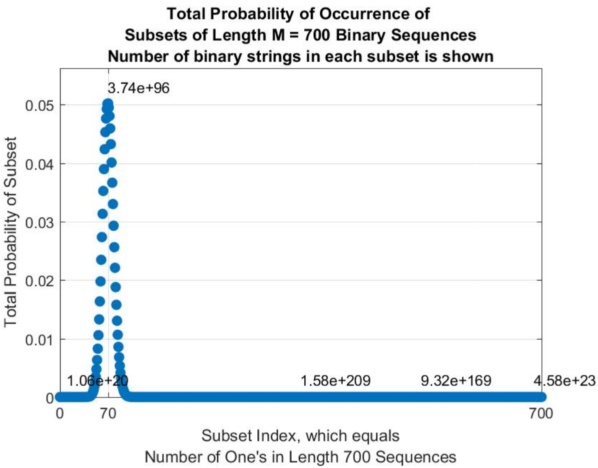
A Sequence of M DMS symbols

$M = 600, p = 0.1$



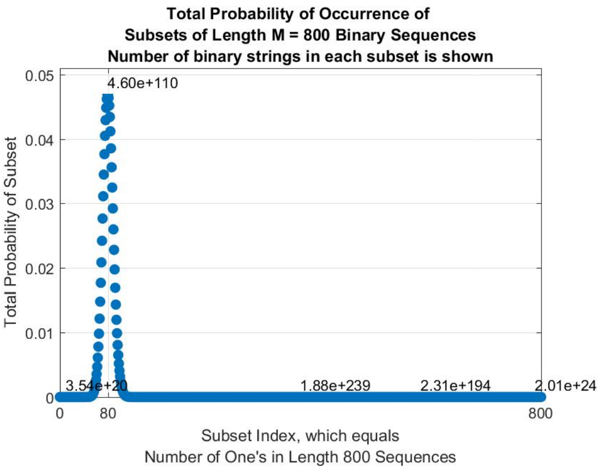
A Sequence of M DMS symbols

$M = 700, p = 0.1$



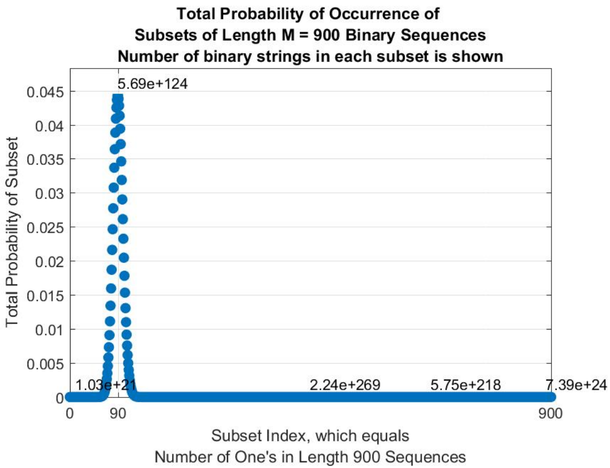
A Sequence of M DMS symbols

$M = 800$, $p = 0.1$



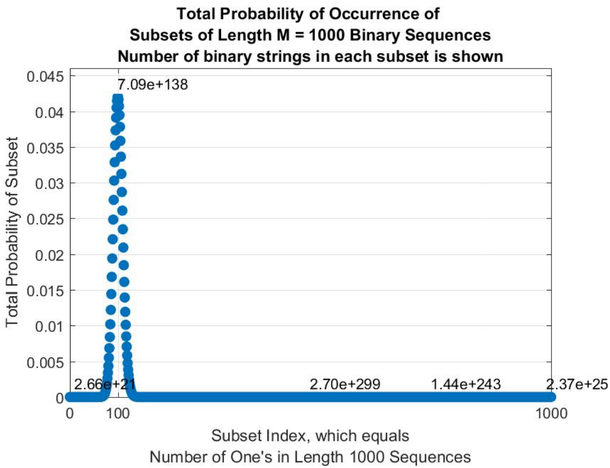
A Sequence of M DMS symbols

$M = 900, p = 0.1$



A Sequence of M DMS symbols

$M = 1000, p = 0.1$



A Few More Questions

- What do we expect to see as we keep on increasing M?

A Few More Questions

- What do we expect to see as we keep on increasing M?
- A claim: only one subset will survive!

A Few More Questions

- What do we expect to see as we keep on increasing M?
- A claim: only one subset will survive!
- Which one?

A Few More Questions

- What do we expect to see as we keep on increasing M?
- A claim: only one subset will survive!
- Which one?
- The subset formed by *all* binary sequences of length M which have $M_1 = p \times M$ ones

A Few More Questions

- What do we expect to see as we keep on increasing M?
- A claim: only one subset will survive!
- Which one?
- The subset formed by *all* binary sequences of length M which have $M_1 = p \times M$ ones
- Why?

A Few More Questions

- What do we expect to see as we keep on increasing M?
- A claim: only one subset will survive!
- Which one?
- The subset formed by *all* binary sequences of length M which have $M_1 = p \times M$ ones
- Why?
- Because that is *exactly* the definition of probability p:
  $p = \lim_{M \to \infty} \frac{M_1}{M}$

Coming to a Conclusion

- As $M \to \infty$, only one subset of 2^M binary sequences survives
  - $\rightarrow$ Total probability of this subset $\to 1$
  - $\rightarrow$ Total probability of all other (nonsurviving) subsets $\to 0$
- No matter which binary sequence is picked from this subset,
  - $\rightarrow$ it has $p \times M$ ones and $(1-p) \times M$ zeros, and
  - $\rightarrow$ therefore, the probability of each of these sequences is *identical* and equal to $p^{pM}(1-p)^{(1-p)M}$

Conclusions

- As $M \to \infty$, only one subset of binary sequence survives
- Let the probability of occurrence of this subset be denoted as p_{typ}. As $M \to \infty$, $p_{typ} \to 1$.
- Each of binary sequences picked from the surviving subset has
  - $\to$ $p \times M$ ones and $(1 - p) \times M$ zeros, and
  - $\to$ probability of occurrence which is equal to $p_i = p^{pM}(1 - p)^{(1-p)M}$
- Suppose the size of the surviving subset is K (i.e., it has K binary sequences)
  - $\to$ $p_{typ} = K \times p_i = K \times p^{pM}(1 - p)^{(1-p)M} \to 1$
  - $\to$ Therefore, $K = p^{-pM}(1 - p)^{-(1-p)M}$

Conclusions

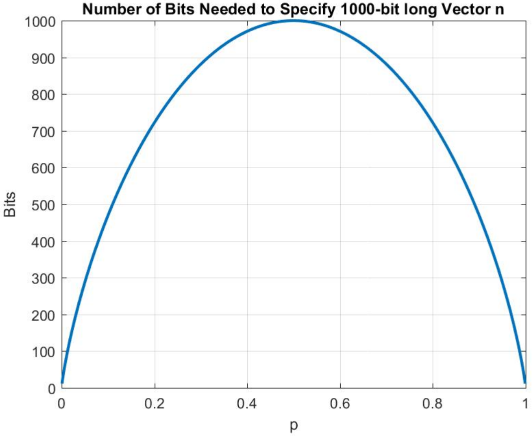
- The size of surviving subset, which we will now call *typical set*, is $K = p^{-pM}(1-p)^{-(1-p)M}$
- Each of its member sequences has equal probability of $p_i = p^{pM}(1-p)^{(1-p)M}$
- Therefore, we can use *fixed-length* coding to represent these K sequences
- This fixed-length code will require *exactly*

$$\begin{aligned}\log_2 K &= \log_2 \left\{ p^{-pM}(1-p)^{-(1-p)M} \right\} \\ &= M \times \left(-p\log_2 p - (1-p)\log_2(1-p)\right) \\ &= M \times H_b(p)\end{aligned}$$

bits. Less than $M \times H_b(p)$ bits will not be sufficient. More than $M \times H(X)$ bits are too many.

Binary Entropy Function
$H_b(p)$

- Entropy function for the binary set $H_b(p) \times M$ ($M = 1000$ bits):



A Question

1. How is the binary Entropy function related to the combinatorial function $\binom{M}{k}$?
2. Generalize: the derivation in the prior slides assumes that the information source generates bits 0 or 1 with probabilities $p_1 = p$ and $p_2 = 1 - p$. Suppose the information source generates one of M symbols having probabilities p_m, where $\sum_{m=1}^{M} p_m = 1$. Derive the typical set formulation and the Entropy function for such non-binary (M-ary) information source.

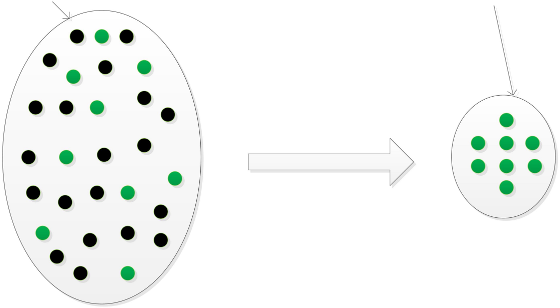
Asymptotic Behavior

of Random Binary Source of Information

- Green markers represent the members of the typical set

Before data compression:
Requires M bits
Total size of the set: 2^{M}

After data compression:
Requires M× H(X) bits
Total size of the set: $2^{M\times H(X)}$



Members of the Typical Set

Remaining, belong to subsets with vanishing probability as N becomes large

Source Coding

Solves the Problem of Data Compression

- Many digital data streams contain a lot of redundant information. For example, a digital image file may contain more zeros than ones: $0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1$
- We wish to squeeze out the redundant information to minimize the amount of data needed to be stored or transmitted
- Definition of Data Compression Problem:
  1. What are the good algorithms that achieve the maximal data compression?
  2. What is the maximum data compression that can be achieved if we want to recover the exact bit sequence after decompression?
- Importance of Data Compression Problem:
  - ▷ Cannot be overstated given so much data is getting uploaded/downloaded and stored in today's world of YouTube, Facebook, etc.